



HAL
open science

Evolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel

Katy Paul

► **To cite this version:**

Katy Paul. Evolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel. Génétique animale. Université Paris-Saclay, 2022. Français. NNT : 2022UPASB067 . tel-03948602

HAL Id: tel-03948602

<https://pastel.hal.science/tel-03948602>

Submitted on 20 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel

*Evolution of genetic diversity in farmed
rainbow trout populations*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 581 agriculture, alimentation, biologie,
environnement et santé (ABIES)

Spécialité de doctorat : Génétique animale

Graduate School : Biosphera. Référent : AgroParisTech

Thèse préparée dans l'UMR **GABI** (Université Paris-Saclay, INRAE, Agroparistech), sous
la direction de **Florence PHOCAS**, Directrice de Recherche

Thèse soutenue à Paris-Saclay, le 22 décembre 2022, par

Katy PAUL

Composition du Jury

Membres du jury avec voix délibérative

Xavier ROGNON

Professeur, AgroParisTech (Université Paris-Saclay)

Président

Bertrand BED'HOM

Professeur, MNHN

Rapporteur & Examineur

Zulma VITEZICA

Professeure, INP Toulouse

Rapporteur & Examinatrice

Pascale LEROY

Directrice de Recherche, INRAE (Bretagne-Normandie)

Examinatrice

François ALLAL

Chercheur, Ifremer

Examineur

Titre : Evolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel

Mots clés : Aquaculture, domestication, sélection, consanguinité, dominance, variabilité génétique

Résumé : L'amélioration génétique de populations de truites arc-en-ciel a permis d'accroître l'efficacité de la production, mais a contribué à une forte diminution de la variabilité génétique. Cette perte de diversité génétique peut causer de la dépression de consanguinité mais aussi mettre en péril le progrès génétique à long terme et l'adaptation des populations aux changements environnementaux. Les niveaux de consanguinité observés (via des indicateurs génomiques) au sein de nos populations de truites françaises, nous interrogent sur les causes et les impacts de tels niveaux de consanguinité. Ces interrogations font l'objet des travaux de thèse présentés. Au-delà des effets de la sélection et de la dérive génétique liée à la faible taille des populations en élevage, la perte de diversité génétique au sein des populations de truite arc-en-ciel s'explique aussi par des déséquilibres de contribution des mères à la génération suivante associés à une surmortalité embryonnaire observée de la fécondation à l'éclosion dans les descendances de certaines mères. L'étude des effets locaux de la consanguinité sur les performances de taille, reproduction et survie de la truite, montre que localisée dans certaines parties du génome, la consanguinité peut avoir des effets notables, sans que l'on observe forcément un effet à l'échelle de tout le génome. Si, ces effets locaux peuvent être négatifs, ils sont toutefois

positifs dans certaines zones du génome. Le principal phénomène observé est un impact négatif de la consanguinité récente, expliqué par l'accumulation d'allèles récessifs dont les effets sont défavorables, en lien direct avec l'estimation d'effets de dominance importants sur les caractères de reproduction et de survie. Cependant la consanguinité, en particulier accumulée dans des générations passées, engendre aussi des effets positifs. Ainsi neuf régions d'homozygotie forte ont été identifiées sous sélection positive commune à trois lignées de truites françaises et une population américaine. Les gènes annotés dans ces régions sont liés à la domestication (fonctions de croissance, reproduction, comportement et immunité) ainsi qu'à la fitness naturelle en lien avec la survie des individus (embryogenèse, organisation cellulaire). C'est pourquoi, comprendre l'impact et le rôle de la diversité génétique locale est essentiel, afin de mieux la gérer dans les programmes de sélection. L'utilisation d'indicateurs génomiques de la consanguinité permettra aux sélectionneurs de mieux maîtriser les compromis entre valeurs génétiques des reproducteurs et effets indésirables de la consanguinité de leur progéniture. Cela permettra aussi d'exploiter les effets d'hétérosis, et ainsi améliorer durablement les performances des cheptels de production.

Title : Evolution of genetic diversity in farmed rainbow trout populations

Keywords : Aquaculture, domestication, selection, inbreeding, dominance, genetic variability

Abstract: Genetic improvement of rainbow trout populations has increased production efficiency, but has contributed to a sharp decrease in genetic variability. This loss of genetic diversity may induced inbreeding depression but also compromised long-term genetic progress and population adaptation to environmental changes. The inbreeding levels observed (through genomic indicators) in our French trout populations, question us on the causes and impacts of such inbreeding levels. These questions are the subject of the thesis work presented. In addition to the effects of selection and genetic drift due to the small size of breeding populations, the loss of genetic diversity in rainbow trout populations may also be explained by imbalanced contribution of mothers to the next generation associated with excess of embryonic mortality observed from fertilization to hatching in the offspring of some mothers. The study of the local effects of inbreeding on the size, reproduction and survival performance of trout, shows that localized in certain regions of the genome, inbreeding may have significant effects, without necessarily observing a genome-wide effect. While these local effects may be negative, they may also be positive in some genomic regions. The main

phenomenon observed is a negative impact of recent inbreeding, explained by the accumulation of recessive alleles with unfavorable effects, directly related to the estimation of important dominance effects on reproduction and survival traits. However, inbreeding, especially accumulated in past generations, also generates positive effects. Thus, nine regions of strong homozygosity have been identified under positive selection common to three French trout lines and one American population. Genes annotated in these regions are related to domestication (growth, reproduction, behavior and immunity functions) as well as to natural fitness related to individual survival (embryogenesis, cell organization). Therefore, understanding the impact and role of local genetic diversity is essential in order to better manage it in breeding programs. The use of genomic indicators of inbreeding will allow breeders to better control the trade-offs between genetic values of the breeding stock and the undesirable effects of inbreeding on their progeny. This will also allow to exploit heterosis effects, and thus improve the performance of production stocks in a sustainable way.

REMERCIEMENTS

La thèse est un long chemin qui ne se fait pas seul. Mes premiers remerciements vont donc à ma directrice de thèse, **Florence Phocas**. Merci pour ton encadrement au-delà de mes espérances. J'ai beaucoup appris grâce à toi ; merci pour toute ton attention, pour le temps que tu as consacré à me relire, à discuter, à me soutenir et à m'encourager, notamment lors des présentations orales de mes travaux. Merci pour toutes tes remarques, toujours constructives, ayant permis d'améliorer mon travail. Ma thèse n'aurait jamais eu cette qualité sans ton aide.

Je souhaite également remercier l'école doctorale **ABIES** et le **département de Génétique Animale** de l'INRAE pour avoir financé ma thèse. Je tiens à remercier l'ensemble du personnel d'ABIES pour sa gentillesse, son suivi et pour le temps accordé à chacun de ses doctorants.

Merci à mes rapporteurs, **Bertrand Bed'Hom** et **Zulma Vitezica**, d'avoir acceptés de relire et d'évaluer mon travail de thèse. Merci également à **Xavier Rognon**, **François Allal** et **Pascale Le Roy** d'avoir accepté d'en être les examinateurs.

Merci aux membres de mon comité de thèse pour leurs conseils, **Xavier Rognon**, **Silvia Rodríguez-Ramilo**, **Gwendal Restoux** et **Pierre Boudry**. Merci pour votre bienveillance, ainsi que de m'avoir suivie tout au long de ces trois années de thèse.

Merci à l'unité **GABI** et tout particulièrement l'équipe **GenAqua** qui m'a accueillie et a surtout supporté mes monologues pendant 3 ans. Donc, merci à **Mathilde**, « l'amoureuse » des chats et des mojitos ; **Delphine**, la 'vraie' parisienne, pour ton sourire et ta gentillesse ; **Jonathan**, pour toutes ces pauses-café et tes talents de pâtissier, ma première année de thèse n'aurait pas été pareille sans toi ; **Henri** le « Roi de la blague » qui m'a permis d'avoir des plantes vivantes dans le bureau ; **Simon**, pour les pauses-café, tes précieux conseils, corrections et pour m'avoir rappelé de temps en temps que le sport ce n'est pas pour moi ; **Mickaël**, pour m'avoir appris à être presque incollable au puzzle lors des pauses-café, on pourra retenir que j'ai encore du travail avant de dépasser le niveau d'Elodie ; **Tristan**, pour toutes ces discussions passionnantes toujours autour d'un café ou d'une chicorée ; **Francine**, pour ta gentillesse et surtout tes petits gâteaux aux marrons ; **Elodie**, comparse de Doc'Jeux et créatrice de gâteaux bien trop caloriques ; **Clémence**, maintenant expat' et plus pipelette que moi ; **Valentin**, peut-être futur doctorant ('on croise les doigts') ; **Audrey**, pour toutes nos longues discussions sur tout : musique, dessins, japon et chats. Merci à **Carole** et aussi à **Nicolas**, le « zen » de l'équipe, pour leur gentillesse, j'espère que tout vas pour le mieux. Merci aussi à **Martin**, **Floriane** et **Olivia**, supers post-doc et stagiaires venus quelques mois à GenAqua.

Et un grand merci à l'équipe poisson de Palavas pour les moments de convivialité passés lors de quelques séminaires (**Marc, François, Maeva, Emilie, Damien, Isabelle, Xavier, Frédérique**, etc...). Je tiens également à remercier l'équipe GIBBS (**Denis, Tatiana, Xavier, Marco, Wendy, Michèle, Mathieu, Maria, Gwendal** etc...), qui m'a accueillie pour mon stage de fin d'études et m'a encouragée à réaliser cette thèse. Et surtout merci à **Anna-Charlotte** et **Audrey**, mes premières co' de bureau qui m'avaient précédée dans cette 'bêtise' !

Merci aux doctorants de GABI, et aux participants des Doc'Jeux, je tiens à remercier particulièrement **Alicia**, une superbe rencontre. Mais je n'oublie pas **Céline, Elitza, Solène, Maria, Quentin** et **Fanny**.

Merci à mes 'vieilles' copines, **Valentine** et **Zoé**, pour nos week-ends jeux de société jusqu'au bout de la nuit (promis JB on ne fera plus de Laser-Game). Et aussi à **Marie D.** partie à l'autre bout de la France faire sa thèse !

Merci à mes copines de prépa : **Marie !L!**, qui m'a suivie dans la folie de la thèse ; **Élise**, la plus courageuse des vétos, promis j'arrive à Coimbra ! ; **Nathalie**, la baroudeuse de l'incroyable ; **Louise**, la plus gentille ; **Sarah**, la plus cool ; **Gwen**, la plus Bretonne ; **Lisa** la plus drôle. En-tout-cas, ce qui vous définit toutes, c'est bien votre gentillesse et surtout votre humour. J'ai hâte de toute vous revoir bientôt !

Merci à ma famille, merci à **mes parents**, merci de m'avoir permis de faire ces années d'études et merci pour votre soutien sans failles (même sans comprendre tout ce que je fais). Merci à mes sœurs, **Audrey, Julie** (correctrice des coquilles) et à ma jumelle, **Sandrine**, d'être là depuis toujours. Bien sûr, merci aussi à tous **mes beaux-frères** et à mes deux (bientôt trois) **neveux et nièces** adorés. Merci à mon **papy Paul**, à ma **mamie Guillou**, qui m'avaient accueillie chez eux, à tour de rôle, il y a quelques années pour des semaines de révision. Merci aussi à mon **papy Guillou** et à ma **mamie Paul** à qui je pense fort, et qui veillent sur moi de là-haut. Merci à mes **oncles, tantes, cousins** et **cousines** d'être simplement vous.

Merci à **ma belle-famille**, de m'avoir si bien accueillie dans votre grande famille.

Merci à toi **Charles**, d'être simplement là. Merci de me faire rire si souvent, pour nos discussions infinies, sur tout et (surtout !) n'importe quoi. Alors ne change jamais, j'ai hâte de vivre encore plein d'aventures avec toi.

Merci à tous ceux que je ne cite pas ici, mais que j'ai pu côtoyer au cours de ma thèse et que je porte dans mon cœur.

Merci aussi à toi Ô lecteur égaré qui pour une raison que j'ignore a décidé de lire cette thèse.

TABLE DES MATIERES

Résumés	II
Remerciements.....	IV
Table des matières	VI
Liste des abbréviations.....	X
Valorisation : Communications et papiers	XII
Chapitre 1 : Introduction générale	1
1.1. Les populations de truite arc-en-ciel sauvages et domestiquées	1
1.1.1. Intérêt pour l'étude de la truite arc-en-ciel.....	1
1.1.2. Expansion de la truite arc-en-ciel dans le monde	3
1.1.3. La reproduction de la truite arc-en-ciel	5
1.1.4. Le génome de la truite arc-en-ciel	8
1.1.5. Les outils génomiques disponibles	11
1.2. Diversité génétique, et évolution des populations	12
1.2.1. Biodiversité et diversité génétique	12
1.2.2. Evolution des populations et de leur diversité génétique	14
1.2.2.1. Les forces évolutives	15
1.2.2.1.1. <i>La mutation</i>	15
1.2.2.1.2. <i>La migration</i>	16
1.2.2.1.3. <i>La dérive génétique</i>	16
1.2.2.1.4. <i>La sélection</i>	17
1.2.2.2. Les théories de l'évolution	20
1.2.2.2.1. <i>La théorie adaptationniste</i>	21
1.2.2.2.2. <i>La théorie neutraliste</i>	21
1.2.2.3. La domestication.....	21
1.2.3. Conséquences de l'augmentation de la consanguinité.....	23
1.3. Méthode d'évaluation de la diversité génétique et d'identification de signatures de sélection.....	26
1.3.1. Evaluation de la diversité génétique.....	26
1.3.1.1. Indicateurs de diversité génétique	26
1.3.1.1.1. <i>Hétérozygotie observée (H_o) et hétérozygotie attendue (H_e)</i>	26
1.3.1.1.2. <i>Indices de fixation de Wright</i>	27
1.3.1.1.3. <i>La taille efficace (N_e)</i>	28
1.3.1.1.4. <i>Les coefficients de consanguinité et de parenté</i>	30

1.3.1.2. Calculer des coefficients de consanguinité	30
1.3.1.2.1. <i>Sur la base du pedigree</i>	30
1.3.1.2.2. <i>Sur la base de données génomiques</i>	31
1.3.1.2.3. <i>Les ilots de consanguinité (ROH)</i>	32
1.3.2. Estimation des effets de la consanguinité sur des performances.....	39
1.3.3. Méthodes et outils d'identification de signatures de sélection.....	42
1.3.3.1. Signatures de sélection intra-population	44
1.3.3.1.1. <i>Exploitation du spectre de fréquences alléliques</i>	44
1.3.3.1.2. <i>Exploitation du déséquilibre de liaison</i>	45
1.3.3.1.3. <i>Exploitation de la diversité génétique</i>	46
1.3.3.2. Signatures de sélection inter-populations : différenciation entre deux populations.....	46
1.4. Questions soulevées	49
Chapitre 2 : Etude du succès reproductif et de la survie embryonnaire dans une lignée domestiquée de truite arc-en-ciel	53
Introduction et conclusion de l'étude	53
Article	55
Annexes de l'étude	85
Chapitre 3 : Estimation de la consanguinité locale, de ses effets positifs ou négatifs, récents ou anciens	91
Introduction et conclusion de l'étude	91
Article	93
Annexes de l'étude	111
Chapitre 4 : Identification de régions sous signatures de sélection positive ou équilibrante, communes à plusieurs populations de truites arc-en-ciel, liées à la domestication et à la fitness	117
Introduction et conclusion de l'étude	117
Article	119
Listes des annexes de l'étude	169

Chapitre 5 : Discussion générale	173
5.1. Les objectifs des programmes d'amélioration génétique	175
5.1.1. Choix de la stratégie d'amélioration génétique	175
5.1.2. La définition de l'objectif de sélection.....	175
5.1.3. Limiter la dépression de consanguinité et l'érosion de la diversité génétique	176
5.2. L'évaluation génétique des reproducteurs.....	178
5.2.1. Prise en compte des effets de dominance	179
5.2.2. Evaluation des effets de dépression de consanguinité et des anomalies génétiques.....	181
5.2.3. Prise en compte de l'originalité génétique des individus	182
5.3. Le choix des reproducteurs	183
5.3.1. Interaction Génotype x Environnement et conservation de la diversité génétique dans le pool de reproducteurs	183
5.3.2. Elimination des candidats à la reproduction porteurs d'allèles délétères	184
5.3.3. Conservation d'un nombre minimal de reproducteurs	185
5.4. les plans d'accouplement	186
5.4.1. Choix du plan d'accouplement	186
5.4.2. Gestion de la consanguinité.....	187
5.4.2.1. La localisation de la consanguinité dans le génome, une information clef.....	187
5.4.2.2. Quelles sont les méthodes disponibles pour une gestion de la diversité génétique à l'échelle locale des génomes ?	189
5.4.3. Valorisation de l'hétérosis	190
5.4.3.1. Prise en compte des effets d'hétérosis dans un programme d'amélioration génétique en lignée pure.....	190
5.4.3.2. Prise en compte des effets d'hétérosis dans un programme d'amélioration génétique en croisement.....	191
5.4.4. Purge du fardeau génétique	192

5.5. Gestion des déséquilibres de contribution des reproducteurs à la génération suivante	193
5.6. Que faire en cas de situation critique ?	194
Conclusion	196
Références	199
Annexes	221
Liste des Annexes	221
<u>Annexe 1</u> : Annexes du Chapitre 4	222
<u>Annexe 2</u> : Article Bernard et al., 2022 (Associé au Chapitre 4)	253
<u>Résultat annexe 3</u> : Effets de la consanguinité chromosomique sur les performances de production d'une lignée de truite arc-en-ciel françaises	269
<u>Résultat annexe 4</u> : Effets locaux de la consanguinité sur la survie du stade oeillé jusqu'à trois semaines après la première alimentation dans la lignée synthétique INRAE	273

LISTE DES ABREVIATIONS

Les mots en *italique* dans le texte sont des mots non-traduits de l'anglais

ADN : Acide désoxyribonucléique

BV : Valeur génétique additive

CLR : Rapport de vraisemblance composite

d² : Part de variance phénotypique expliquée par les effets génétiques de dominance

DL : Déséquilibre de Liaison

EHH : *Extended Haplotype Homozygosity*

F : Coefficient de consanguinité

FAO : Organisation des Nations Unies pour l'alimentation et l'agriculture

FLK : Test de Lewontin-Krakauer

F_{ROH} : Coefficient de consanguinité basé sur la mesure des ROH

FST : Indice de fixation de Wright

(G)BLUP : *(Genomic) Best Linear Unbiased Prediction*

(G)EBV : *(Genomic) Estimated Breeding value*

GV : Valeur génotypique

GRM : Matrice d'apparentement génomique

GWAS : *Genome Wide Association Study*

h² : Héritabilité au sens strict.

HD : Haute densité

He : Hétérozygotie attendue

Ho : Hétérozygotie observée

IBD : Identique par descendance

iHS : *integrated Haplotype Score*

INRAE : Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

IUCN : Union Internationale pour la Conservation de la Nature

ISU : Index de Synthèse Unique

kb | **Mb** : kilobases (= kilo-paire de bases) | Mégabases

MAF : Fréquence de l'allèle mineur

MD: Moyenne densité

Ne : Taille ou effectif efficace d'une population

OCS : *Optimal contribution selection*

ORI : Indice d'originalité

PROSPER : PROcédure de Sélection Par Epreuves Répétées

QTL : *Quantitative Traits Locus*

rEHH : *relative Extended Haplotype Homozygosity*

ROH : *Runs of Homozygosity* ou ilots de consanguinité

SFS : *Site Frequency Spectrum*

SNP : *Single Nucleotide Polymorphism* ou Polymorphisme Nucléotidique

Ss4R : Quatrième duplication complète du génome des téléostéens

SY : Lignée synthétique et expérimentale INRAE

SYSAAF : Syndicat des Sélectionneurs Avicoles, Aquacoles et Entomocoles Français

TAC : Truite arc-en-ciel

XP-EHH : *Cross-population Extended Haplotype Homozygosity*

VALORISATION : COMMUNICATIONS ET PAPIERS

Articles publiés ou soumis dans une revue à comité de lecture

- **Paul, K.**, d'Ambrosio, J., & Phocas, F. (2021). Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout. *Evolutionary applications*, 15(4), 645-662. Special Issue : 'Genomics for Aquaculture and Livestock breeding improvement' (2022). <https://doi.org/10.1111/eva.13308>
- Bernard, M., Dehaullon, A., Gao, G., **Paul, K.**, Lagarde, H., Charles, M., ... & Phocas, F. (2022). Development of a high-density 665 K SNP array for rainbow trout genome-wide genotyping. *Frontiers in Genetics*, 13, 1664-8021. <https://doi.org/10.3389/fgene.2022.941340>
- **Paul, K.**, Pelissier, P., Goardon, L., Dechamp, N., Danon, J., Jaffrelo, L., Poncet, C., Dupont-Nivet, M., Phocas, F. (*under review*). Maternal and genetic effects on embryonic survival from fertilization to swim up stage and reproductive success in a farmed rainbow trout line. *Aquaculture Reports*.
- **Paul, K.**, Restoux, G., Phocas, F. (*under review*) Genome-wide detection of positive selection signatures shared by four domesticated *Oncorhynchus mykiss* populations. *Molecular ecology*.

Communications orales

- **Paul, K.**, D'Ambrosio, J., Phocas, F. (décembre 2020). Effets de la consanguinité et de la dominance sur les caractères de reproduction femelle chez la truite arc-en-ciel. 3èmes Journées Techniques du Sysaaf, en ligne
- **Paul, K.**, Pelissier, P., Goardon, L., Dechamp, N., Dupont-Nivet, M., Phocas, F. (octobre 2021). Effets de la dominance et des effets maternels sur la survie des œufs embryonnés. 4èmes Journées Techniques du Sysaaf, présentiel Rennes.
- **Paul, K.**, D'Ambrosio, J., Phocas, F. (avril 2021). Molecular inbreeding and genetic load for female reproduction traits in a rainbow trout selected line. European Aquaculture Society 2020, web conference
- **Paul, K.**, D'Ambrosio, J., Phocas, F. (may 2021). Old and recent inbreeding effects on female size and reproduction traits in a rainbow trout selected line. 6th international symposium on Genomics In Aquaculture, web conference

- **Paul, K.**, Phocas, F. (may 2021). Evolution of genetic diversity in farmed rainbow trout lines. Séminaire des doctorants du département de génétique animale, web conference.

- **Paul, K.**, Restoux, G., Phocas, F. (may 2022). Detecting signatures of selection between three french domesticated and North-American wild populations of rainbow trout. 7th international symposium on Genomics In Aquaculture, in person, Grenade.

Poster

-**Paul, K.**, Pelissier, P., Goardon, L., Dechamp, N., Dupont-Nivet, M., Phocas, F. (october 2021) Sweepstakes reproductive success from fertilization to hatching in a domesticated line of *Oncorhynchus mykiss*. European Aquaculture Society 2021, in person, Madera.

Chapitre 1

Introduction Générale

1.1. Les populations de truite arc-en-ciel sauvages et domestiquées

1.1.1. Intérêt pour l'étude de la truite arc-en-ciel

La truite arc-en-ciel (TAC), *Oncorhynchus mykiss*, est une espèce d'eau douce, originaire de la péninsule du Kamtchatka en Russie et d'Amérique du Nord (de l'Alaska jusqu'au Mexique ; Guyomard & Vandeputte, 2010). Elle est présente dans la quasi-totalité des cours d'eau nord-américains (Okazaki, 1984 ; Smith & Stearley, 1989). Selon la localisation des rivières et les conditions environnementales rencontrées (distance à la mer, présence d'obstacles à la migration ou de lacs à l'intérieur d'un bassin versant), les populations de TAC peuvent être de forme résidente des lacs et des cours d'eau ou de forme anadrome (allant à la mer) (McPhail, 2007), voire constituées d'une proportion variable d'individus anadromes (Kendall et al., 2015). La TAC est actuellement élevée dans tous les continents sauf l'Antarctique (Figure 1). La truite arc-en-ciel fait partie des quinze espèces de poissons qui dominent la production mondiale piscicole continentale, marine et côtière (FAO, 2022). De plus, la TAC est la principale espèce de poisson élevée en eau douce froide dans le monde avec 739 500 tonnes produites en 2020 (en eau douce) (FAO, 2022). La production de TAC peut également se faire dans des élevages en mer, mais 77% de la production mondiale totale de TAC en 2020 (959 600 tonnes) provenait d'élevages continentaux en eau douce (FAO, 2022).

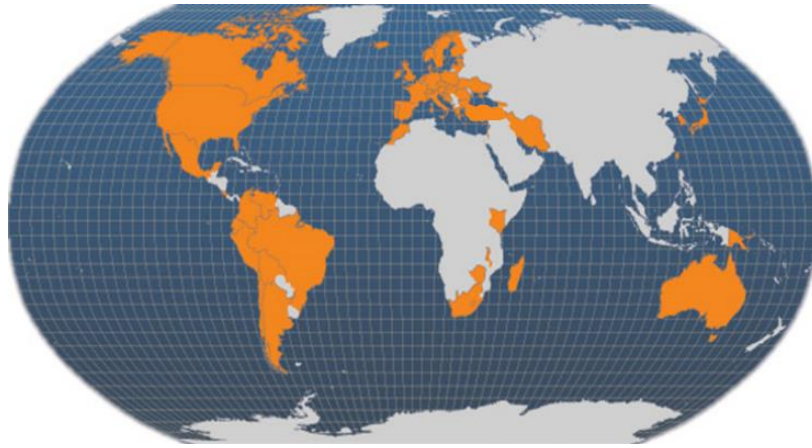


Figure 1. Principaux pays producteurs de truite arc-en-ciel (Source : *Statistiques des Pêches, FAO, 2022*).

La France se classe 8^{ème} parmi les principaux pays producteurs de TAC (élevage en eau douce et/ou en mer), le premier étant l’Iran (191 820 tonnes produite en 2020) suivi de la Turquie (144 283 tonnes), de la Norvège (96 131 tonnes), et du Chili (87 724 tonnes) (https://www.fao.org/fishery/statistics-query/en/aquaculture/aquaculture_quantity). Alors qu’en Iran, en Turquie ou en France la TAC est principalement élevée en eau douce, en Norvège et au Chili la production de TAC se fait principalement en mer. En France, il s’agit de la principale espèce piscicole élevée, et représente à elle seule 78% de la production (Agreste, 2020 ; FAO, 2022).

La perte de diversité génétique fait partie des nombreux enjeux auxquels l’aquaculture doit faire face, car elle met en péril sa résilience face aux changements globaux (FAO, 2022). Chez la TAC, les premiers programmes de sélection commerciale ne débutèrent que dans les années 1970 en Europe du Nord. En France, ils se sont mis en place au début des années 1990 (Haffray, 2019). Aujourd’hui, quatre entreprises de sélection françaises (les Sources de l’Avance filiale du groupe Aqualande, Milin Nevez filiale de Bretagne Truite, Viviers de Sarrance et Les Fils de Charles Murgat) exportent à travers le monde des œufs et alevins améliorés. La perte de diversité génétique observée dans les populations françaises domestiquées, qu’elles soient des lignées commerciales sélectionnées ou des lignées expérimentales (D’Ambrosio et al., 2019) nous interroge sur les moyens à mettre en œuvre rapidement pour préserver leur diversité génétique et garantir leur adaptation et amélioration génétique à long terme face aux changements globaux.

Par ailleurs, la TAC est un excellent modèle amplement étudié pour conduire de multiples recherches (cancérologie, toxicologie, immunologie, physiologie, nutrition, biologie du développement), en raison de sa facilité d'élevage, sa fécondation externe ainsi que sa grande taille qui facilite notamment l'étude de phases précoces du développement ou certaines opérations chirurgicales (Thorgaard et al., 2002). En génétique, la truite est également intéressante du fait de l'originalité de son génome, et de la possibilité de produire des animaux au génome complètement homozygote (haploïde doublé) ou avec plusieurs jeux de chromosomes (polyploïdes) (Fraslin, 2018). En France, il existe des lignées expérimentales utilisées dans diverses études pour améliorer les connaissances biologiques actuelles, notamment sur les caractères de santé (Fraslin et al., 2018) et d'adaptation (Lagarde et al., 2023) ou bien encore quantifier les différences génétiques associées aux mécanismes épigénétiques (Lallias et al., 2021).

1.1.2. Expansion de la truite arc-en-ciel dans le monde

La truite arc-en-ciel (TAC), *O. mykiss*, appartient à la famille des salmonidés (*Salmonidae*) et à la sous-famille des *Salmoninae*. La sous-famille des *Salmoninae* regroupe trois principaux genres : *Salvelinus* (les ombles), *Salmo* (regroupant les saumons et truites d'Atlantique) et *Oncorhynchus* (regroupant les saumons et les truites du Pacifique). Initialement appelée *Salmo gairdneri*, pour les TAC d'Amérique du Nord, son appellation scientifique a été changée pour *Oncorhynchus mykiss* sur la base d'informations moléculaires (Okazaki, 1984 ; Smith & Stearley, 1989). L'espèce était à l'origine rattachée au genre *Salmo* en raison de sa ressemblance physiologique avec le saumon de l'Atlantique. Or, il a été démontré qu'en réalité la TAC était génétiquement plus proche du saumon du Pacifique que du saumon de l'Atlantique (Guyomard et Vandeputte, 2010).

Les premières traces de domestication de la TAC remontent aux années 1870 au nord de la Californie. Elle aurait été domestiquée à partir de la population sauvage présente dans le bassin de la rivière McCloud (Hershberger, 1992 ; Gall & Crandell, 1992). Les souches de TAC actuellement domestiquées dériveraient principalement de populations prélevées dans ce bassin (Dollar & Katz, 1964 ; Thorgaard, 1983). Dans les années qui suivirent sa domestication, la TAC fut exportée à travers le monde (Teletchea, 2018). Selon Bean (1895), la France a été l'un des premiers pays à l'introduire, via une cargaison expédiée à l'aquarium du Trocadéro à Paris. Cette introduction aurait été suivie par de nombreuses autres, d'origines légèrement différentes

(Guyomard et Vandeputte, 2010). Au même moment, des cargaisons d'œufs ont également été expédiées au Japon et en Allemagne (von Behr, 1882 ; Stanković et al., 2015). Ainsi, que dans d'autres pays d'Europe quelques années plus tard à la fin du XIX^{ème} siècle (Royaume-Uni, Suisse, Belgique, Italie, Portugal, Irlande) (Baird, 1886 ; Mather, 1886 ; McDonald, 1893 ; Ravenel, 1899). En parallèle, des transferts de TAC intra-Europe ont eu lieu, principalement à partir des écloseries allemandes vers l'Autriche, la Bulgarie, le Danemark, la Pologne, la Suède, la Suisse, la Russie, la Slovaquie, la Slovénie, la République tchèque, etc. (Crawford & Muir, 2008). Des transferts vers d'autres parties du monde ont également rapidement suivi, depuis la Grande-Bretagne vers le Sri Lanka (Jinadasa et al., 2005), la Nouvelle-Zélande (Scott et al., 1978), l'Argentine (Riva Rossi et al., 2004) et la Chine (Zhao et al., 2008). Au début du XX^{ème} siècle, de nouveaux transferts de TAC depuis les Etats-Unis eurent lieu vers l'Autriche, les Pays-Bas, la République tchèque, la Slovaquie, la Pologne et la Bosnie (Ehrenbaum, 1926 ; Leach, 1924 ; O'Malley, 1927).

Dans tous ces pays, la TAC a été introduite en milieu naturel et s'est naturalisée dans de nombreuses régions du monde (Stanković et al., 2015). Une population non endémique est considérée comme naturalisée si elle prospère dans un état de vie libre, sans soutien et indépendamment d'interventions humaines continues (Copp et al., 2005). Les populations naturalisées restent rares en France ; la seule bien répertoriée se trouve dans le réservoir artificiel des Bouilloises, dans les Pyrénées où elle a été introduite en 1958 (Dumas, 1976 ; Baran et al., 1995).

Cette naturalisation de la TAC peut engendrer des impacts généralement négatifs sur la faune piscicole indigène, particulièrement pour les autres salmonidés ; la TAC devient alors une espèce invasive (Cambray, 2003 ; Stanković et al., 2015). Par exemple, cette naturalisation a posé problème en Croatie où la prédation de la TAC sur les populations endémiques de vairons a conduit à l'extinction de certaines populations (Stanković et al., 2015). Des observations contradictoires ont été menées quant aux risques de compétition entre TAC et truite fario (*Salmo trutta forme fario*) indigène en Europe. Honsing-Erleburg (2005) décrit une compétition forte entre les deux espèces en Autriche, avec une inversion totale des effectifs en faveur de la TAC 10 ans après son introduction. C'est pourquoi certains pays comme la Norvège ou l'Autriche, ont classé la TAC dans la liste noire des espèces invasives (Stanković et al., 2015). Cependant une autre étude montre qu'au contraire c'est la truite fario qui parvient par compétition interspécifique à supplanter la TAC (Baran et al., 1995). De plus, il a été montré que la TAC

présente un faible risque pour le biote et les écosystèmes aquatiques indigènes en raison de son incapacité à établir des populations autonomes dans la majeure partie des environnements. Elle est ainsi parfois recommandée à l'introduction pour la pêche de loisir dans les rivières françaises pour soulager sensiblement la pression de pêche sur les truites fario indigènes (<https://peche-hautes-alpes.com/hautes-alpes-la-truite-arc-en-ciel-est-elle-un-danger-pour-notre-truite-fario/>), expliqué par le fait que la TAC semble plus facile à pêcher que la truite fario (Dumas, 1976). De plus, cela est rendu possible par l'incapacité des hybrides issus du croisement entre la TAC et la truite fario à survivre dès les premiers stades de développement (Suzuki & Fukuda, 1971). Plus généralement, les possibilités d'hybridation de la TAC avec d'autres salmonidés sont assez faibles (Chevassus et al., 1975). Par ailleurs, les risques liés aux échappements dans le milieu naturel de TAC issues d'élevages de production sont maintenant également limités en France par le fait qu'une grande part de la production vient de l'élevage de TAC triploïdes stériles ou de populations monosexes femelles, plus intéressantes que les populations bisexuées pour la production (moindre mortalité, meilleure croissance et meilleure qualité du produit final). Le caractère invasif de la TAC est donc à mettre en regard avec le lieu de sa naturalisation, la faune présente ainsi que le type de truites introduites (diploïdes femelles ou triploïdes stériles).

1.1.3. La reproduction de la truite arc-en-ciel

Comme chez tous les poissons ovipares, la TAC se reproduit par fécondation externe et la rencontre des gamètes doit intervenir rapidement après leur émission dans l'eau. En milieu naturel, la femelle truite arc-en-ciel peut pondre de façon séquentielle, lui permettant de se reproduire avec plusieurs mâles. Généralement, les femelles pondent jusqu'à 2000 œufs par kilogramme de poids vif. Après l'ovulation, les œufs sont conservés dans la cavité abdominale des femelles, ils baignent dans le liquide cœlomique (milieu isotonique présentant des activités antiprotéase et antibactérienne) permettant de les conserver (Bobe et al., 2010). L'ovulation et la ponte des œufs peuvent intervenir à plusieurs jours d'intervalle. Toutefois, après 8 jours dans la cavité abdominale, la qualité des ovules diminue significativement. Après la ponte, sans intervalle de temps, la fécondation des œufs advient (Bobe et al., 2010).

Habituellement, la durée du cycle reproducteur est d'un an, même si cette période peut être réduite sous photopériode artificielle (Bobe et al., 2010). Les populations naturelles de

TAC d'Amérique du Nord se reproduisent au printemps. Or, en Europe, la plupart des populations de TAC domestiquées ont une période de reproduction différente, qui s'étend du milieu de l'automne au début de l'hiver. Cela est probablement dû aux températures automnales plus clémentes en Europe qu'en Amérique du Nord, autorisant une ponte plus précoce car la gamétogenèse n'est pas bloquée (Bobe et al., 2010). Néanmoins, il existe également en Europe des lignées dont la reproduction se déroule au printemps, comme pour les populations naturelles d'Amérique du Nord.

L'âge de la puberté des TAC dépend avant tout de leur sexe et de leur croissance associée aux conditions d'élevage (alimentation et température), mais aussi de leur génétique (Breton et al., 2010). Les mâles atteignent généralement la puberté un an avant les femelles, c'est-à-dire à un ou deux ans contre deux ou trois ans pour les femelles. Les TAC sont des poissons gonochoriques, soit mâles, soit femelles qui gardent le même sexe tout au long de leur vie. Chez la truite arc-en-ciel, le sexe est gouverné par un système chromosomique simple (XX/XY) dont le déterminant majeur est connu (gène *sdY*, Yano et al., 2012). Cependant, des individus spontanément masculinisés sont fréquemment observés au sein des populations XX. Des travaux expérimentaux ont montré que le taux de masculinisation spontanée dépendait à la fois de facteurs génétiques et de la température d'alevinage (Quillet et al., 2002 ; Valdivia et al., 2014 ; Fraslin et al., 2020).

Par ailleurs, le sexe phénotypique (mâle ou femelle) peut être inversé par rapport au sexe génétique (XY ou XX) par traitement hormonal à la première alimentation des alevins. Ainsi, la masculinisation est possible grâce aux androgènes, ainsi que la féminisation grâce aux œstrogènes. En élevage, la masculinisation est recherchée afin de produire ce qu'on appelle des néo-mâles. Ce sont des individus avec un phénotype mâle mais un génotype femelle (XX). Ainsi, lorsqu'on les croise avec des femelles normales (génétique et phénotype), ces néo-mâles ne produisant que des gamètes « X », permettent d'obtenir des populations monosexes femelles (génotype + phénotype femelle). Actuellement, pour des raisons de facilité de gestion des élevages, mais également pour garantir une meilleure qualité de produit, due à une maturation sexuelle plus tardive, la production des TAC se fait en grande majorité en élevage mono-sexe femelles (90%).

En élevage, la TAC ne rencontre pas les conditions pour une reproduction naturelle, et les juvéniles sont donc obtenus par fécondation artificielle. Pour cela, les gamètes matures sont prélevés par massage abdominal sous légère anesthésie, puis une fécondation artificielle est

réalisée par l'éleveur (Bobe et al., 2010). Le processus de fécondation artificielle permet de réaliser facilement les plans d'accouplements souhaités, qu'ils soient hiérarchiques (chaque père est croisé avec une ou plusieurs femelles) ou factoriels (chaque père est croisé avec chacune des mères et inversement). Il est ainsi aisé d'accoupler les reproducteurs choisis et donc de gérer facilement le progrès génétique et la diversité génétique au sein des populations (Dupont-Nivet et al., 2006). En France, les accouplements des reproducteurs au sein des programmes de sélection se font sur le modèle de plans factoriels incomplets (cf. Chapitre 5). Lors des fécondations artificielles, la ponte de chacune des femelles est divisée en autant de lots que de mâles utilisés pour féconder les œufs, car plusieurs études ont mis en évidence une compétition spermatique entre les mâles. Cela se traduit par une contribution supérieure de certains mâles lorsque les laitances de plusieurs mâles sont mélangées (Parker, 1998 ; Kaspar et al., 2008). Cela permet donc de s'affranchir d'un mécanisme responsable d'une variance importante du succès reproductif sur la voie paternelle engendrant des tailles de famille variables susceptibles d'induire une perte de diversité génétique au cours des générations. De plus, dans le cadre des schémas de sélection de type PROSPER, Chevassus et al. (2004) ont suggérés de féconder et de regrouper ensemble les œufs des mères aux tailles d'œufs similaires. Et cela afin de s'affranchir de la variance du succès reproductif sur la voie maternelle due à une croissance initiale variable entre les alevins originaires d'œufs de tailles différentes.

Le moment précis de l'ovulation des femelles est peu prévisible. Or, un prélèvement trop tardif des ovules peut poser des problèmes de vieillissement post-ovulatoire excessif, entraînant une baisse de leur qualité, se traduisant par une chute de la fécondabilité, ainsi que l'apparition de graves défauts de développement (mortalité ou malformations embryonnaires). Chez la truite, une durée de cinq jours à 10-12°C entre l'ovulation et la fécondation permet d'obtenir une qualité ovocytaire maximale (Sakai, 1975 ; Springate et al., 1984). Sakai (1975) observe que les premiers jours après l'ovulation, le taux de fécondation est réduit, probablement dû à quelques œufs n'ayant pas fini leur maturation, ainsi que cela a été confirmé suite à l'examen d'ovules de poissons sacrifiés juste après ovulation (Wallace & Selman, 1978). Pour s'affranchir du problème du vieillissement post-ovulatoire, il est possible de synchroniser les pontes des femelles, grâce à une induction hormonale (Bobe et al., 2010). Cela est fait en élevage, et permet de connaître le moment précis de l'ovulation des femelles, qui se fera simultanément. Cela permet également d'éviter la manipulation répétée des femelles afin de vérifier l'ovulation de chacune d'elles.

Après fécondation, le développement de l'embryon de TAC peut être divisé en quatre phases jusqu'au stade de l'alevin (Vernier, 1969) : la segmentation, la gastrulation, la période du bourgeon causal et la résorption de la vésicule vitelline. La durée des étapes du développement embryonnaire dépend de la température ambiante. C'est pourquoi, dans la littérature, les durées de développement sont généralement exprimées en degré - jours. De la fécondation à l'éclosion, on considère qu'il faut entre 290 et 370 degrés - jours selon la lignée (Bobe et al., 2010). Sachant que la plage de températures de développement embryonnaire normale se situe entre 4 et 13°C, il faudra entre minimum 22 jours et maximum 92 jours de la fécondation à l'éclosion. Ce n'est qu'après l'éclosion que les branchies et la bouche des alevins se développent, lors de la résorption de la vésicule vitelline (Bobe et al., 2010). Les alevins commencent alors une période de nourriture mixte (endogène et exogène), qui durera entre 42 et 63 degrés - jours (Escaffre & Bergot, 1984).

1.1.4. Le génome de la truite arc-en-ciel

Chez tous les vertébrés, on retrouve la trace de deux duplications complètes du génome (hypothèse 2R initialement proposé par Ohno, 1970) ayant eu lieu chez les Chordés il y a environ 450-500 millions d'années (Dehal & Boore, 2005). Chez les poissons téléostéens, une troisième duplication du génome entier est survenue (Ts3R). Et une 4^{ème} et dernière duplication complète du génome (Ss4R) a eu lieu chez les salmonidés il y a environ 96 millions d'années (Figure 2; Berthelot et al., 2014 ; Macqueen & Johnston, 2014).

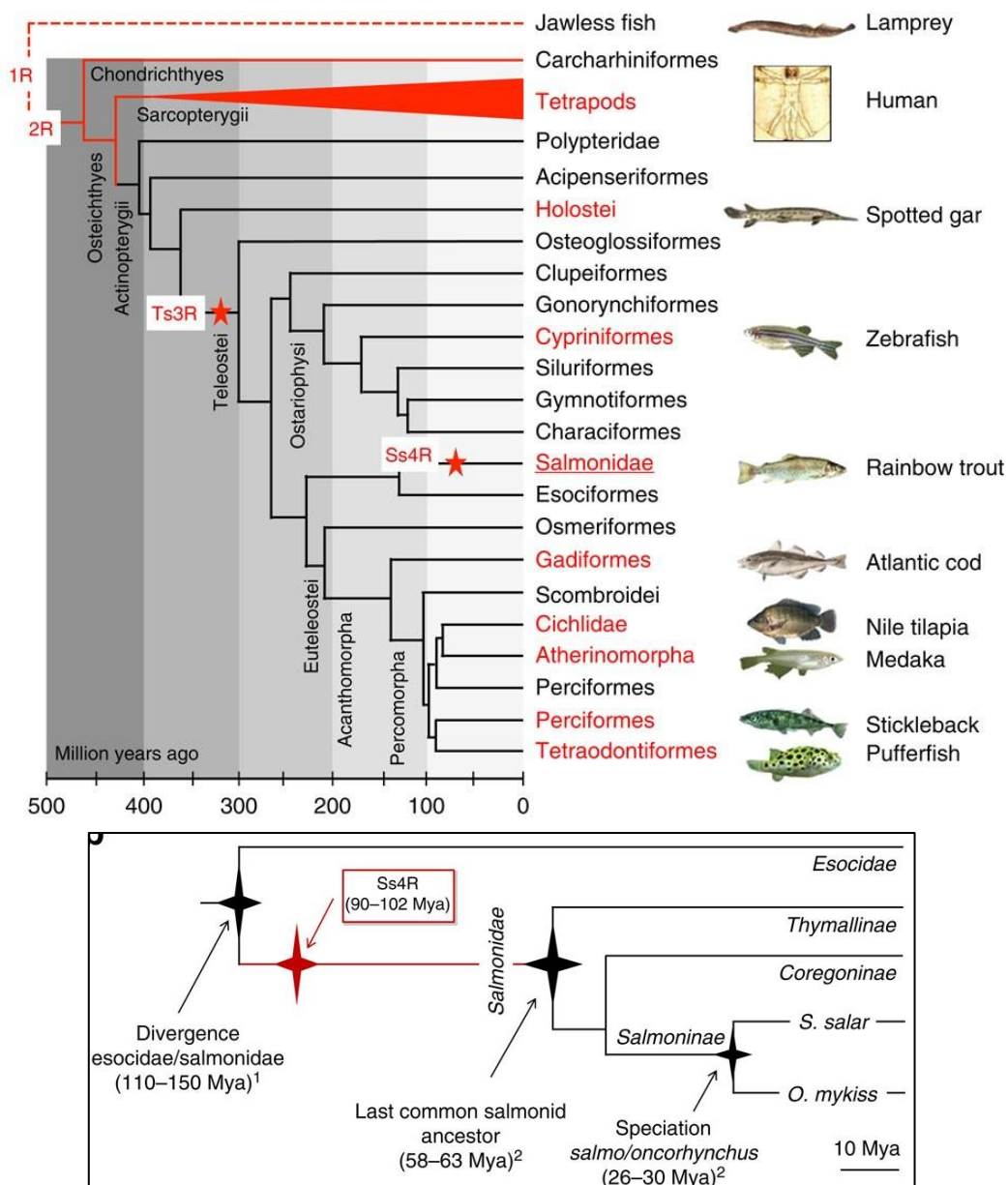


Figure 2. Datation des différentes duplications complètes du génome subies par les salmonidés (Source : Figure issue de Berthelot et al., 2014).

La taille du génome de la truite arc-en-ciel est estimée à environ $2,4 \cdot 10^9$ paires de bases (Hardie & Hebert, 2003 ; Gao et al., 2021). Le génome de la TAC est très plastique et un nombre variable de chromosomes est observé entre populations de TAC, avec un nombre de paires de chromosomes variant entre 29 et 32. Toutefois, toutes les truites arc-en-ciel possèdent 104 bras chromosomiques, c'est le polymorphisme robertsonien (fusion/cassure de chromosomes au niveau des centromères) qui fait varier le nombre de paires chromosomiques, comme la

cytogénétique a pu le mettre en évidence (Guyomard et Vandeputte, 2010). Le génome ancestral est supposé être à $n=29$ chromosomes haploïdes (Thorgaard, 1983). S'il existe d'autres espèces pour lesquelles on observe des variations inter-individuelles du nombre des chromosomes (Lukhtanov et al., 2011), ce phénomène est rare, et pour la majorité des espèces, le nombre et la structure des chromosomes ne varient pas.

Ces différences génomiques entre populations de TAC ont conduit à considérer deux génomes de références différents et donc construire deux assemblages distincts du génome de la TAC. Ces assemblages ont été réalisés sur la base du séquençage d'individus de deux lignées double haploïdes androgéniques (YY) américaines (Swanson et Arlee) (Gao et al., 2016, 2021) :

- Le premier génome de référence assemblé a été réalisé à partir d'un individu de la lignée Swanson, originaire de la péninsule de Kenai en Alaska (Omyk_1.0 : GCA_002163495.1, Gao et al., 2016). Cette lignée a $n=29$ paires de chromosomes et est considérée comme proche du génome ancestral de la TAC. Un premier assemblage avait été réalisé à l'INRA (AUL_PRJEB4421_v1, Berthelot et al., 2014) et un second assemblage plus précis à l'USDA (Omyk_2.0, Gao et al., 2016).
- Le second génome de référence, récemment assemblé par l'USDA (Omyk_1.1 : GCA_013265735.3, Gao et al., 2021) a été réalisé sur un individu de la lignée domestiquée Arlee originaire de la côte Nord de la Californie et est constitué de $n=32$ paires de chromosomes.

Les chromosomes omy4, omy14 et omy25 du génome de référence Swanson ont été scindés en deux pour constituer respectivement les chromosomes 30, 32 et 31 du génome de référence Arlee.

Palti et al. (2014) ont réalisé un arbre phylogénétique de diverses lignées de TAC, comparant notamment les lignées Swanson et Arlee, ainsi que des lignées isogéniques INRAE. On peut constater que les lignées françaises sont génétiquement plus proches de la lignée Arlee provenant de Californie. Cela est cohérent avec la connaissance de la diffusion de la TAC en Europe à partir de cet état américain, ainsi qu'avec les analyses de Guyomard et al. (2012) et D'Ambrosio et al. (2019) indiquant qu'au moins le chromosome omy25 du génome de référence Swanson correspondait à deux groupes de liaison distincts dans les populations de TAC françaises.

1.1.5. Les outils génomiques disponibles

Ces dix dernières années, la baisse importante des coûts de génotypage et séquençage a permis l'accès à des ressources génomiques importantes pour les espèces aquacoles, en particulier pour les salmonidés.

Deux puces de génotypages sont ainsi actuellement disponibles pour la TAC. La première, développée par l'USDA, est une puce moyenne densité (MD - 57K marqueurs SNPs) (Palti et al., 2014). La seconde est une puce haute densité (HD - 665K marqueurs SNPs) développée en 2020 à INRAE (Bernard et al., 2022) en collaboration avec l'USDA et le SYSAAF. La puce MD a servi à mes premiers travaux de thèse (Chapitres 2 et 3) et la puce HD à mes derniers travaux de thèse présentés au Chapitre 4. Avec ce second outil, j'ai pu caractériser finement le déséquilibre de liaison entre marqueurs à faible distance dans les populations de mon étude du Chapitre 4. Cette description est présentée comme première valorisation scientifique de la puce HD (Bernard et al., 2022 en Annexe 1).

A Retenir – 1.1. Les populations de truites arc-en-ciel sauvages et domestiquées

- La truite-arc-en-ciel, originaire de la côte Pacifique d'Amérique du Nord et de la péninsule du Kamtchatka en Russie, a été disséminée à travers le monde.
- La domestication de la truite arc-en-ciel est récente (fin du XIX^{ème} siècle).
- La truite arc-en-ciel fait partie des 15 espèces majeures de la production piscicole mondiale (FAO, 2022). L'Union Européenne est le second producteur mondial de TAC, juste derrière l'Iran.
- Le nombre de chromosomes varie de 29 à 32 paires d'une population de TAC à l'autre.
- Deux génomes de référence sont à l'heure actuelle séquencés et assemblés à l'échelle des chromosomes : Swanson (n=29 chromosomes) et Arlee (n=32 chromosomes).
- Deux puces de génotypage sont disponibles : une puce moyenne densité (57K SNPs) et une puce haute densité (665K SNPs).

1.2. Diversité génétique et évolution des populations

1.2.1. Biodiversité et diversité génétique

En juin 1992, le Sommet de la Terre de Rio de Janeiro alerte les Etats et l'opinion publique sur les changements climatiques et les atteintes à la biodiversité, en s'efforçant de définir la notion de développement durable définie par l'ONU (rapport Brundtland de 1987) : « Le développement durable est un développement qui répond aux besoins du présent sans compromettre la capacité des générations futures à répondre à leurs propres besoins ». Cette conférence de Rio reconnaît la préservation de la diversité biologique comme l'un des grands enjeux environnementaux de la planète. Elle met au chapitre 15 de l'Agenda 21 la préservation de la diversité biologique et l'utilisation durable des ressources biologiques en objectifs essentiels des actions à conduire tant au niveau international que national ou local. La biodiversité concerne non seulement la diversité des espèces vivantes, mais aussi de leurs caractéristiques génétiques (variabilité des gènes et des populations intra-espèces), ainsi que la diversité des écosystèmes. La diversité génétique décrit le niveau de la diversité intra spécifique par le nombre (variété) des allèles des gènes présents au sein d'une même espèce, c'est-à-dire par le niveau de polymorphisme des gènes observé dans l'espèce. Elle se distingue de la variabilité génétique, qui mesure le degré de variation du génotype (ensemble des caractéristiques génétiques) entre individus d'une même population (ou espèce) associé ou non à une variation de phénotype (caractère évalué). La diversité génétique représente le réservoir génétique d'une espèce, et peut être définie comme une mesure dérivée de la biodiversité permettant de l'étudier et de la gérer (Pereira et al., 2013).

Les différentes versions et associations alléliques des gènes apparaissent par mutation ou bien par réarrangements chromosomiques (recombinaisons). Actuellement, l'étude de la diversité génétique et du polymorphisme des gènes se fait essentiellement grâce à la détection de sites (= *loci*) polymorphes, et cela grâce à des marqueurs moléculaires. Les marqueurs moléculaires sont associés à des fragments d'ADN correspondant à un locus ou plusieurs *loci* polymorphes. Il y a quelques années, les marqueurs utilisés en routine étaient les microsatellites. Ce sont des motifs très courts (moins de 10 paires de base) hautement répétés (en tandem) et bien répartis le long du génome (chez les salmonidés, les microsatellites sont présents tous les 20-80 kb, Estoup et al., 1993). Le nombre de répétitions de microsatellites à un site donné varie d'un individu à l'autre, permettant d'accéder au polymorphisme de répétitions de chaque individu. Ces dernières années un autre type de marqueur s'est développé, les SNPs (Single

Nucleotide Polymorphism). Ce sont des marqueurs basés sur le polymorphisme d'une base unique, ils sont nombreux et bien répartis le long du génome, en moyenne toutes les 64 paires de bases (Gao et al., 2018) et représentent donc une large partie de toutes les différences génétiques entre individus. Ces caractéristiques font des SNPs le type de marqueurs de prédilection aujourd'hui utilisé. A chaque site polymorphe au niveau de la population, un individu sera soit homozygote si ses deux allèles sont identiques, soit hétérozygote s'il possède deux versions différentes. La diversité génétique varie selon les régions du génome, les gènes considérés ainsi que d'une population à l'autre (Ellegren & Galtier, 2016).

Finalement, la diversité génétique peut être définie comme l'exact opposé de la consanguinité (Lefort-buson, 1988). La consanguinité est issue de l'accouplement d'individus apparentés, c'est-à-dire de deux individus ayant au moins un ancêtre commun. Les descendants sont donc homozygotes à plus de *loci* que ceux issus d'un accouplement d'individus non apparentés. Un individu ayant un génotype (= combinaison d'allèles à un locus) homozygote, du fait de sa consanguinité, peut aussi être appelé « autozygote ». Autrement dit, un génotype autozygote est composé d'allèles identiques (homozygotes) issu d'un même allèle ancestral (= IBD, identique par descendance). A l'échelle de la population, un niveau de consanguinité trop important conduit à une réduction du nombre de sites polymorphes. Toutefois, la consanguinité ne se transmet pas à la descendance (non héritable), si deux individus consanguins mais non apparentés se reproduisent, leurs descendants ne sont pas consanguins.

Il est reconnu que la diversité génétique offre aux espèces une plus grande capacité d'adaptation et de survie, surtout à des changements environnementaux conséquents (Barrett & Schluter, 2008). Un niveau important de diversité génétique contribue par exemple à la capacité de réponse immune à la diversité spatio-temporelle des pathogènes (cas du complexe majeur d'histocompatibilité dans la réponse immunitaire, Bed'Hom & Chazara, 2011), ainsi que contribuer à des mécanismes divers impliqués dans la survie ou la reproduction (Durland et al., 2021). Un large *pool* génétique est également le garant du maintien du progrès génétique à moyen et long terme en sélection. Cela permet d'avoir une probabilité plus importante que certains individus de la population aient des allèles mieux adaptés aux fluctuations de l'environnement et soient donc susceptibles de survivre et de transmettre à leur descendance des caractéristiques génétiques favorables. De plus, il a été observé dans un grand nombre d'espèces animales ou végétales que les individus consanguins présentent une moindre valeur sélective (ou *fitness*), avec un plus grand nombre de malformations, ainsi qu'une survie, une

croissance et une fertilité moins bonnes que les descendants de parents non apparentés. Ce phénomène est appelé dépression de consanguinité (Aulstad & Kittelsen, 1971 ; Kincaid, 1976 ; Su et al., 1996 ; Leroy, 2014 ; Doekes et al., 2019 ; Baes et al., 2019).

Une méta-analyse des effets de dépression de consanguinité menée par Doekes et al. (2021) sur sept espèces d'animaux de rente (bovin, porc, poulet, mouton, chèvre, cheval et lapin) démontre un niveau de dépression de consanguinité similaire ($\approx - 0.28\%$ sur la moyenne du caractère lorsque le coefficient de consanguinité F augmente de 0,01 point) pour divers types de caractères dans les populations animales (survie, reproduction, croissance, production et santé). Les diminutions de performances associées à l'augmentation de consanguinité induisent des pertes économiques plus ou moins importantes pour les sélectionneurs et/ou les producteurs (Leroy et al., 2014).

1.2.2. Evolution des populations et de leur diversité génétique

On définit l'évolution comme la modification des fréquences des allèles au sein d'une espèce ou d'une population. En l'absence de pression des forces évolutives, une population dite « idéale » (c'est-à-dire d'effectif infini, fermée (isolée), avec une reproduction panmictique, sans sélection ni mutation) conserve indéfiniment des fréquences alléliques stables ; elle est dite en équilibre de Hardy-Weinberg (Hardy, 1908 ; Weinberg, 1908). Toutefois, une population « réelle » est soumise à quatre forces évolutives (la dérive, la sélection, la mutation et la migration) qui influencent l'évolution de sa diversité génétique au cours des générations.

Parmi les forces évolutives, certaines peuvent jouer en faveur de la diversité génétique dans les populations. C'est le cas de la mutation, et dans certains cas de la migration. En revanche, dérive génétique, sélection et effet fondateur induisent des pertes de diversité à l'échelle des populations isolées.

Encart 1. Notion d'équilibre d'Hardy-Weinberg

Le modèle d'Hardy-Weinberg est un modèle central de la génétique des populations. La notion d'équilibre d'Hardy-Weinberg (HW) est conditionné à plusieurs hypothèses : La population doit être idéale, de taille infinie et constante, se reproduisant en générations discrètes (pas de chevauchement entre les générations) et en panmixie totale (les couples se forment au hasard). La population doit être non sélectionnée, sans migration, ni mutation. Ainsi sous équilibre d'Hardy-Weinberg, la diversité génétique de la population se maintient et tend vers un équilibre stable.

De cet équilibre d'HW, découle la loi de distribution des génotypes associée à l'allèle A de fréquence p et l'allèle a de fréquence $q=1-p$ (permettant aux fréquences alléliques de rester constantes) : $p^2 + 2pq + q^2 = 1$

Le génotype AA sera de fréquence de p^2 ; le génotype Aa sera de fréquence $2pq$; et le génotype aa sera de fréquence q^2 dans la population.

1.2.2.1. Les forces évolutives

1.2.2.1.1 La mutation

L'apparition de nouvelles mutations permet d'ajouter différentes versions alléliques et donc d'augmenter la diversité génétique au sein d'une espèce ou population. Ce processus est à la base de la variabilité génétique observée dans les populations. Les mutations génétiques peuvent survenir aléatoirement au cours de la réplication de l'ADN lors des divisions cellulaires, mais elles peuvent également être produites lors d'une exposition à des agents mutagènes présents dans l'environnement (UV, agents chimiques, virus, etc). Il existe trois grands types de mutations : le remplacement (substitution), la suppression (délétion) ou l'ajout (insertion) d'une ou plusieurs paires de base (mutations ponctuelles ou tandem de paires de base).

Les remaniements chromosomiques sont des modifications des structures chromosomiques et sont considérés comme des mutations chromosomiques. Ces processus

peuvent correspondre à des duplications, des délétions, des translocations (= échange de fragments d'ADN entre chromosomes), des variations du nombre de chromosomes (fusions, fissions), ou des inversions de segments d'ADN (= changement d'orientation des segments modifiant l'ordre des gènes). La polyploïdisation des génomes est également une forme de mutation (mutations génomiques).

Seules les mutations dans les cellules germinales peuvent être transmises aux descendants de l'individu et donc dans la population. Ces mutations peuvent avoir un effet neutre sur l'organisme, mais également des effets favorables ou défavorables.

1.2.2.1.2. La migration

La migration peut être définie comme une partie (plus ou moins importante) d'une population d'une espèce allant fonder une nouvelle population, ou faire simplement référence aux flux de gènes entre plusieurs populations pré-existantes. Si la migration concerne la création d'une nouvelle population, cette dernière peut être de diversité génétique réduite par rapport à la population initiale car seul un nombre limité de reproducteurs va engendrer les descendants de la nouvelle population ; ce phénomène est appelé effet fondateur (Mayr, 1954, 1963). Généralement, au sein de populations sauvages, c'est l'isolement géographique d'une partie de la population initiale qui peut induire la création d'une nouvelle population. Dans le cadre de la domestication, c'est l'intervention humaine, isolant certains individus d'une population sauvage, qui induit la création d'une nouvelle population.

A l'inverse, dans le cadre de flux de gènes entre deux populations, de nouveaux variants peuvent être introduits dans la population cible si la population migratrice est génétiquement suffisamment différente (différence ayant pu être induite par un isolement géographique), augmentant alors la diversité génétique de la population cible.

1.2.2.1.3. La dérive génétique

La dérive génétique est l'évolution d'une population ou d'une espèce causée par des événements aléatoires, impossibles à prévoir (Kimura, 1968). Dans une très grande population, dite de taille « infinie », fermée et non soumise à sélection, les fréquences alléliques sont relativement stables d'une génération à l'autre. Alors que dans une population de taille finie, les fréquences alléliques sont modifiées aléatoirement d'une génération à l'autre, par un effet

d'échantillonnage appelé « dérive génétique ». Elles augmentent ou diminuent alors de manière aléatoire, à la manière d'un tirage aléatoire sans remise des gènes d'une génération à l'autre. Au fil des générations, on observe la fixation ou la perte aléatoire de certains allèles, aboutissant au monomorphisme des variants dans une population de taille finie (Masel, 2011 ; De Meeûs, 2012). La dérive génétique concerne tous les allèles même si l'impact sur les allèles neutres (c'est-à-dire ne conférant ni avantage ni inconvénient) est plus important. Malgré tout, un allèle favorable peut disparaître ou un allèle défavorable se fixer dans une population par dérive, ce qui est fréquent pour des populations de taille très faible. Plus la taille de la population est réduite, plus le phénomène de dérive est important. Une forte dérive peut rapidement conduire au déclin de la population et dans certains cas à son extinction, en raison d'un risque plus élevé de fixer au hasard des allèles délétères (Smith & Haigh, 1974). Ce mécanisme prend une part importante dans le processus du vortex d'extinction (Frankham et al., 2002).

1.2.2.1.4. La sélection

La sélection naturelle a été pour la première fois décrite par Charles Darwin, naturaliste anglais, dans *l'Origine des espèces* publié en novembre 1859, où il tente d'expliquer l'évolution des espèces ainsi que leur diversité en décodant les mécanismes sous-jacents à la sélection naturelle. Au sein des populations, la sélection naturelle favorise la survie et la reproduction des individus ayant la meilleure valeur sélective (encore appelée valeur adaptative ou *fitness* en anglais), c'est-à-dire les mieux adaptés à leur environnement. En biologie évolutive, la *fitness* est un concept central, mentionné pour la première fois par Herbert Spencer (1864), puis associé au terme de sélection naturelle par Darwin dans la 5^{ème} édition de *l'Origine des espèces* (1869). Dans le cas particulier de la sélection artificielle, ce sont les hommes qui choisissent d'avantager la reproduction de certains individus sur la base de spécificités phénotypiques intéressantes pour l'élevage et la production ou simplement pour des raisons esthétiques (critères de couleur, forme et taille).

Les individus se reproduisant le plus transmettent une part génétique de leurs aptitudes à un plus grand nombre de descendants, permettant sur le long terme l'adaptation de la population (East, 1918 ; Fisher, 1958). La valeur sélective d'un individu est une mesure de son succès reproducteur (Sober, 2001) ; elle dépend principalement de sa survie entre le stade zygote et le stade adulte, et de sa fertilité (nombre de descendants viables capables de se reproduire). Des verrous reproductifs, pouvant induire une variation du succès reproductif entre

les individus d'une population, sont observés à différents stades du cycle de reproduction (maturation des gamètes, ponte, fécondation, développement embryonnaire, croissance des larves, recrutement de la population adulte reproductrice). Ainsi, le nombre de reproducteurs participant à la génération suivante est généralement inférieur au nombre de reproducteurs initial et à la taille de la population réelle. Cela peut conduire à une perte de diversité génétique au sein de la population (Hedgcock & Pudovkin, 2011). Ainsi, chez des espèces marines très fécondes, une très forte variance du succès reproductif individuel est observée, liée aux fortes variations environnementales subies en milieu marin, et impactant la contribution des reproducteurs à la génération suivante. Ce phénomène est appelé « Loterie » du Succès Reproductif ou « *Sweepstakes Reproductive Success* » (SRS) et a été proposé par Hedgcock et al. (1994). Les SRS correspondent à un important différentiel entre la taille réelle de la population (N) et le nombre efficace de reproducteurs pour la génération suivante (Hedgcock & Pudovkin, 2011).

Pour qu'il puisse y avoir sélection sur un caractère, trois conditions sont nécessaires et indispensables : le caractère doit être variable dans la population, héritable et il doit y avoir une corrélation entre le nombre attendu de descendants (*fitness*) et le caractère (Levins & Lewontin, 1985). Quelle que soit la nature de la sélection, naturelle ou artificielle, quatre types de sélection existent, et peuvent affecter les fréquences alléliques d'un locus ou plusieurs *loci* de diverses manières, laissant des signatures différentes dans le génome :

- a) **La sélection directionnelle** est liée à la présence de variants favorables dans la population. Cette forme de sélection est la plus intuitive. En effet, l'allèle conférant une meilleure *fitness* va inévitablement augmenter en fréquence dans la population. Cette forme de sélection favorise des variants allant dans le même sens et agit donc sur la valeur moyenne du caractère. Ce type de sélection peut être positive ou négative, c'est-à-dire sélectionner certains variants avantageux ou au contraire éliminer les variants délétères (sélection purifiante). La sélection artificielle est très généralement un processus de sélection directionnelle.
- b) **La sélection stabilisante** (ou optimisante) favorise les phénotypes intermédiaires en éliminant les phénotypes extrêmes et permet donc le maintien de caractères à des valeurs moyennes au cours des générations. Il en résulte que les variants aux effets les plus extrêmes sont éliminés de la population. Cette sélection agit en réduisant la variance du caractère sans changer sa valeur moyenne.

- c) **La sélection disruptive** (ou diversifiante) favorise au contraire les phénotypes (et les variants génétiques) extrêmes et élimine les intermédiaires. Ce type de sélection se produit lorsque les phénotypes extrêmes ont un avantage sur les intermédiaires (Levene, 1953; Marchinko et al., 2014). Par exemple, une population de prédateurs peut avoir deux types de proies, la sélection disruptive correspond au processus par lequel deux types d'individus deviendront majoritaires, chacun spécialisé sur l'une des deux proies. Les individus intermédiaires seront éliminés car trop généralistes.
- d) **La sélection équilibrante** (ou balançante) favorise les allèles en fréquences intermédiaires (Charlesworth, 2006). Cette sélection peut être expliquée par le mécanisme de surdominance, c'est-à-dire l'avantage à l'hétérozygote. L'exemple de surdominance le plus connu chez l'humain est lié à la maladie génétique de la drépanocytose, due à une mutation récessive qui induit une durée de vie limitée aux individus homozygotes pour cette mutation. Cependant, en parallèle, cette mutation confère une immunité au paludisme, due à la modification de la forme des cellules sanguines. C'est pourquoi, les individus hétérozygotes sont avantagés. Un second processus induisant une sélection équilibrante est la sélection fréquence-dépendante, c'est-à-dire l'avantage à l'allèle rare. Par exemple, le phénotype du camouflage est souvent dépendant de la fréquence de celui-ci dans la population. En effet, si la fréquence des individus avec un camouflage spécifique augmente dans la population, alors la probabilité que les prédateurs développent des stratégies de contournement de ce camouflage augmente, diminuant alors la valeur sélective de l'allèle et avantageant une autre stratégie, plus rare, de camouflage (Huneman, 2011).

Pour résumé, excepté, dans le cadre de la sélection équilibrante, la sélection conduit à une perte de diversité génétique *a minima* au niveau des *loci* sous sélection. En effet, lorsqu'une mutation favorable apparaît dans le génome, sa fréquence dans la population augmente par sélection. Les allèles neutres, en déséquilibre de liaison avec la mutation favorable, sont eux aussi sélectionnés par effet d'entraînement. On parle d'effet auto-stop ou de « hitch-hiking » (Smith & Haigh, 1974). Le phénomène inverse peut également être observé lors de la purge d'allèles fortement délétères, on parle alors de sélection négative (Charlesworth et al., 1995). La région sous sélection subit une réduction de sa variance génétique, et laisse une empreinte spécifique dans le génome qu'on appelle balayage sélectif (ou « *selective sweep* ») (Figure 3). Seul le processus de recombinaison homologue peut rompre les associations d'allèles et en créer de nouvelles pour générer de la diversité génétique à l'échelle chromosomique.

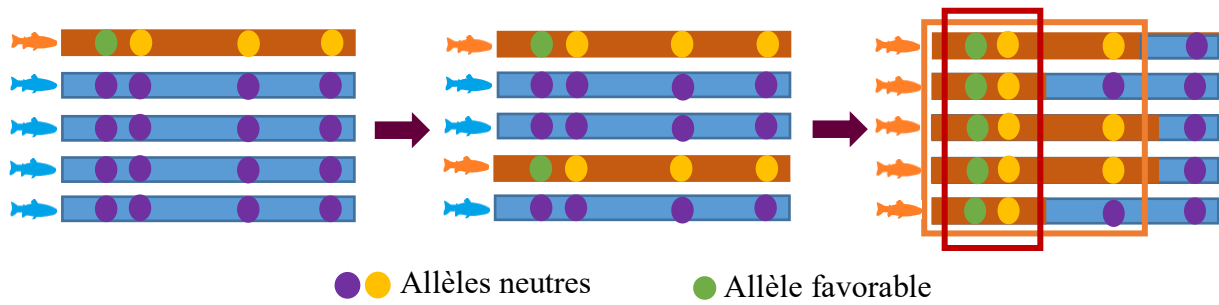


Figure 3. Représentation schématique d'un balayage sélectif après l'apparition d'une mutation favorable dans la population. *La mutation favorable (vert) est en déséquilibre de liaison plus ou moins important avec différentes mutations neutres (jaunes). Au cours des générations, la mutation favorable augmente en fréquence jusqu'à fixation. La mutation neutre la plus proche (jaune, en fort déséquilibre de liaison) est alors elle aussi sélectionnée et augmente en fréquence dans la population, voire devient fixée. Deux allèles en fort déséquilibre de liaison auront moins de chances d'être « séparés » par un événement de recombinaison, que des allèles ayant un déséquilibre de liaison plus faible.*

1.2.2.2. Les théories de l'évolution

Afin de comprendre les mécanismes génétiques de l'évolution des populations ou des espèces, la génétique des populations a été initiée dans les années 1920 à 1930 (Fisher, 1919 ; Haldane, 1932 ; Wright, 1931). Ce n'est qu'avec la re-découverte des lois de Mendel (1822–1884) en 1900 par plusieurs biologistes (Hugo de Vries, Karl Correns, Erich von Tshermak), que Fisher (1919) relie les théories darwiniennes sur la sélection naturelle à la notion d'hérédité mendélienne (Mendel, 1865 ; Bateson & Mendel, 2013) et établit la base de la théorie de la synthèse moderne de la biologie évolutive.

Le terme théorie synthétique de l'évolution (ou néodarwinisme), a été introduit par Julian Huxley (1942). Cette théorie a été admise comme paradigme commun à l'ensemble de la biologie et a permis de démontrer que l'évolution des espèces et des populations est induite à la fois par la sélection naturelle et le hasard. Elle fut construite sur la base de deux théories générales acceptées : la théorie adaptationniste et la théorie neutraliste.

1.2.2.2.1. La théorie adaptationniste

Dans cette théorie, l'évolution des espèces et des populations est principalement due aux pressions de sélection, la dérive n'a alors que très peu d'impact sur l'évolution des populations. Par exemple, pour les adaptationnistes, la multitude de convergences évolutives parmi plusieurs espèces ne peut pas être due au hasard et doit probablement impliquer une pression de sélection commune sur certains caractères.

1.2.2.2.2. La théorie neutraliste

Dans cette théorie Kimura (1983) conjecture que l'évolution d'une population est principalement expliquée par sa taille efficace et son taux de mutation.

Sans renoncer au rôle de la sélection naturelle et des processus adaptatifs (mutations non neutres) dans l'évolution des populations, Kimura estime que la principale force évolutive est la dérive génétique car la plupart des mutations ne sont pas adaptatives, mais neutres et ne procurent donc ni avantage ni désavantage : il n'y a pas de pression de sélection.

1.2.2.3. La domestication

La domestication est un processus évolutif correspondant à l'adaptation génétique au fil des générations d'une population sauvage à la manipulation par l'Homme et à l'élevage en milieu captif (Darwin, 1859, 1868 ; Price, 1984). Ce processus est associé à la sélection des individus les plus adaptés aux besoins de l'Homme (alimentation, caractéristiques physiques). Cela peut conduire à une baisse importante de la diversité génétique au sein des espèces et populations due au goulot d'étranglement intrinsèque à la sélection d'un nombre limité d'individus sauvages pouvant survivre à la captivité. Néanmoins, la domestication permet de réduire la pression de sélection naturelle sur les populations, les animaux passant d'un milieu naturel sauvage à un milieu anthropisé, plus protecteur. Ainsi, lors de la domestication, l'Homme a éloigné les prédateurs, fourni de la nourriture, mais a privilégié les animaux les plus dociles et sociables (Price, 1999). C'est pourquoi, la domestication peut parfois permettre le développement dans la population de caractères qui seraient « mal adaptés » dans des conditions sauvages.

La domestication a entraîné des changements biologiques fondamentaux chez les animaux domestiques par rapport à leurs homologues sauvages (Darwin, 1859 ; Zeder, 2012 ; Larson & Fuller, 2014 ; Wright, 2015). Dans des populations domestiquées de poissons, des

modifications du comportement, de la reproduction, de la croissance et de l'immunité des individus ont pu être observées (Milla et al., 2021). Au niveau comportemental, la domestication réduit la réactivité des poissons pour échapper à des prédateurs (Pasquet, 2018).

Pour illustrer ces différences phénotypiques, une comparaison de la réponse immunitaire entre des populations sauvages et domestiquées a été réalisée dans plusieurs espèces de poissons. Chez le saumon d'Atlantique, aucune différence dans la réponse à *Aeromonas salmonicida* n'a été observée entre les souches sauvages et domestiques (Glover et al., 2009). Cependant, chez la perche, une réponse immunitaire plus élevée a été observée chez les poissons domestiqués par rapport à la population sauvage après diverses infections (Doux fils et al., 2011 ; Mandiki et al., 2011). De la même manière, Palińska-Żarska et al. (2021) montrent que les larves de perches eurasiennes issues de géniteurs domestiqués ont une plus forte expression des gènes liés à la fonction immunitaire, par rapport à des larves issues de géniteurs sauvages. Ainsi, il semble qu'à cause des plus fortes densités de poissons en élevage, les individus domestiqués soient indirectement plus sélectionnés pour la résistance aux virus, bactéries ou pathogènes que les individus des populations sauvages.

La domestication des animaux a commencé il y a plus de 15 000 ans avec la domestication du loup (Galibert et al., 2011). Cette dernière a précédé celle de plusieurs espèces animales terrestres (chèvre, mouton, porc, chat il y a environ 11 000 ans ; etc.). La domestication des espèces aquatiques est bien plus tardive et a commencé par celle de la carpe il y a environ 2 000 ans. Cependant, la date et le lieu précis de la domestication de la carpe fait encore l'objet de débats : Chine néolithique ou Europe centrale à l'époque romaine (Balon, 1995 ; Balon, 2004). La plupart des espèces de poissons d'élevage ont été domestiquées très récemment au cours du siècle dernier (Teletchea & Fontaine, 2014). Pour rappel, la truite arc-en-ciel n'a été domestiquée que dans les années 1870 en Californie (cf. partie 1.1.2 ; Hershberger, 1992). En 1950, seulement 42 espèces de poissons étaient considérées comme domestiquées (FAO, 2003). Aujourd'hui, 313 espèces de poissons sont élevées dans le monde (Metian et al., 2020 ; FAO, 2022).

Les processus de domestication et de sélection artificielle peuvent induire des pertes importantes de diversité génétique au sein des populations ou des espèces domestiquées. Dans la partie suivante, je vous présente donc les conséquences d'une augmentation de la consanguinité et les mécanismes sous-jacents à la dépression de consanguinité.

1.2.3. Conséquences de l'augmentation de la consanguinité

L'augmentation de la consanguinité dans une population conduit à une perte d'hétérozygotie le long du génome. Cela augmente donc le risque de trouver des individus homozygotes pour des allèles délétères récessifs. L'expression de ces allèles délétères (fardeau génétique) est alors accrue et la *fitness* des individus consanguins est réduite, c'est ce qu'on appelle la dépression de consanguinité (Aulstad & Kittelsen, 1971 ; Charlesworth & Charlesworth, 1999). Si les *loci* à l'origine de la dépression de consanguinité interagissent de manière indépendante et additive, la valeur sélective va décliner de façon linéaire pour des consanguinités croissantes (Figure 4). La pente (négative) de cette droite quantifie l'importance de la dépression de consanguinité pour une unité d'accroissement de la consanguinité.

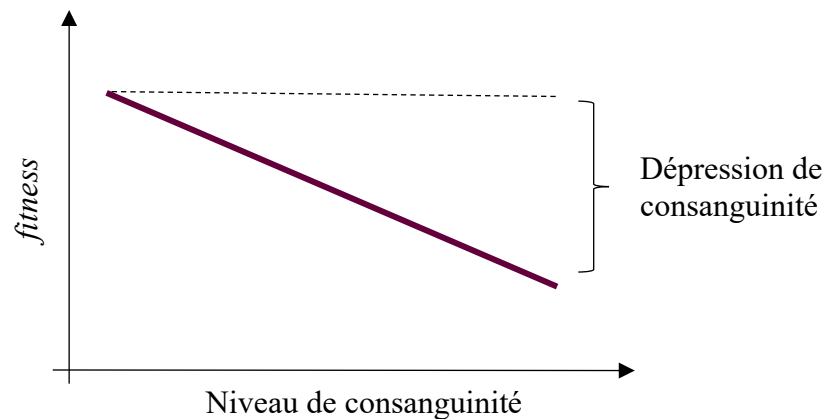


Figure 4. Illustration de l'effet de la dépression de consanguinité.

Plusieurs mécanismes génétiques ont été avancés (e.g. Charlesworth & Charlesworth, 1999) pour expliquer le phénomène de dépression de consanguinité (Figure 5) :

- 1- L'hypothèse de dominance partielle (expression des allèles récessifs délétères ; Davenport, 1908).
- 2- L'hypothèse de superdominance (ou *overdominance*) (supériorité de l'hétérozygote sur les deux types d'homozygotes ; East, 1908).
- 3- L'hypothèse de pseudo-overdominance (combinaisons des effets de plusieurs *loci* liés à effets récessifs délétères ; Jain & Allard, 1966).

Chez des espèces végétales allogames comme le maïs, le phénomène de superdominance (phénomène d'hétérosis ou de vigueur hybride) joue un rôle majeur dans la dépression de consanguinité (Birchler et al., 2003). Toutefois, chez le riz, espèce autogame, le phénomène

pourrait être majoritairement expliqué par l’hypothèse de pseudo-overdominance, mais cela reste difficile à valider par rapport à l’hypothèse de superdominance (Luo et al., 2001). Hormis dans de rares cas, le rôle de la superdominance reste controversé. Les manifestations d’hétérosis seraient plus dues à des effets de dominance (addition des effets de gènes favorables) que de superdominance. Ainsi, chez les espèces animales, la dépression de consanguinité est principalement expliquée par le mécanisme de dominance partielle, caractérisé par l’accumulation d’allèles récessifs délétères (Caballero & Keightley, 1998 ; Fu & Ritland, 1994 ; Charlesworth & Willis, 2009 ; Ceballos et al., 2018).

Dans certains cas rares, ce sont les deux génotypes homozygotes qui sont les plus avantageux par rapport au génotype hétérozygote, c’est ce qu’on appelle la sous-dominance (Charlesworth & Willis, 2009).

Toutes les hypothèses sous-jacentes aux effets de la consanguinité, qu’ils soient positifs (sous-dominance) ou négatifs (dominance partielle, superdominance ou pseudo-overdominance) impliquent des effets génétiques de dominance. Ainsi, Misztal et al. (1997) observent chez différents caractères évalués chez le bovin, que plus la variance de dominance était élevée, plus la dépression de consanguinité était forte.

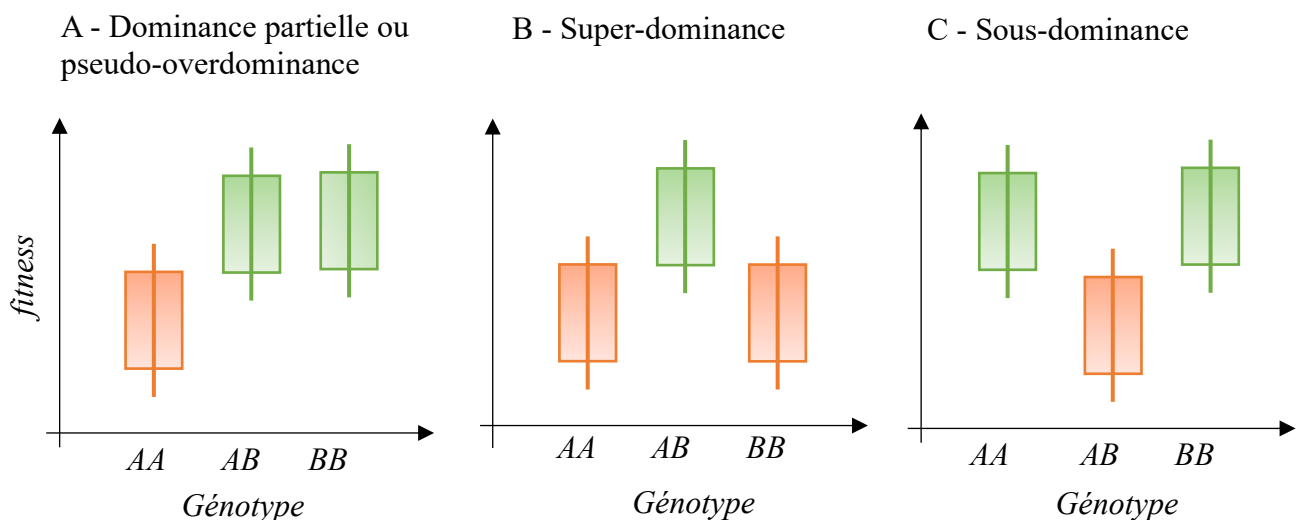


Figure 5. Illustration des différents mécanismes impliqués dans les effets de la consanguinité. A - La dominance partielle ou la pseudo-overdominance, effet d’un allèle récessif délétère qui s’exprime à l’état homozygote (orange). B - Superdominance, avantage de l’hétérozygote (vert) par rapport aux deux homozygotes (orange). C - Sous-dominance, avantage des deux homozygotes (vert) par rapport à l’hétérozygote (orange).

A Retenir – 1.2. Diversité génétique et évolution des populations

- Le maintien de la diversité génétique au sein des populations animales est essentiel, non seulement pour limiter les phénomènes de dépression de consanguinité sur les performances de survie, reproduction, production et santé des animaux, mais aussi pour garantir un potentiel adaptatif et une amélioration génétique des populations domestiques à plus long terme.
- Les quatre forces évolutives (mutation, migration, dérive génétique et sélection) participent à l'évolution des populations ou des espèces, induisant des modifications génétiques et physiologiques.
- La domestication a eu de profondes conséquences sur la physiologie, le comportement et la diversité génétique des espèces élevées.
- Il est essentiel de mieux comprendre les mécanismes génétiques impliqués dans les processus de domestication et de sélection des populations ainsi que dans les phénomènes de dépression de consanguinité.
- Identifier *in fine* les régions génomiques impliquées dans ces processus doit permettre de mieux gérer la diversité génétique dans les populations sélectionnées.

1.3. Méthode d'évaluation de la diversité génétique et d'identification de signatures de sélection

1.3.1. Evaluation de la diversité génétique

Dans cette partie, je vais décrire les diverses méthodes permettant d'évaluer la diversité génétique au sein des populations animales. Par la suite, je détaillerai plus précisément la méthode basée sur les ROH (cf. partie 1.3.1.2.3.), car il s'agit de celle que j'ai utilisée dans ma thèse pour étudier la diversité génétique chez la TAC.

1.3.1.1. Indicateurs de diversité génétique

1.3.1.1.1. Hétérozygotie observée (H_o) et hétérozygotie attendue (H_e)

D'après la définition de la consanguinité, les individus consanguins présentent un niveau d'hétérozygotie plus bas que les individus non consanguins. Ainsi, une mesure classique en génétique des populations est la comparaison entre l'estimation de l'hétérozygotie observée (H_o) et attendue (H_e) c'est-à-dire attendue sous l'hypothèse d'HW.

L'hétérozygotie attendue (H_e), correspond à l'hétérozygotie moyenne attendue à un ou plusieurs *loci* sous l'hypothèse d'HW. On l'estime sur la base de l'indice de diversité de Nei (1973), qui est la probabilité, dans une population, de tirer pour un locus deux allèles différents :

$$h_e = 1 - \sum_{n=1}^N f_n^2$$

Avec f_n la fréquence allélique de l'allèle n .

Le taux d'individus hétérozygotes d'une population (H_e) sera la moyenne de toutes les valeurs h_e , L étant le nombre de *loci* considérés :

$$H_e = \frac{\sum h_e}{L}$$

L'hétérozygotie observée (H_o) est simplement calculée à partir des fréquences des hétérozygotes à chaque locus observé dans la population. C'est donc la moyenne sur les

L loci étudiés du nombre d'individus hétérozygotes ($n_{ind_{HET}}$) divisé par le nombre total d'individus (N) :

$$H_o = \frac{1}{L} \sum_{i=1}^L \left[\frac{n_{ind_{HET}}}{N} \right]$$

Les niveaux d'hétérozygotie observés dans les populations de TAC françaises (D'Ambrosio et al., 2019) sont similaires à ceux observés dans diverses espèces aquacoles (Villanueva et al., 2022), ainsi que des populations bovines (Signer-Hasler et al., 2017) ou porcines (Grossi et al., 2017).

Cette méthode est précise, mais nécessite un bon échantillonnage de la population, afin d'estimer des fréquences alléliques proche de la population réelle, au risque de fausser l'évaluation de la diversité génétique. Cette méthode est également très dépendante du choix des marqueurs.

1.3.1.1.2. Indices de fixation de Wright

Les indices de fixation de Wright (1965) (F_{is} , F_{st} et F_{it}) permettent d'évaluer la diversité génétique dans une population subdivisée en sous-populations et également d'estimer la différenciation génétique entre populations.

Le F_{is} mesure l'écart à la panmixie, c'est-à-dire la déviance de l'hétérozygotie observée par rapport à l'hétérozygotie attendue en cas de reproduction aléatoire. Cet écart à la panmixie est dû à l'augmentation de la consanguinité au sein de la population (ou d'une sous-population) (Nei & Chesser, 1983).

$$F_{is} = \frac{H_e - H_o}{H_e} = 1 - \frac{H_o}{H_e}$$

Les valeurs du F_{is} sont comprises entre -1 et +1. Un F_{is} négatif est associé à un excès d'hétérozygotes et un F_{is} positif à un déficit en hétérozygotes.

Le F_{st} mesure la différenciation génétique entre populations et sous-populations (Weir & Cockerham, 1984 ; Nei & Chesser, 1983).

$$F_{st} = 1 - \frac{H_o}{H_T}$$

Avec H_T , l'hétérozygotie attendue si tous les individus de toutes les populations s'accoupleraient au hasard. Lorsque le F_{st} est proche de 0, cela signifie que les deux populations comparées sont génétiquement proches. A l'inverse lorsque le F_{st} est proche de 1, cela traduit une forte différenciation génétique entre les populations (suggérant peu ou aucun flux de gènes entre populations) (cf partie 1.3.3.2.).

Le F_{it} est un indicateur de la différenciation des individus par rapport à la population totale et permet de décrire l'écart à l'hétérozygotie entre les individus de plusieurs sous-populations d'une espèce. Le F_{it} permet donc d'évaluer la diversité génétique conservée dans les populations au niveau de l'espèce. On le calcule de la manière suivante (Nei & Chesser, 1983) :

$$F_{it} = 1 - \frac{H_e}{H_T}$$

1.3.1.1.3. La taille efficace (N_e)

La taille ou l'effectif efficace (N_e) est la taille d'une population idéale (sous équilibre d'HW), panmictique qui présenterait le même niveau de diversité ainsi qu'une évolution des fréquences alléliques identique que dans la population observée, et cela sous l'influence de la dérive seule (Wright, 1931). On peut estimer l'effectif efficace d'une population à l'aide : i) du nombre de reproducteurs (mâles et femelles) ii) de données génomiques comme le déséquilibre de liaison (DL) ou bien iii) de coefficients de consanguinité (évalués sur pedigree ou apparemment génomique).

i) A partir du nombre de reproducteurs (Fisher, 1930 ; Wright, 1931) :

$$N_e = \frac{4N_f N_m}{N_f + N_m}$$

Avec N_f le nombre de femelles reproductrices et N_m le nombre de mâles reproducteurs dans la population. Ici on suppose que chacun des reproducteurs de chaque sexe a une contribution

équilibrée. Dans le cas de déséquilibre de contribution connu, d'autres formules prennent en compte la possibilité d'avoir un nombre variable de gamètes apportées par chacun des individus pour constituer la génération suivante.

- ii) A partir du déséquilibre de liaison (DL) entre marqueurs le long du génome (Sved, 1971) :

$$E(r^2) = \frac{1}{1 + 4N_e c}$$

Avec c , le taux de recombinaison et S la taille de l'échantillon. Le DL est mesuré par la corrélation au carré (r^2) entre les allèles positionnés aux différents *loci* (Hill, 1981). On en déduit ensuite la taille efficace de la population :

$$N_e \approx \frac{1 + \frac{1}{S} - E(r^2)}{4c \left[E(r^2) - \frac{1}{S} \right]}$$

- iii) A l'aide de l'évolution du taux de consanguinité (Falconer & Mackay, 1996)

$$N_e = \frac{1}{2\Delta F}$$

Avec, ΔF le taux d'accroissement de la consanguinité d'une génération à l'autre :

$$\Delta F = \frac{F(\text{génération } n) - F(\text{génération } n - 1)}{(1 - F(\text{génération } n - 1))}$$

Les coefficients de consanguinité peuvent être estimés sur la base de diverses données (pedigree et génomique ; cf. partie 1.3.1.2.).

La FAO conseille de conserver un N_e minimal de 50 afin de limiter la dépression de consanguinité au sein des populations animales (FAO, 2013). Une baisse importante du N_e a été observée au sein des populations domestiquées de TAC françaises. Sur la base du DL, le N_e a été estimé entre 24 et 48 dans les générations récentes de truite arc-en-ciel (D'Ambrosio et al., 2019). Ces valeurs de N_e sont globalement cohérentes avec ceux d'autres espèces aquacoles où des N_e inférieures à 50 ont été estimées dans des populations ayant subies plusieurs générations de sélection (ormeau – Chen et al., 2017, poisson-chat – Garcia et al., 2018, saumon – Yáñez et al., 2014 ; Barria et al., 2018 ; Martinez et al., 2022 ; daurade - Villanueva

et al., 2022 ; bar - Šegvić-Bubić et al., 2016 ; Villanueva et al., 2022). Les tailles efficaces estimées dans les populations aquacoles domestiquées sont nettement inférieures à celles observées dans les populations bovines et porcines (Danchin-Burge, 2020 ; Ganteil et al., 2020), qui sont majoritairement comprises entre 70 et 90.

1.3.1.1.4. Les coefficients de consanguinité et de parenté

Les coefficients de consanguinité et de parenté des individus d'une population permettent d'en apprécier la diversité génétique. En effet, la probabilité entre deux individus qu'un allèle tiré au hasard chez l'un soit identique par descendance à un allèle au même locus tiré au hasard chez l'autre, est plus élevée lorsque l'apparentement augmente entre les individus, car cela est associé à une baisse de la diversité allélique au sein de la population.

Si l'on souhaite estimer les coefficients de consanguinité (ou de parenté), deux catégories d'informations sont utilisables : le pedigree et les informations génomiques (cf. partie 1.3.1.2.).

1.3.1.2. Calcul des coefficients de consanguinité

1.3.1.2.1. Sur la base du pedigree

Deux méthodes probabilistes permettent d'estimer les coefficients de consanguinité des individus d'une population à partir de l'information généalogique : la méthode des chemins (Wright, 1922) et la méthode tabulaire (Cruden, 1949). Le coefficient de consanguinité est la probabilité pour un individu de porter deux copies identiques d'un même allèle (hérité d'un ancêtre commun).

Le coefficient de consanguinité (F_k) d'un individu k estimé sur la base du pedigree, correspond également au coefficient de parenté (Φ_{ij}) de ses parents (i et j). A l'aide d'information généalogiques, le calcul des coefficients de consanguinité et de parenté par la méthode des chemins se fait de la manière suivante :

$$\Phi_{ij} = F_k = \sum_{\text{ancêtre}} \left(\frac{1}{2}\right)^n (1 + F_a)$$

Avec, pour chaque ancêtre commun identifié entre i et j , le nombre total d'individus n dans le pedigree jusqu'à l'ancêtre commun, et F_a le coefficient de consanguinité de chaque ancêtre commun.

Le calcul des coefficients de consanguinité et de parenté, par la méthode tabulaire se fait de la manière suivante :

$$\Phi_{ij} = F_k = \frac{1}{4} \Phi_{pp'} + \frac{1}{4} \Phi_{pm'} + \frac{1}{4} \Phi_{mp'} + \frac{1}{4} \Phi_{mm'}$$

Avec m et p les parents de i et m' et p' les parents de j ($\Phi_{pp'}$ est donc le coefficient de parenté entre p (père de i) et p' (père de j)).

L'utilisation du pedigree pour mesurer les coefficients de consanguinité est basée sur la théorie de l'identité par descendance (IBD) et considère que la consanguinité est répartie de manière homogène le long du génome. Cependant, on atteint vite les limites d'une estimation basée sur le pedigree étant donné que l'on mesure seulement la consanguinité théorique et attendue des individus et cela ne permet pas de différencier deux pleins-frères (car cette méthode ne prend pas en compte le déséquilibre de liaison et l'aléa de méiose). De plus, la fiabilité de la mesure des coefficients de consanguinité dépend de la fiabilité et de la profondeur des données généalogiques. Un pedigree incomplet ou faussé entraîne une sous-estimation des coefficients de consanguinité (Engelsma et al., 2011).

1.3.1.2.2. Sur la base de données génomiques

Plusieurs méthodes basées sur des informations génomiques ont été développées pour estimer les coefficients de consanguinité. La première est basée sur la méthode des moments et permet d'estimer le coefficient de consanguinité de chacun des individus (Ritland, 1996).

$$F_k = \frac{1}{L} \sum_{i=1}^L \left(\frac{s_{A,i} - p_i^2}{p_i} - \frac{s_{B,i} - q_i^2}{q_i} \right)$$

Avec, F_k le coefficient de consanguinité de l'individu k , basé sur L loci. Si l'individu est homozygote pour l'allèle A ou B du locus i , alors $s_{A,i}$ ou $s_{B,i}$ sont respectivement égaux à 1. Si le locus est hétérozygote alors $s_{A,i}$ et $s_{B,i}$ sont égaux à 0. La fréquence allélique de l'allèle A (allèle majoritaire) est p_i , et la fréquence de l'allèle B est q_i .

Sur la base de la relation entre les individus, des matrices d'apparentements génomiques G peuvent être calculées (GRM, VanRaden, 2007) :

$$G = \frac{ZZ'}{\sum_{i=1}^L 2p_i q_i}$$

Avec L le nombre de locus considérés, p_i la fréquence dans la population de l'allèle majoritaire et $q_i (=1-p_i)$ la fréquence de l'allèle minoritaire. La matrice Z' est la transposée de la matrice Z , qui est construite de la manière suivante : $Z = M - P$. M est une matrice de dimension N (= nombre d'individus) par L , spécifiant les allèles présents chez les individus génotypés. La matrice P est de dimension N par L contenant les fréquences alléliques dans la population de ces mêmes allèles ($P[n,l]=2p_i \times 1_N$).

Plus l'apparentement entre les individus est élevé, plus le risque d'avoir de hauts niveaux de consanguinité dans la descendance l'est également. Cette méthode est basée sur les fréquences alléliques, ainsi le coefficient d'apparentement génomique estimé sera plus élevé si les individus partagent un allèle rare qu'un allèle commun. C'est pourquoi, cette méthode nécessite un bon échantillonnage de la population au risque de fausser l'estimation des coefficients de consanguinité.

Une dernière méthode basée sur les îlots de consanguinité (ou ROH) tend à devenir la méthode de référence dans l'estimation des coefficients de consanguinité. Etant donné que cet indicateur a été celui utilisé dans cette thèse, je vais maintenant détailler les raisons qui font des ROHs un indicateur de prédilection de la consanguinité ainsi que les calculs permettant de mesurer les coefficients de consanguinité à partir des ROHs.

1.3.1.2.3. Les îlots de consanguinité (ROH)

Les ROHs sont définis comme de longs segments homozygotes hérités d'un ancêtre commun et donc identiques par descendance (Figure 6 ; McQuillan et al., 2008).

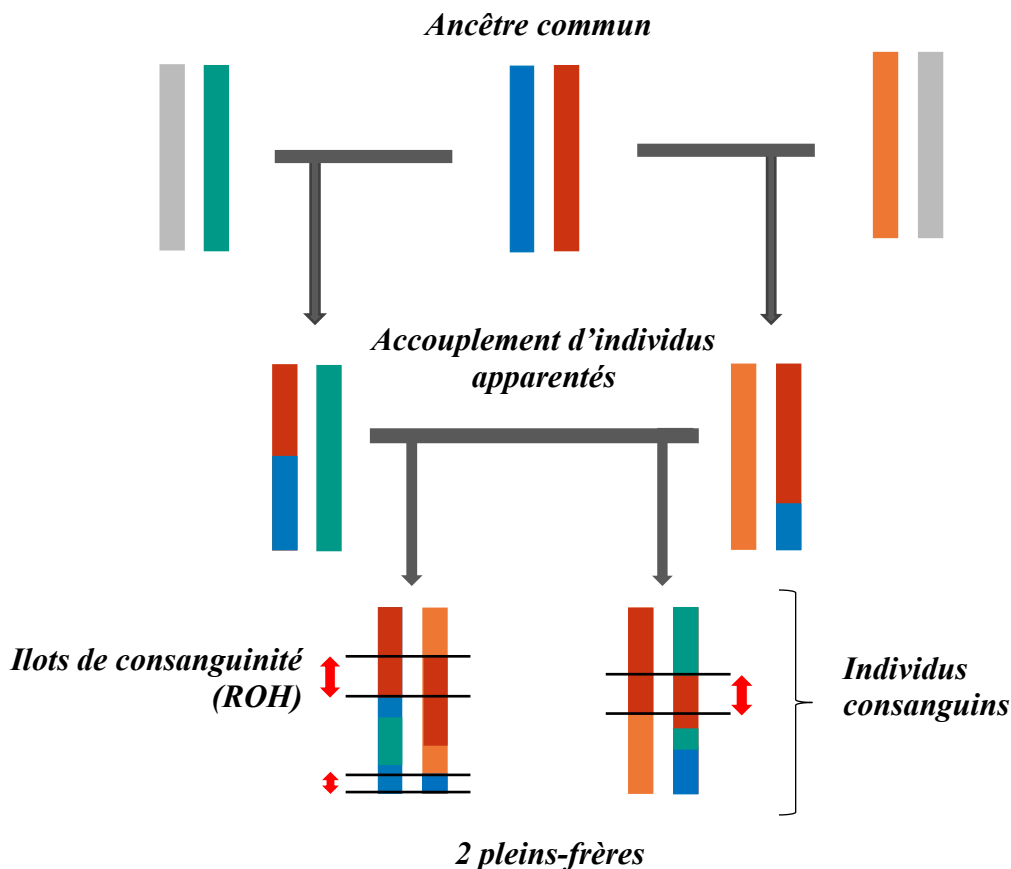


Figure 6. Illustration des îlots de consanguinité (ROH). Si l'on considère un ancêtre commun avec deux chromosomes (1 bleu et 1 rouge), celui-ci transmettra l'un de ses chromosomes à ses descendants (+ recombinaison lors de la méiose). Lors d'un accouplement d'individus apparentés, on pourra retrouver des fragments des chromosomes de l'ancêtre commun à l'état homozygote chez l'individu consanguin, c'est ce qu'on appelle un ROH.

Une fois les ROHs identifiés pour chacun des individus, un coefficient de consanguinité peut être estimé pour chacun d'eux de la manière suivante :

$$F_{ROH,i} = \frac{\sum L_{ROH,i}}{L_{auto}}$$

Avec, $\sum L_{ROH,i}$, la somme totale des ROHs présents dans le génome de l'individu i et L_{auto} , la taille totale du génome où il est possible de détecter un ROH.

Il a été démontré que les niveaux de consanguinité des individus sont variables le long du génome des animaux (Curik et al., 2017 ; D'Ambrosio et al., 2019). Ainsi, des individus avec le même niveau de consanguinité globale peuvent avoir une répartition de leur

consanguinité différente et n'expriment donc probablement pas les mêmes effets de dépression de consanguinité. Il est donc essentiel de pouvoir localiser la consanguinité, ce que permettent les ROHs (Keller et al., 2011). Il est alors envisageable d'estimer le coefficient de consanguinité d'une région spécifique du génome en considérant les ROHs contenus seulement dans celle-ci (Figure 7).

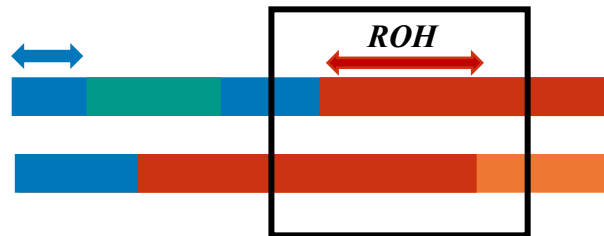


Figure 7. Illustration du calcul d'un coefficient de consanguinité local (fenêtre noire). Ici on ne considère que les ROHs présents dans la fenêtre étudiée, donc le ROH rouge et pas le bleu (d'après Doublet, 2020).

Le calcul du coefficient de consanguinité se fait alors sur la base de la fenêtre considérée et non pas sur la taille totale du génome.

$$F_{ROH,region\ r,i} = \frac{\sum L_{ROH,region\ r,i}}{L_{REGION}}$$

Avec, $\sum L_{ROH,region\ r,i}$, la somme totale des ROHs de l'individu i présent dans la région r , L_{REGION} , est la taille de la région étudiée.

Grâce à la taille des ROHs, il est également possible d'estimer l'âge des événements ayant engendré de la consanguinité. Plus un fragment homozygote est long, moins il a été cassé par recombinaisons successives et donc plus on estime son apparition récente. A l'opposé, les courts fragments ROHs correspondent à des événements anciens ayant engendré la consanguinité (Pryce et al., 2014 ; Doekes et al., 2019).

Plusieurs logiciels et méthodes associées permettent d'identifier les ROHs. Tout d'abord, le logiciel PLINK (Purcell et al., 2007 ; Chang et al., 2015) permet de facilement les détecter à l'aide d'une fenêtre coulissant le long du génome. Cependant, ce logiciel nécessite un paramétrage pertinent pour détecter les ROHs, paramétrage qui peut faire varier la taille et la position des ROHs. Ce paramétrage est également à ajuster en fonction des espèces et de la densité des génotypages (Purfield et al. 2012 ; Chang et al., 2015). Les paramètres initialement

demandés dans le logiciel sont : le nombre minimal de SNPs homozygotes dans un ROH, le nombre maximal de marqueurs hétérozygotes autorisés (en lien avec les possibles erreurs de génotypage), la taille minimale d'un ROH, la distance maximale entre deux marqueurs consécutifs et la densité de marqueurs dans un ROH. Une fois les paramètres correctement ajustés, cet outil donne de bons résultats dans l'estimation des coefficients de consanguinité (Forutan et al., 2018). Sur le même principe, une détection des ROHs peut être réalisée à l'aide du package R `detectRUNS` ; ce package permet également d'identifier des régions constituées seulement de marqueurs hétérozygotes (appelés ROHet).

Une autre méthode, probabiliste, a été proposée pour identifier les ROHs, basée sur un modèle de Chaines de Markov Cachées avec deux états possibles à chaque position : présence ou absence de ROH (Leutenegger et al., 2003 ; Druet & Gautier, 2017). Druet & Gautier (2017) ont également considéré l'âge d'apparition de la consanguinité, en classant les ROHs selon leur taille. Cette méthode peut être mise en place grâce au package R, `RZooRoH`, et présente de bons résultats dans l'estimation des coefficients de consanguinité. Une forte corrélation a été observée entre les coefficients de consanguinité calculés grâce aux ROHs détectés par le logiciel `PLINK` et ceux calculés à partir du package `RZooRoH` (Meyermans et al., 2020). La méthode basée sur le package `RZooRoH` est coûteuse en mémoire et en temps de calcul, par rapport à une détection de ROHs réalisée à l'aide du logiciel `PLINK` ou du package R `detectRUNS`.

Depuis la fin des années 2010, grâce à l'essor des outils génomiques, de nombreux généticiens ont cherché à estimer les coefficients de consanguinité réels (sur la base d'informations moléculaires) des individus de populations animales domestiquées. Chez la TAC, sur la base des ROHs, des coefficients de consanguinité entre 12,2 et 19,5% ont été estimés dans plusieurs populations françaises de truites domestiquées/sélectionnées (D'Ambrosio et al., 2019). Les populations commerciales de truites ont les niveaux moyens de consanguinité les plus élevés (F moyen compris entre 16,6% et 19,5%), alors qu'une population non-sélectionnée de truites arc-en-ciel a un F moyen d'environ 10%. Les estimations des coefficients de consanguinité estimés dans d'autres populations de truite arc-en-ciel sont basés uniquement sur le pedigree (Wang et al., 1999 ; 2002). Pante et al. (2001), à partir du pedigree, montra qu'après 6 générations de sélection, les niveaux de consanguinité des populations de truite arc-en-ciel se trouvaient entre 3 et 10%. Dans d'autres espèces piscicoles les coefficients

de consanguinité moyens sont, dans certains cas, comparables (saumon et turbot) aux F moyens des populations de truites françaises (Tableau 2), mais sont parfois bien plus bas (tilapia). Les premières études traitant des coefficients de consanguinité génomiques des espèces aquacoles sont très récentes par rapport aux études concernant les espèces animales terrestres.

Au sein des populations d’animaux terrestres sélectionnées, des niveaux de consanguinité, évalués par ROH soit plus faibles, similaires ou plus élevés à ceux des populations de truites françaises, ont pu être observés dans différentes espèces, races ou lignées (Tableau 2). Les niveaux de consanguinité au sein des populations domestiquées sont très variables et peuvent atteindre des niveaux élevés, et cela particulièrement dans les lignées sélectionnées (exemple : US Jersey ; lignées de porc italien ; poules pondeuses ; dindes) ou à faibles effectifs (exemple : races locales de porc, chèvre et poules). Il est donc essentiel de caractériser et comprendre les effets de la consanguinité pour la gérer au mieux dans les programmes de sélection.

Tableau 2. Récapitulatif des niveaux moyens de consanguinité moléculaire observés dans plusieurs espèces animales aquatiques terrestres (estimations toutes établies à l’aide des îlots de consanguinité, ROH). *N* correspond au nombre de populations considérées.

Espèce	Race/type de populations	Niveaux moyens de coefficients de consanguinité (F_{ROH})	Références
Truite arc-en-ciel	Lignées sélectionnées et/ou commerciales (N = 4)	de 12,2 à 19,5%	D’Ambrosio et al., 2019
	Lignée non sélectionnée (N = 1)	10%	
Saumon	Lignées commerciales (N = 2; A et B, lignées sélectionnées chilienne)	de 14,2 à 15,2%	Yoshida et al., 2020
	Lignée issue du croisement A x B	0,40%	
Tilapia	Lignée sélectionnée (N = 1, Lignée Sukamandi avec 23 générations de sélection)	en moyenne 2,7% (de 0,9 à 12% par génération)	Yu et al., 2022
Bovin	US Jersey (N = 4)	de 14,4 à 22,5%	Makanjuola et al., 2020a; Lozada-Soto et al., 2022 ; Srikanth et al., 2022

(Suite Tableau 2)

Espèce	Race/type de populations	Niveaux moyens de coefficients de consanguinité (F_{ROH})	Références
Bovin	Holstein (N = 4)	de 12% à 15,6%	Doublet et al ., 2019 ; Makanjuola et al., 2020a; Makanjuola et al., 2020b ; Lozada-Soto et al., 2022
	Montbéliarde & Normande (N = 2)	~ 10%	Doublet et al ., 2019
	Red Dairy Cattle (N = 3 ; original Finnish Ayrshire breed, Scandinavian and North American red breeds)	de 6 à 7%	Martikainen et al., 2017
	Ayrshire (N = 1)	11%	Lozada-Soto et al., 2022
	Brown Swiss (N = 1)	15%	Lozada-Soto et al., 2022
	Guernsey (N = 1)	15%	Lozada-Soto et al., 2022
Porc	Race Landrace (N = 9)	de 2,9 à 9,4%	Zanella et al., 2016 ; Grossi et al., 2017 ; Yang et al., 2017
	Race Large white (N = 5)	de 4,4 à 10,6%	Zanella et al., 2016 ; Yang et al., 2017
	Race Duroc (N = 5)	de 3 à 9,5%	Grossi et al., 2017 ; Yang et al., 2017
	Race Yorkshire (N = 1)	5%	Grossi et al., 2017
	Race Pietrain (N = 3)	de 4,5 à 7,6%	Yang et al., 2017
	Lignées commerciales (N = 3; Italian Large White, Italian Duroc et Italian Landrace)	de 11,7 à 26,7%	Schiavo et al., 2020, 2021, 2022
	Races autochtones provenant de plusieurs pays européens (N = 24)	de 7,2 à 33,2%	
	Populations domestiquées dans le monde (N = 100)	de 0,2 à 22,8%	Yang et al., 2017
	Population sauvages dans le monde (N = 22)	de 0,2 à 11,3%	
		Race 'Valle del Belice' (N = 1)	8,40%
	Race Barbaresca (N = 1)	8,70%	Mastrangelo et al., 2017b
Mouton	Races du consortium des moutons italiens (N = 20)	de 1,6 à 9,9%	Mastrangelo et al., 2018
	Races locales suisses (N = 8)	de 2,15 à 15,32%	Signer-Hasler et al., 2019
	Races françaises sélectionnées (N = 5)	de 3,5 à 4,4%	Rodríguez-Ramilo et al., 2019
	Races ovines (N = 68 - international sheep genomics consortium)	de 0,9 à 24,4%	Nosrati et al., 2021
	Race 'German White-headed' (N = 1)	de 4,26 à 4,64%	Addo et al., 2021

CHAPITRE 1 – INTRODUCTION GENERALE

(Suite Tableau 2)

Espèce	Race/type de populations	Niveaux moyens de coefficients de consanguinité (F_{ROH})	Références
Chèvre	Races suisses (N = 8)	de 13,7 à 22,6%	Burren et al., 2016
	Races suisses (N = 9)	de 3 à 9%	Signer-Hasler et al., 2022
	Population sauvage - Bezoar	26,40%	
Poule	Lignées commerciales de poules de chair (N = 4)	de 12,2 à 17,6%	Talebi et al., 2020 ; Restoux et al., 2022
	Lignées commerciales de poules pondeuses (N = 4)	de 23,2 à 36,1%	Talebi et al., 2020 ; Restoux et al., 2022
	lignées commerciales (N = 3; Rhode Island Red, White Leghorn, Cornish)	de 22 à 28%	Zhang et al., 2020
	Races françaises locales (N = 19)	de 12,5 à 41,9%	Restoux et al., 2022
	Races françaises locales, non inclus dans un programme de gestion de la diversité génétique (N = 4)	de 19,8 à 27,6%	Restoux et al., 2022
	Race locale européenne (N= 2, Houdan)	de 24,4 à 35%	Zhang et al., 2020 ; Restoux et al., 2022
	Races chinoises présentes dans des programmes de conservation (N = 3 ; Baier Yellow, Beijing You and Langshan)	de 5,4 à 7,4%	Zhang et al., 2018
	race locales chinoises (N = 5; Beijing, Hongshan, Shouguang, Silkie, Tibet)	de 5 à 23%	Zhang et al., 2020
	racas italiennes locales de poulet (N = 8)	de 16,1 à 47,8%	Cendron et al., 2021
	Lignée sauvage (N = 2; red junglefowl)	de 8 à 11,4%	Talebi et al., 2020 ; Zhang et al., 2020
Dinde	Lignées commerciales de dindes (N = 3)	de 26 à 28%	Adams et al., 2021

1.3.2. Estimation des effets de la consanguinité sur les performances

A l'échelle d'un individu, la dépression de consanguinité se mesure en comparant la performance d'un individu consanguin (croisement plein-frère) \overline{W}_I avec celle d'un individu issu d'un croisement aléatoire (Charlesworth & Charlesworth, 1999) :

$$\overline{W}_R : \delta = \frac{\overline{W}_R - \overline{W}_I}{\overline{W}_R}$$

Il est possible d'évaluer à l'échelle d'une population les effets de la dépression de consanguinité à partir des coefficients de consanguinité calculés pour chaque individu.

Pour cela, des modèles linéaires ou non-linéaires permettent de décrire l'effet de la consanguinité sur la valeur sélective ou les performances des individus au sein d'une population. Plusieurs études ont comparé les deux approches (linéaire et non-linéaire) et sont arrivées à la conclusion que pour des valeurs modérées de consanguinité ($F < 0,20$), les modèles linéaires étaient adaptés pour estimer les effets de la dépression de consanguinité (Carrillo & Slewerdt, 2010 ; Malhado et al., 2013). Classiquement, un modèle linéaire BLUP - modèle animal est utilisé avec en covariable le vecteur des coefficients de consanguinité (cf Encart 2). L'estimation des effets de la dépression de consanguinité peut se faire sur la base du niveau global de consanguinité, ou bien à l'aide d'estimations des niveaux de consanguinité locaux (Martikainen et al., 2018 ; Botero-Delgadillo et al., 2020 ; Doekes et al., 2020). Doekes et al. (2020) ont estimé l'effet de la consanguinité sur chacun des SNPs à l'aide d'une GWAS en associant une régression linéaire à chaque SNP. Pour cela, ils ont imputé pour chaque SNP l'information suivante : contenu ou non dans un ROH.

Au vu des mécanismes génétiques pouvant expliquer le phénomène de dépression de consanguinité (cf. partie 1.2.3), il est important de considérer les effets de dominance dans le modèle utilisé pour estimer les effets de la consanguinité (Toro & Varona, 2010 ; Vitezica et al., 2016). Il a aussi été montré que les phénomènes de dépression de consanguinité sont plus souvent observés suite à des événements récents d'apparition de consanguinité (Doekes et al., 2019 ; Mekanjuola et al., 2020b), où les effets défavorables de la consanguinité n'ont pas encore pu être éliminés par la purge des allèles récessifs délétères au sein de la population (Kristensen & Sørensen, 2005).

Encart 2 : Modèles de la génétique quantitative

En génétique quantitative, on considère qu'un caractère quantitatif est expliqué par une multitude de gènes ayant tous de très faibles effets. Ainsi, pour comprendre et expliquer les effets génétiques d'un caractère, il est nécessaire de décrire un modèle génétique sous-jacent aux variations des phénotypes. Le modèle général permettant d'expliquer le phénotype d'un caractère quantitatif (y) se décrit simplement comme la combinaison linéaire et additive d'un effet génétique (g) (ou valeur génétique) et un effet de l'environnement (pe) (e , correspond aux effets résiduels) : $y_{ij} = g_i + pe_i + e_{ij}$

On peut décomposer la valeur génétique en trois composantes majeures : les effets additifs (a) et les effets génétiques non additifs de dominance (d) et d'épistasie (i) : $g = a + d + i$

Si l'on considère seulement les effets additifs, définis comme la somme des effets moyens des gènes, la valeur génétique additive d'un descendant sera en espérance : $g_i = E(a_i) = \frac{1}{2}a_m + \frac{1}{2}a_p$, le $\frac{1}{2}$ traduisant le fait qu'un parent transmet en espérance la moitié de sa valeur génétique à ses descendants.

On définit la dominance comme la somme des effets d'interaction entre allèles de chaque gène impliqué dans l'expression du caractère et l'épistasie comme la somme des effets d'interaction entre allèles de *loci* différents. Il faut garder en mémoire que ces effets sont pour l'essentiel recréés à chaque génération car ils dépendent d'association particulière d'allèles (= recrée lors de l'accouplement).

Dans le but de simplifier les modèles génétiques, classiquement seul les effets additifs sont pris en compte. Toutefois, aujourd'hui des outils performants permettent de prendre en compte les effets non-additifs sans trop de difficultés. Notamment grâce au programme '*parallele90*' développé par Vitezica et al. (2016) qui permet de construire la matrice de dominance simplement à partir des génotypes utilisés dans le modèle. L'idéal est donc de prendre en compte *a minima*, les effets additifs ainsi que les effets de dominance pour décrire les performances.

L'héritabilité (h^2) permet d'estimer ce que chaque parent transmet à sa descendance. C'est donc la part de la variance phénotypique (σ_y^2) expliquée par la variance génétique (σ_g^2) (Lush, 1937) : $h^2 = \frac{\sigma_g^2}{\sigma_y^2}$

De plus, si l'on considère que la variance génétique est principalement due à la variance génétique additive (on néglige la variance non-additive), on pourra écrire : $h^2 = \frac{\sigma_a^2}{\sigma_y^2}$ que l'on appelle hérabilité au sens strict.

De la même manière, on peut estimer la part des effets de dominance expliquant la variance phénotypique par le ratio : $d^2 = \frac{\sigma_d^2}{\sigma_y^2}$ Un d^2 fort est un indice concernant l'importance potentielle des effets de la consanguinité.

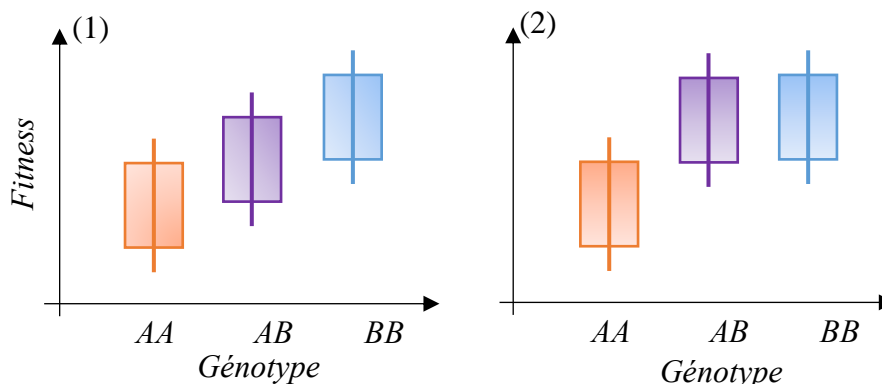


Figure 7. Illustration d'effets additifs (1) ou de dominance (2) pour locus bi-allélique (allèles A et B). Observation de la variation phénotypique en fonction du génotype.

Chez la TAC, des effets significatifs de dépression de consanguinité ont été observés sur le poids des femelles avec une réduction entre 1,7 et 5% du poids de la femelle pour une augmentation de 0,10 point du coefficient de consanguinité (Gjerde et al., 1983 ; Kincaid, 1983 ; Su et al., 1996 ; Pante et al., 2001). Des effets de dépression de consanguinité ont également été estimés sur des caractères de reproduction de la truite arc-en-ciel (Kincaid, 1983 ; Su et al., 1996), ainsi que sur la survie des descendants (Kincaid, 1983). Néanmoins, toutes ces estimations d'effets de dépression de consanguinité chez la truite arc-en-ciel sont soit basées sur la comparaison des performances d'un groupe d'individus consanguins (croisement plein-

frères) avec celles d'un groupe d'individus issus d'un croisement aléatoire, soit sur un modèle de régression linéaire basé sur des coefficients de consanguinité théoriques calculés à partir des pedigrees. Or, de nombreuses études ont montré que les estimations des effets de la dépression de consanguinité sur la base des informations généalogiques étaient moins précises que celles réalisées à partir de données génomiques (Forstmeier et al., 2012 ; Pryce et al., 2014 ; Kardos et al., 2016 ; Howard et al., 2017). C'est pourquoi réaliser, sur la base d'une information moléculaire, de nouvelles estimations des effets de la consanguinité chez la truite arc-en-ciel permettraient de préciser et d'améliorer les connaissances des effets de la consanguinité chez la truite arc-en-ciel.

1.3.3. Méthodes et outils d'identification de signatures de sélection

Le concept fondateur de la recherche de traces de sélection est l'idée formalisée par Lewontin & Krakauer (1973) que certaines différences entre populations sont le fruit d'une sélection positive susceptible d'avoir laissé des patrons de variation des fréquences alléliques localement distincts dans le génome de populations divergentes. La caractéristique clef de la sélection positive est qu'elle provoque une augmentation inhabituellement rapide de la fréquence des allèles autour du variant favorable, se produisant sur un temps suffisamment court pour que les recombinaisons ne cassent pas substantiellement l'haplotype dans lequel la mutation sélectionnée s'est produite. Une signature de la sélection positive est donc un allèle ayant un DL à longue distance inhabituel compte tenu de sa fréquence dans la population. Différentes méthodes et logiciels associés (Table 3) ont été développés pour révéler des signatures de sélection, et donc identifier des zones localement distinctes par rapport au reste du génome, soit au sein d'une unique population (intra-population), soit par différenciation entre populations (inter-population).

Table 3. Outils disponibles pour les différentes méthodes de détection de signatures de sélection.

Méthode	Détection	Type d'informations	Outils disponibles
D de Tajima	<i>intra-population</i>	Spectre des fréquences alléliques	Arlequin (<i>Excoffier et al., 2005</i>), SweepFinder2 (<i>Nielsen et al., 2005</i>), VariScan (<i>Vilella et al., 2005</i>), VCFTools (<i>Danecek et al., 2011</i>)
CLR	<i>intra-population</i>		SweeD (<i>Pavlidis et al., 2013</i>), SweepFinder2
Statistique H (de Fay et Wu)	<i>intra-population</i>		DnaSP (<i>Rozas et al., 2017</i>), VariScan
ROH	<i>intra-population</i>	Réduction du polymorphisme	DetectRUNS (<i>Biscarini et al., 2018</i>), PLINK (<i>Purcell et al., 2007 ; Chang et al., 2015</i>), RZooRoH (<i>Bertrand et al., 2019</i>), ZANARDi (<i>Marras et al., 2017</i>)
rEHH, iHS XP-EHH	<i>intra-population / inter-populations</i>	Déséquilibre de liaison et haplotypes	Hapbin (<i>MacLean, 2014</i>), Rehh (<i>Gautier et al., 2017</i>), Selscan (<i>Szpiech & Hernandez, 2014</i>)
Fst	<i>inter-populations</i>	Différenciation locale (SNP)	BayeScan (<i>Foll & Gaggiotti, 2008</i>), HierFstat (<i>Goudet, 2005</i>), PLINK, SNP & Variation Suite (<i>Golden Helix, Bozeman, MT, United States</i>), VCFTools, Bayescenv (<i>Villemereuil & Gaggiotti, 2015</i>), OutFLANK (<i>Whitlock & Lotterhos, 2015</i>)
FLK HapFLK	<i>inter-populations</i>	Différenciation locale basée sur les SNP Haplotypes	HapFLK (flk : <i>Bonhomme et al., 2010 ; hapflk : Fariello et al., 2013</i>)

Dans le cadre d'une détection de signatures de sélection intra-population, les régions sous balayage sélectif auront une plus faible diversité génétique que la moyenne avec un excès d'allèles rares, et pourront présenter une concentration d'allèles potentiellement favorables impliqués dans des phénomènes d'adaptation, de sélection ou de domestication (Smith & Haigh, 1974). La recherche de signatures de sélection intra-population peut utiliser trois types d'informations génomiques : le spectre de fréquences alléliques, le déséquilibre de liaison ou la réduction de la diversité génétique (Vitti et al., 2013 ; Weigand & Leese, 2018 ; Saravanan et al., 2020). Pour détecter des signatures de sélection intra-population, il est nécessaire de partir de l'hypothèse d'une évolution neutre du génome (modèle neutre), c'est-à-dire une évolution du génome en l'absence de sélection. On cherche alors à identifier les régions du génome qui s'éloignent significativement de ce modèle nul, que l'on considérera donc sous sélection.

1.3.3.1. Signatures de sélection intra-population.

1.3.3.1.1. Exploitation du spectre de fréquences alléliques

Plusieurs méthodes de détection de signatures de sélection se basent sur les traces que la sélection laisse sur le spectre de fréquences alléliques.

Le D de Tajima (Tajima, 1989) permet d'estimer le polymorphisme par fenêtres génomiques. Le D de Tajima est calculé comme la différence entre deux mesures : le nombre moyen de séquences différentes deux à deux et le nombre total de sites polymorphes trouvés sur l'ensemble des séquences d'une population. Ainsi, un D négatif correspondra à un excès d'allèles rares associé à une réduction de la diversité génétique, et indiquant donc une trace de sélection. Un D positif correspondra à des événements de sélection équilibrante (le nombre moyen de différences entre séquences sera supérieur à la mesure de diversité basée sur le nombre de site polymorphes).

Une deuxième méthode est le H de Fay et Wu (2000), évalué comme la différence entre la mesure de la diversité basée sur le nombre moyen de différences entre séquences, et une mesure de la diversité sensible aux allèles dérivés à haute fréquences (*versus* la fréquence des allèles ancestraux). Par rapport au D de Tajima, cette méthode permet de différencier un événement de sélection positive et négative. Un H négatif correspondra à une région ayant un excès d'allèles dérivés à hautes fréquences, indiquant donc une trace de sélection positive.

Le rapport de vraisemblance composite (CLR) est une troisième approche qui permet d'évaluer l'asymétrie du spectre des fréquences alléliques parmi plusieurs *loci* (fenêtre génomique), tout en incorporant le taux de recombinaison local afin d'éliminer les faux positifs dus aux événements démographiques (Nielsen, 2005).

Malgré l'ancienneté de ces méthodes qui ont fait leurs preuves, les tests basés sur les fréquences alléliques sont grandement affectés par l'histoire démographique des populations. Une variation importante de la taille de la population (goulots d'étranglement, expansion rapide de la population) peut artificiellement faire augmenter la fréquence des allèles rares, induisant une détection de signature de sélection artificielle. De la même manière un « biais de recrutement » lié à l'étude d'un échantillon de la population, peut fausser les fréquences alléliques réelles de la population et donc l'identification de signatures de sélection.

1.3.3.1.2. Exploitation du déséquilibre de liaison

Pour ce type de méthodes, la notion de liaison génétique est sous-jacente. En effet, dans le cadre de signatures de sélection récente, la recombinaison génétique n'a pas le temps de casser les haplotypes sélectionnés. La diversité génétique est donc réduite dans l'entourage du locus sous sélection, entraînant la création de longs haplotypes à faible diversité génétique dans la population, c'est ce qu'on appelle un balayage sélectif (cf. partie 1.2.2.1.4. la sélection). Les méthodes basées sur le déséquilibre de liaison (DL) sont puissantes pour détecter des balayages sélectifs forts (c'est-à-dire, de longs haplotypes en forte fréquence) en cours d'établissement (non fixés).

La mesure EHH (= *Extended Haplotype Homozygosity*) permet d'évaluer le degré de DL entre un SNP focal et les SNPs situés à différentes distances physiques x du locus d'étude. L'EHH est définie comme la probabilité que deux chromosomes choisis au hasard et qui portent l'haplotype focal, soient identiques par descendance pour l'ensemble de l'intervalle entre le SNP focal et le point x (Sabeti et al., 2002). La décroissance DL, et donc l'échelle relative du DL à "courte" et "longue" distance, dépend des taux de recombinaison locaux. Notre test de sélection positive consiste à trouver un haplotype central avec une combinaison de fréquence élevée et d'EHH élevée, par rapport à d'autres haplotypes centraux au locus. Une statistique nommée *iHS* (*integrated Haplotype Score*) a été proposée par Voight et al (2006) pour comparer les profils EHH intégrés obtenus pour un SNP dans sa version ancestrale versus mutée. Les valeurs extrêmes de l'*iHS* sont des traces de sélection positive car un haplotype de base avec un EHH inhabituellement élevé et une fréquence élevée dans la population indique la présence d'une mutation qui s'est propagée à travers la population plus rapidement que prévu sous un modèle d'évolution neutre.

Un aspect attrayant de cette approche est que les divers haplotypes de base à un locus servent de contrôles internes les uns pour les autres, en ajustant toute irrégularité dans le taux de recombinaison local et donc en permettant de distinguer les traces de sélection des signaux démographiques.

Cette mesure nécessite le phasage des données de génotypage, c'est-à-dire la reconstruction des haplotypes pour chacun des individus. Cette étape est cruciale, car elle va conditionner la précision de la détection des signatures de sélection.

Une autre méthode basée sur le DL a été développée, le LDD (*Linkage Disequilibrium Decay*). Contrairement à la statistique EHH, cette méthode ne nécessite pas d'informations haplotypiques (phasage) mais s'appuie sur les mêmes principes que l'EHH.

1.3.3.1.3. Exploitation de la diversité génétique

Deux méthodes se focalisent sur l'identification de régions génomiques ayant subi une réduction du polymorphisme par rapport au reste du génome. Ces méthodes sont basées sur les îlots de consanguinité (ROH) (McQuillan et al., 2008) ou bien sur des groupements d'hétérozygotie (Rubin et al., 2010).

Dans le cadre de l'utilisation des ROHs, on recherche des régions où une partie importante de la population partage de longs segments homozygotes, ce qui correspond à une baisse de la diversité génétique de la population sur cette portion du génome.

Concernant les groupements de variants hétérozygotes, cela consiste à simplement compter le nombre d'hétérozygotes dans une fenêtre par rapport à la moyenne du nombre d'hétérozygotes dans le génome (Rubin et al., 2010), permettant d'identifier les régions en déficit ou bien au contraire en excès d'hétérozygotes.

1.3.3.2. Signatures de sélection inter-populations : différenciation entre deux populations

Les signatures de sélection détectées entre populations permettent d'identifier les points de divergence, c'est à-dire de forte différenciation entre deux populations (Zhao et al., 2015). On peut alors mettre en lumière des régions génomiques sélectionnées dans une population mais sous évolution neutre (= absence de sélection) au sein de l'autre population. Pour cela, il est possible d'utiliser soit simplement l'information des *loci* (SNPs), soit celle des haplotypes.

Deux méthodes existent pour différencier les populations en approche par locus : l'indice de fixation de Wright (F_{st} ; Wright, 1931) et le test FLK (Test de Lewontin-Krakauer ; Bonhomme et al., 2010). Un indice de fixation (F_{st}) localement extrême par rapport à la distribution des F_{st} le long du génome suggère la présence d'une sélection positive ou équilibrante dans l'une des deux populations. Tout comme les méthodes de détection intra-population basées sur les fréquences alléliques, cette méthode peut conduire à la détection de

faux positifs du fait de la structure des populations (Nei & Maruyama, 1975). La statistique FLK correspond à un F_{st} implémenté d'une matrice de parenté entre les populations. Cet ajout permet de corriger le test pour la structure des populations. Ainsi, la statistique FLK tient compte des changements dynamiques de la taille effective des populations au fil du temps et réduit donc le risque de détecter des faux positifs par rapport à l'approche F_{st} .

Concernant les approches haplotypiques de détection des signatures de sélection, une première méthode est basée sur les EHH (cf. section 1.3.3.1.2.). La statistique XP-EHH (Cross Population EHH, Sabeti et al., 2007) compare les profils intégrés d'EHH de deux populations pour identifier les différences majeures. Cette méthode est utilisée pour détecter les balayages sélectifs dans lesquels un allèle a approché ou atteint la fixation dans une population tout en restant polymorphe dans l'autre. La seconde méthode est hapFLK (Fariello et al., 2013), qui est une extension haplotypique de la statistique FLK. Cette méthode, contrairement à la statistique XP-EHH, ne nécessite pas forcément le phasage des données.

En fonction des méthodes utilisées, plusieurs problèmes et biais peuvent apparaître. En effet, il n'y a pas un seul modèle d'évolution neutre (évolution des génomes non soumis à la sélection) mais de multiples modèles d'évolution du génome dépendant de l'histoire des populations. Effectivement, une population peut ne pas être soumise à sélection mais subir différents événements démographiques (goulots d'étranglements, expansion rapide, etc.) agissant alors sur le profil du polymorphisme sur tout le génome, induisant la détection de faux positifs lors de l'identification des signatures de sélection. En parallèle, la sélection peut être forte sur le phénotype, mais agir sur une multitude de gènes à effet faible, la sélection ne laisse donc pas toujours les mêmes signatures dans le génome. C'est pourquoi, l'utilisation d'approches variées pour détecter des signatures de sélection permet d'identifier des traces de sélection ne pouvant pas être détectées par toutes les métriques.

A Retenir – 1.3. Méthodes d'évaluation de la diversité génétique et d'identification de signatures de sélection

- Les ROHs permettent de calculer des coefficients de consanguinité individuels réels. Cette mesure génomique devient la méthode de référence pour estimer des coefficients de consanguinité.
Les populations commerciales de truite arc-en-ciel ont des niveaux de consanguinité globalement élevés et très variables le long des génomes.
- Les effets de dépression de consanguinité chez la truite arc-en-ciel n'ont jusqu'à présent jamais été estimés sur la base d'informations génomiques.
- La prise en compte des effets de dominance semble importante pour comprendre les mécanismes génétiques sous-jacents à la dépression de consanguinité.
- La sélection laisse des traces dans le génome : des signatures de sélection identifiables à l'aide d'approches et d'outils variés.
- Les régions génomiques sous sélection peuvent contenir des gènes d'intérêt liés à des phénomènes tels que la *fitness* des individus, la domestication ou la sélection des populations.

1.4. Questions soulevées

L'étude réalisée préalablement à mon projet de thèse (D'Ambrosio et al., 2019) concluait que les processus récents de domestication et de sélection avaient amené à réduire significativement la diversité génétique des populations domestiquées et sélectionnées de truite arc-en-ciel, amenant les populations sélectionnées à des niveaux élevés de consanguinité (~20%) et des effectifs efficaces inférieurs à 50. Ces résultats demandent à être investigués en profondeur car les plans d'accouplement actuellement réalisés sont des factoriels partiels faisant intervenir à chaque génération environ 70 à 100 reproducteurs de chaque sexe, soit des effectifs efficaces espérés compris entre 140 et 200 reproducteurs (Dupont-Nivet et al., 2006). De plus, dû aux niveaux importants de consanguinité dans les populations de truites, des phénomènes de dépression de consanguinité peuvent y être observés notamment sur des caractères de taille et de reproduction (Kincaid, 1976 ; Su et al., 1996 ; Rye & Mao, 1998 ; Pante et al., 2001 ; Wang et al., 1999, 2002 ; Naish et al., 2013 ; Waters et al., 2020).

Par ailleurs, les travaux préliminaires des localisations des ROHs par Jonathan D'Ambrosio (résultats non publiés) laisse supposer que les zones d'homozygotie partagées par de nombreux individus intra ou entre populations ne sont pas réparties aléatoirement le long du génome, mais correspondent à des zones potentiellement enrichies en gènes impliquées dans les processus de développement embryonnaire, immunité et reproduction.

Dans ce contexte, mon projet de thèse est d'étudier **l'évolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel**. La finalité de mes travaux est triple :

- 1- Comprendre les verrous reproductifs amenant à un fort accroissement de la consanguinité en élevage (même sans sélection artificielle)
- 2- Identifier les principales régions du génome qui contribuent à la dépression de consanguinité
- 3- Identifier les signatures de sélection liées à la domestication et à la sélection

Dans un premier temps, nous avons cherché à valider l'hypothèse émise dans l'étude de D'Ambrosio et al. (2019) afin d'expliquer la perte de diversité génétique dans des populations non-sélectionnées de truites arc-en-ciel. L'hypothèse concernait la présence de mécanismes lors du développement embryonnaire des descendants, induisant un déséquilibre de contribution parentale à la génération suivante. Ensuite nous avons essayé de comprendre les mécanismes

impliqués dans la survie des alevins. Ainsi, pour cela, les questions suivantes ont été traitées dans le chapitre 2 : **quels sont les verrous reproductifs conduisant à une perte de diversité génétique dans une population domestiquée de truite arc-en-ciel ? Peut-on identifier des effets de dépression de consanguinité sur la survie embryonnaire ? Quels sont les composantes génétiques de la survie précoce ? Quelles sont les caractéristiques maternelles associées à une meilleure survie des alevins ?**

Nous savons qu’au sein des populations de truites arc-en-ciel françaises, les niveaux de consanguinité sont très variables le long du génome (D’Ambrosio et al., 2019). L’objectif du chapitre 3, était de caractériser les effets locaux de la consanguinité dans une population commerciale de truites. Trois questions importantes ont été traitées dans ce chapitre. La première question était de **quantifier les effets globaux de la consanguinité sur les performances de taille et de reproduction des truites**. La seconde question était **d’identifier les zones du génome associées à ces variations de performance : Des régions particulières sont-elles plus souvent impliquées ? Sont-elles liées à des événements récents d’accouplements entre reproducteurs apparentés ?** La dernière question était de savoir si **nos observations corroboraient l’hypothèse privilégiée pour expliquer le phénomène de dépression de consanguinité**, à savoir le rôle majeur d’allèles récessifs à effet délétère.

La dernière étape de ce projet consistait à identifier les signatures de sélection, positives ou équilibrantes, communes entre plusieurs populations de truites arc-en-ciel domestiquées. Cette étude est présentée dans le chapitre 4, nous avons pu répondre aux questions suivantes : **Existe-t-il des régions du génome des truites arc-en-ciel, très homozygotes, sous sélection positive conservées entre plusieurs populations domestiquées ? Ou bien des régions hétérozygotes, sous sélection é également partagées entre plusieurs populations domestiquées ? Ces régions sont-elles simplement en lien avec des fonctions essentielles dans la survie et la reproduction des truites ? Ou sont-elles fortement liées au processus de domestication des populations ?**

Pour finir, dans un 5^{ème} chapitre et à la lumière des connaissances acquises dans cette thèse, je proposerai des éléments de réflexion pour valoriser et gérer au mieux la diversité génétique à chaque étape d’un programme de sélection truiticole.

Chapitre 2

Etude du succès reproductif et de la survie embryonnaire dans une lignée domestiquée de truite arc-en-ciel

Une perte de diversité génétique non négligeable a été observée dans une lignée expérimentale de truites arc-en-ciel constituée par l'INRA et entretenue depuis le début des années 1980 sans sélection volontaire (D'Ambrosio et al., 2019). La forte variation du succès reproducteur généralement observée chez les poissons a été suggérée comme mécanisme premier susceptible d'expliquer cette baisse de diversité génétique. La valeur sélective (« fitness ») est un concept central de la biologie évolutive qui décrit le succès reproducteur d'un individu, succès qui dépend de sa fécondité propre et de la survie de sa descendance jusqu'à la maturité sexuelle. Ces deux composantes méritent de l'attention chez la truite arc-en-ciel en raison de leur impact sur l'équilibre ou le déséquilibre induit dans les contributions parentales à la génération suivante. En outre, la survie des individus est un caractère associé à la robustesse particulièrement important à considérer chez les poissons car elle quantifie la résistance globale d'un individu aux multiples facteurs de stress environnementaux

Par conséquent, dans ce chapitre nous avons étudié la variance des tailles de familles, ainsi que les facteurs biologiques ayant une incidence sur la survie précoce chez la truite arc-en-ciel. Plus précisément, nous avons souhaité répondre aux questions suivantes : **Quels sont les verrous reproductifs conduisant à une perte de diversité génétique dans une population domestiquée de truite arc-en-ciel ? Peut-on identifier des effets de dépression de consanguinité sur la survie embryonnaire ? Quels sont les composantes génétiques de la survie précoce ? Quelles sont les caractéristiques maternelles associées à une meilleure survie des alevins ?**

Pour cette étude nous avons réalisé une expérimentation à la Pisciculture Expérimentale INRAE des Monts d'Arrée (PEIMA) afin d'évaluer la présence d'un déséquilibre significatif de contribution parentale (15 mères et 9 pères) dans la cohorte des descendants (945 descendants). Nous avons évalué la survie à trois stades précoces : stade œillé, éclosion et à trois semaines après la première alimentation. Nous avons cherché à identifier les facteurs biologiques (consanguinité individuelle et parentale, phénotype maternel : taille/poids des femelles ; poids de la ponte, taille des œufs, nombre d'œufs dans la ponte) impactant la survie précoce.

Conclusion du Chapitre 2

Cette étude a permis de mettre en évidence un déséquilibre de contribution des mères à la génération suivante au sein d'une lignée domestiquée.

L'analyse génétique de la survie embryonnaire a révélé une très faible héritabilité mais d'importants effets maternels et de dominance. Malgré une variance de dominance élevée, la survie des alevins ne semble pas affectée par leur niveau de consanguinité, ni ceux de leurs parents. Notons toutefois que les niveaux de consanguinité observés étaient modérés dans cette étude (en moyenne 0.10).

La survie des alevins s'est révélée positivement corrélée avec la fécondité des mères (nombre d'œufs) et négativement corrélée avec le poids (après ponte) de la mère. En revanche, survie précoce et taille de l'œuf ne semblent pas corrélées.

En raison des forts effets de dominance observés sur la survie des alevins, l'exploitation des effets d'hétérosis par le croisement de géniteurs génétiquement divers pourrait être intéressant. De plus, la sélection des femelles les plus fécondes pourrait aider les éleveurs à éviter le déséquilibre de contribution maternelle à la génération suivante et ainsi préserver la diversité génétique dans les programmes de sélection.

Les différents résultats obtenus dans ce chapitre ont été présentés sous la forme d'un poster au congrès *European Aquaculture Society* (EAS) et d'une présentation orale aux *Journées Techniques du Sysaaf* en 2021. L'ensemble des résultats obtenus a fait l'objet d'un article « **Maternal and genetic effects on embryonic survival from fertilization to swim up stage and reproductive success in a farmed rainbow trout line** » soumis à la revue *Aquaculture Reports* en octobre 2022 et présenté dans ce chapitre.

Maternal and genetic effects on embryonic survival from fertilization to swim up stage and reproductive success in a farmed rainbow trout line

Paul, K.¹, Pelissier P.², Goardon L.², Dechamp N.¹, Danon J.³, Jaffrelo L.³, Poncet C.³,
Dupont-Nivet M.¹, Phocas, F.¹

1. Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

2. INRAE, UE 0937 PEIMA, 29450 Sizun, France

3. INRAE-UCA, Plateforme Gentyane, UMR GDEC, 63000 Clermont-Ferrand, France

Corresponding author : Florence Phocas - florence.phocas@inrae.fr

Highlights

- Dam contribution to the next generation is moderately unbalanced in farmed trout
- Maternal and non-additive genetic effects are the main factors explaining early survival
- Within the range studied, inbreeding is not a significant player in early survival
- Small females but high fecund females show better survival of their offspring
- Eggs of over-average size produce more abnormal alevins and lower early survival

Abstract

Reproductive success and offspring survival until sexual maturity are essential traits both for fish fitness and aquaculture development. Variation in offspring's survival among family results in unbalanced parental contributions to the next generation and may explain the loss of genetic diversity observed in some farmed populations. Therefore, we studied the variance in parental contributions to a progeny cohort, as well as the biological factors impacting offspring early survival in rainbow trout. The data consisted of 945 individual survival observations from fertilization to the juvenile stage from 135 full-sib families of the INRAE experimental synthetic line. Survival was assessed at eyed-egg stage, hatching, and 3 weeks after first feeding. We used a full-factorial mating design to partition phenotypic variance in early survival traits into maternal and additive genetic effects under threshold GBLUP models considering the inclusion of genomic information for 32,725 SNP. Average offspring survival proportions were 91.0% at the eyed-egg stage, 87.2% at hatching, and 84.4% three weeks after first feeding. Significant unbalanced dam contributions were observed at the eyed-egg and hatching stages. Low heritability was estimated for early survival traits ($h^2=0.20\pm 0.12$ and 0.13 ± 0.09 for survival from egg-eyed stage and, respectively, hatching and first feeding), revealing that additive genetic variance was not significantly different from zero, while maternal effects explained a larger part ($c^2=0.37\pm 0.16$ and 0.15 ± 0.07 , respectively) of the phenotypic variances. Phenotypically, offspring early survival was positively correlated with dam fecundity, while it was negatively correlated with dam post-spawning weight. Negative, but not significant association was observed between early survival and dam's average egg weight. If a study of genetic correlations confirms the phenotypic trends observed here, promoting high fecund females should help the breeders to increase offspring early survival and to maintain genetic diversity in breeding programs.

Keywords: fish, parental contribution, heritability, inbreeding, maternal effects

1. Introduction

In animal breeding, a major goal in the context of global environmental changes (climate change, acidification, eutrophication, water resources, etc.) is that animals should be robust against multiple environmental disturbances and stressors throughout their life-span. Therefore, there is a growing interest in selecting for robustness in farm species, as it allows promoting less-controlled and low-input systems where rearing conditions are more variable and unpredictable (Phocas et al., 2016). Survival is an ultimate robustness trait in fish breeding because it measures an individual's resistance to multiple mortality factors in a specific environment (Vehviläinen et al., 2008). Nevertheless, the first step to reach this ultimate goal is to ensure a high reproductive success through high fecundity of individuals linked to high survival of embryos and juveniles. In rainbow trout, little is precisely known about survival at embryo and larval stages, but juvenile fingerling survival is estimated at over 90% (Vehviläinen et al., 2010). Males are sexually mature at 1 or 2 years-old, while females attain maturity later at 2 or 3 years-old. In the wild, most fish only spawn once or twice in their lifetime (Christie et al., 2018), but females may reproduce once a year for 2 or 3 years in farming conditions. In good farming conditions, 2-year-old dams can produce 4,700 eggs on average, corresponding to around 2,300 eggs per kg of body weight (D'Ambrosio et al., 2020).

The contributions of genetic and maternal environmental effects to offspring survival can shift during development. Maternal components tend to be larger at early life (egg and alevin) stages because of the influence of maternal investment, such as egg quality, whereas additive genetic effects tend to be larger at later life (fry) stage, as shown in Atlantic salmon (Houde, 2015) and more generally in any animal species (Cheverud et al., 1983). Several studies in salmonids and carps have shown small sire effects but large dam effects on early survival (Herbinger et al. 1995; Fishback et al. 2002; Vandeputte et al. 2004; Houde et al., 2013). The genetic architecture of early survival traits was investigated under sire-dam genetic models in salmonids species (Kanis et al., 1976; Rye et al., 1990; Vehviläinen et al., 2008; Houde et al., 2015). All studies concluded that heritability of early survival was low to moderate, but impact of non-additive genetic factors and maternal environmental effects was important.

In fish breeding, high unbalanced progeny numbers between parents can be observed under mass spawning. Sperm competition under hatchery conditions using pools of sperm is a well-known phenomenon explaining unbalanced sire contributions in salmonids (Gharrett and

Shirley, 1985; Gage et al., 2004) and carps (Kaspar et al., 2008). One of the consequences of the high variation in fish reproductive success and family size is the decrease in effective size and subsequent loss of genetic diversity in the next generation (Wedekind et al., 2007). Such phenomenon has been suggested as a likely factor to explain the observed loss of genetic diversity in a farmed rainbow trout line that was kept unselected since the early 1980s (D'Ambrosio et al., 2019). In order to identify the key factors that may explain variance in reproductive success and thus loss of genetic diversity in a domesticated population of rainbow trout (*Oncorhynchus mykiss*), we designed an experiment to determine maternal and genetic factors impacting offspring survival during embryonic and early larval development (from fertilization to three weeks after the first feeding). This topic has not been covered so much so far, and to our knowledge, no study had genotyped several hundreds of individuals to accurately follow parental contributions and assess genetic parameters of offspring survival. Two main issues are raised in this paper: how do the early larval stages affect the variance in family size in farmed trout? What are the main genetic and maternal biological factors explaining variation in reproductive success and alevin survival in farm conditions?

2. Materials and Methods

Experimentation was conducted at the INRAE experimental facility PEIMA (Sizun, France – Agreement number B29-277-02).

2.1. Population

The INRAE synthetic line was initially developed by inter-crossing several domesticated lines of rainbow trout to create a population with large genetic variability. The line was constituted from a mixture of French farmed populations with some new introductions from Denmark and the USA in the early 1980s and then closed to outside germplasm (D'Ambrosio et al., 2019). Since then, the population has been maintained without intentional selection using a full factorial mating design between randomly chosen parents. About 60 dams and 80 sires were used at 2 years of age to produce each progeny cohort. An incident in the experimental facility led to the production of the 2019 study cohort from parents bred at 3 years of age.

2.2. Experimental mating design

A full factorial mating plan was performed between 15 ovulated dams and 9 fluent sires of the INRAE rainbow trout synthetic line. The 3-year-old dams and sires were randomly divided into 3 groups of 5 females and 3 males, respectively. The groups of males were named Ma, Mb, and Mc, and males were labeled from 1 to 3 within their group name. In the same way, female groups were named Fa, Fb, and Fc, with a label from 1 to 5 for females within their group name. An ovulation checking was performed once a week, and brooders were maintained on average at 9°C during the spawning season, to limit the occurrence of overripe eggs, as stated by Sakai (1975) and Bry (1981). Oocytes from females of the same group were gently mixed before the fertilization process as routinely performed on the INRAE experimental facility. To avoid sperm competition (Pilastro et al., 2002; Wedell et al., 2002), the fertilization of 35 ova from each female group (i.e. 7 ova per female) was carried out male-by-male (Figure 1).

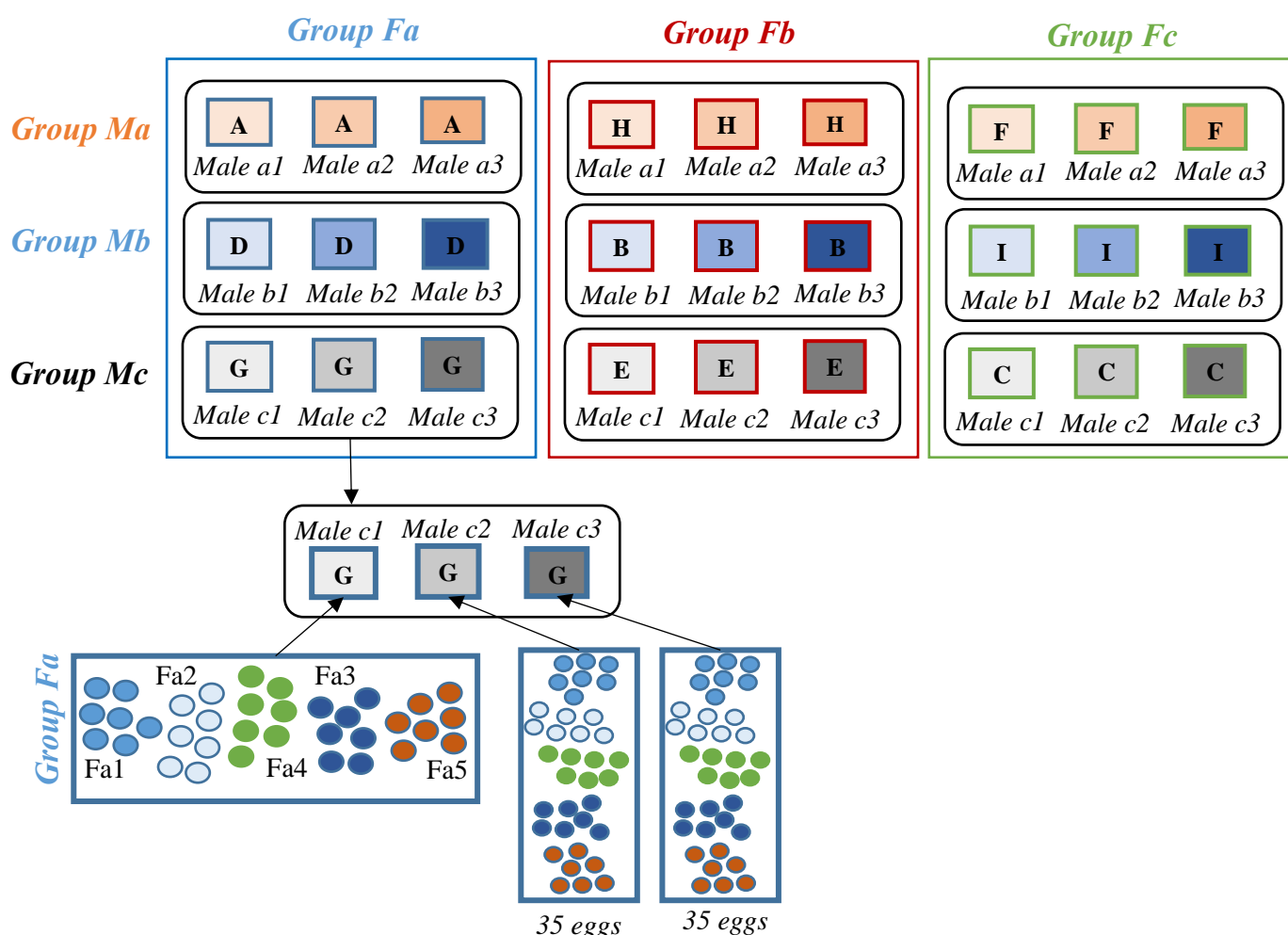


Figure 1. Experimental mating design between the 3 male and 3 female groups and distribution of fertilized eggs in 9 incubators (from A to I).

For each male, 1 mL of sperm was mixed with 1 mL of StorFish® (IMV, France), a dilution and preservation medium for salmonid milt. For each female group fertilization, 0.20 mL of sperm/StorFish mixture of one male was added to the group of 35 eggs. Each pool was then mixed with 15 mL of ActiFish solution (Supplementary Table A.1) to spread sperm over all the oocytes, and spring water was added 5 minutes later to activate oocyte fertilization. One hour after fertilization, all 105 eggs from a small factorial plan were grouped together in the same incubator until the eyed egg stage. The full mating design thus included 9 small factorial plans (each group of 5 females mated to each group of 3 males), generating 9 batches of 105 eggs placed in 9 egg incubators (named A to I) as described in Figure 1. The 9 batches were duplicated to assess the consistency of the quantification of parental contributions at the progeny eyed egg stage. Incubators were positioned in the hatching trough as in Supplementary Figure A.1 with duplicate batches facing each other.

To summarize figures in this experimental mating design, each sire could fertilize 105 eggs in each replicate, and each dam gave 63 eggs, and each pair may have up to seven offspring observed at the eyed egg stage (20 dpf at 12°C). Totally, 945 eggs per replicate were fertilized.

2.3. Dam phenotypes

Dam phenotypes were measured on the day of spawning. Among the 17 females collected, two females were discarded due to the presence of blood in the egg mass for the first one and of over-ripe eggs for the second one.

Female post-spawning weight (PW in g) and the fork length (FL in mm), spawn weight (SW) and the weight of the coelomic fluid (CF) were recorded. Two batches of 50 eggs were randomly collected from each egg mass to obtain the weight of 50 eggs and then, by dividing by 50, an average egg weight (EW) for each female. The ratio SW / EW allowed us to estimate the egg number in the spawn (EN). Then, the relative fecundity of each dam (RF) was calculated as EN/PW .

An additional 20 g sample of each egg mass was taken to get an accurate estimate of egg diameter (ED) and its coefficient of variation to assess spawn heterogeneity. To do so, a picture of the egg sample after 24h of hydration in spring water was obtained using the *VisEgg* system (Cardona et al., 2020), a dedicated shooting system consisting of a light tablet and digital SLR camera (canon EOS 1000D, resolution: 10.1 M pixels). Each picture was analyzed with *Visilog*

7.3 software (Thermo Scientific) to allow the automatic detection and measurement of each egg's diameter. In each 20 g sample of egg mass, there were, on average, 318 oocytes.

2.4. *Offspring phenotypes*

Standard husbandry practices were followed, with a daily bath with bronopol and daily dead eggs removal and counting. At the eyed egg stage, there were still 890 embryos in the first replicate and 858 in the second. All 890 embryos from the first duplicate of the experimental mating were sampled to be genotyped for parentage assignment. This first duplicate was only used to get the first estimate of parental contributions at the eyed egg stage. Due to the low number of eggs considered per mating pair, we were willing to assess the consistency of these early parental contributions by comparing estimates from two replicates. The second set of 858 embryos remained under observation from the eyed egg stage until the end of the experiment, three weeks after the first feeding (65 dpf at 12°C). All dead fry were collected each day, and the presence of malformation (spine, head or caudal fin) was recorded. In addition, fry that did not complete yolk sac absorption were considered malformed. All fry (dead or alive) were sampled for genotyping.

2.5. *Genotypes and parentage assignment*

Genotyping was performed at the INRAE Gentyane Genotyping Platform (Clermont-Ferrand, France). The 890 offspring of the first replicate were genotyped for parentage assignment using 96 SNP in KASPar technology. Dynamic Array™ IFC 96*96 chips were used with Biomark™ HD Reader to perform the competitive PCR and chip reading. The Fluidigm® SNP Genotyping Analysis software was used to analyze the genotyping results. The panel of 96 SNP was chosen among the SNP from the Axiom™ Trout Genotyping array distributed throughout the whole genome and have a minor allele frequency (MAF) above 0.30 in the five French rainbow trout lines described in D'Ambrosio et al. (2019). This 96 SNP-panel is presented in Table A.2.

The 24 parents and the 858 offspring from the second replicate were genotyped for 57,501 SNP using the Axiom™ Trout Genotyping array (Palti et al., 2015). The quality control of genotyped SNP was performed as described by D'Ambrosio et al. (2019) to remove SNP with probe polymorphism and multiple locations on the trout genome. Only the 32,725 SNP with a call rate higher than 0.97, a test for the Hardy Weinberg equilibrium with a p-value >0.0001, and a MAF higher than 0.05 were retained for further analysis. All missing genotypes for the 32,725 SNP were imputed using FImpute software with default parameters (Sargolzaei et al., 2014).

We used the R package *APIS* (Griot et al., 2020) to assign offspring of both replicates to their parents using a likelihood approach with decision rules based on the observed distributions of Mendelian transmission probabilities. We considered a maximum threshold of 1% for parentage assignment error.

2.6. Inbreeding coefficient calculation

Runs of homozygosity (ROH) were identified for each fish using the PLINK v1.9 *homozyg* function (Chang et al., 2015), as defined by D'Ambrosio et al. (2019), with the following parameters: '*--homozyg-kb 1000 --homozyg-window-snp 30 --homozyg-snp 30 --homozyg-gap 1000 --homozyg-density 100 --homozyg-het 1*'.

The inbreeding coefficient (F_i) for any fish was calculated as the sum of ROH lengths ($\sum Length_{ROH,i}$) in this individual i divided by the total length of the autosomal genome covered by SNPs (L_{Genome}):
$$F_i = \frac{\sum Length_{ROH,i}}{L_{Genome}}$$

The total size of the autosomal genome covered by SNPs (= 1,874 Mb) was considered as the length of the genome, removing gaps of more than 1 Mb without any SNP from the total size. Relatedness between pairs of parents was recovered using identity by descent (IBD) estimates from PLINK v1.9 software (*genome* function) (Purcell et al., 2007).

2.7. Statistical analyzes

Parental contributions were derived by counting each parent and pair's number of alive and dead progeny. With the help of R package *stats*, fisher exact tests were performed (*fisher.test* function) to assess potential unbalanced parental contributions to alive offspring at three stages: at the eyed egg stage for both replicates, at hatching, and 3 weeks after the first feeding for the second replicate. To control from multiple testing, the false discovery rate (FDR) method by Benjamini and Yekutieli (2001) was used to compare contributions two-by-two (*p.adjust* function).

In addition, offspring survival was evaluated from the day of fertilization to the time of the three stages for the second replicate: eyed egg stage (Se), hatching (Sh), and 3 weeks after first feeding (Sf). Survival was also assessed between stages from eyed egg to hatching (Seh) and from hatching to 3 weeks after first feeding (Shf). Pearson correlations were computed between maternal phenotypes and offspring survival rates.

To assess the importance of genetic and maternal factors on early survival, we considered two

phenotypes defined as binary traits: alive or dead from eyed-egg to hatching stages and from eyed-egg stage to three weeks after first feeding. From eyed-egg to hatching stages, 36 dead and 822 alive individuals were counted; 62 dead and 796 alive were counted from eyed-egg stage to the end of the experiment. The following genomic best unbiased prediction (GBLUP) models were considered to describe the vector of phenotypes (\mathbf{y}) of the 858 offspring alive at the eyed-stage and genotyped with the 57K chip.

Various genetic models (A- additive; AM- additive + maternal; AMD- additive + maternal dominance effects) were tested either considering (threshold model) or ignoring (linear model) the binary nature of the survival traits. The full tested GBLUP model described \mathbf{y} under a linear (underlying) scale as: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \mathbf{M}\mathbf{c} + \mathbf{b}\mathbf{F}_p + \mathbf{b}_{\text{sire}}\mathbf{F}_{\text{sire}} + \mathbf{b}_{\text{dam}}\mathbf{F}_{\text{dam}} + \mathbf{e}$

where $\boldsymbol{\beta}$, \mathbf{a} , \mathbf{d} , \mathbf{c} and \mathbf{e} are the vectors of fixed incubator effects and random additive genetic effects, dominance deviation effects, maternal effects and residual effects, respectively. \mathbf{X} , \mathbf{Z} , \mathbf{W} and \mathbf{M} are the incidence matrices for $\boldsymbol{\beta}$, \mathbf{a} , \mathbf{d} and \mathbf{c} , respectively. The vectors \mathbf{a} and \mathbf{d} included the additive and dominance genetic values of the 858 individuals related through the genomic additive relationship matrix \mathbf{G} and the dominance relationship matrix \mathbf{D} . The \mathbf{G} and \mathbf{D} genomic matrices were built using the `parallelf90` program (Vitezica et al., 2013).

The regression coefficients \mathbf{b} , \mathbf{b}_{dam} , \mathbf{b}_{sire} on \mathbf{y} of the vectors \mathbf{F}_p , \mathbf{F}_{dam} , \mathbf{F}_{sire} , respectively, account for individual (progeny) and parental inbreeding effects on the studied performance.

Whatever the model, there were no significant effects of parental inbreeding levels, so these effects were removed for the analysis. In addition, while the AMD linear model converged and gave the lowest AIC (Table A.4), we could not make converged the AMD threshold model.

Therefore, the final model retained as main result for the analysis and discussion of the results was: $\mathbf{y} = \boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{M}\mathbf{c} + \mathbf{b}\mathbf{F}_p$

The two traits were analyzed separately using BLUPf90 package (Misztal et al., 2014) and the AIREMLF90 program (Thompson et al., 2005) for linear models or THRGIBBSf90 program (Tsurata & Misztal, 2006) for threshold models. With the threshold model, the variance components were estimating using a Gibbs sampler with 200,000 cycles, 10,000 cycles of burn-in and 1 sample was kept every 40 cycles for posterior analysis. Convergence was checked by graphical inspection of both samples and posterior distributions of the parameters. All of these checks were implemented in POSTGIBBSf90 program (Aguilar et al., 2019).

3. Results

3.1. Genomic similarities among parents

The average inbreeding coefficients were close to each other for dams (12%) and sires (11%), with similar minimum values around 7% (Table 1). However, one dam had a very high inbreeding coefficient (28%) compared to all other parents (Figure 2).

Table 1. Summary statistics of inbreeding coefficients for dams, sires and their progeny.

	Fdam	Fsire	Fp
Mean	0.121	0.110	0.102
SD	0.055	0.038	0.030
Minimum	0.071	0.066	0.003
Maximum	0.279	0.181	0.249

	Fa1	Fa2	Fa3	Fa4	Fa5	Fb1	Fb2	Fb3	Fb4	Fb5	Fc1	Fc2	Fc3	Fc4	Fc5	Ma1	Ma2	Ma3	Mb1	Mb2	Mb3	Mc1	Mc2	Mc3	
Fa1	0.09	0.00	0.03	0.02	0.00	0.1	0.00	0.00	0.00	0.05	0.08	0.00	0.06	0.11	0.00	0.00	0.00	0.02	0.00	0.00	0.08	0.07	0.00	0.00	
Fa2	0.13	0.00	0.09	0.00	0.09	0.00	0.00	0.00	0.00	0.06	0.09	0.12	0.13	0.04	0.00	0.00	0.00	0.00	0.04	0.12	0.04	0.00	0.00	0.00	0.00
Fa3	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Fa4	0.07	0.03	0.14	0.13	0.09	0.13	0.09	0.13	0.03	0.08	0.09	0.13	0.12	0.1	0.08	0.00	0.12	0.1	0.08	0.1	0.04	0.00	0.11		
Fa5	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	
Fb1	0.09	0.05	0.07	0.08	0.09	0.1	0.06	0.04	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.1	0.09	0.13	0.01	0.06	0.00	0.03	
Fb2	0.08	0.07	0.09	0.00	0.06	0.08	0.03	0.00	0.03	0.08	0.03	0.03	0.08	0.03	0.08	0.03	0.11	0.09	0.06	0.07	0.05	0.00	0.04		
Fb3	0.12	0.09	0.00	0.05	0.03	0.04	0.00	0.04	0.04	0.00	0.04	0.04	0.00	0.04	0.04	0.00	0.03	0.00	0.12	0.07	0.07	0.00	0.00	0.00	0.00
Fb4	0.07	0.00	0.1	0.21	0.1	0.00	0.1	0.09	0.00	0.11	0.09	0.06	0.00	0.02	0.00	0.02	0.11	0.09	0.06	0.00	0.02	0.00	0.12		
Fb5	0.14	0.28	0.14	0.00	0.13	0.00	0.00	0.08	0.00	0.08	0.00	0.08	0.00	0.03	0.01	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fc1	0.1	0.14	0.12	0.15	0.07	0.00	0.27	0.11	0.03	0.08	0.06	0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.08	0.06	0.05	0.00	0.00	0.03	
Fc2	0.09	0.03	0.09	0.1	0.14	0.00	0.09	0.02	0.12	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.27	0.07	0.00	0.05		
Fc3	0.09	0.12	0.05	0.00	0.00	0.05	0.04	0.07	0.27	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.13	0.07	0.00	0.00	0.00	
Fc4	0.12	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.13	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.07	0.13	0.07	0.00	0.00	0.00	
Fc5	0.12	0.1	0.00	0.05	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.08		
Ma1	0.15	0.00	0.08	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.05	0.05	0.00	0.00	0.00	0.06	
Ma2	0.16	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	
Ma3	0.08	0.12	0.09	0.08	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.09	0.08	0.15	0.00	0.04		
Mb1	0.09	0.00	0.04	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.04	0.05	0.00	0.15		
Mb2	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.08		
Mb3	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.27	0.00	0.00	0.00	0.00		
Mc1	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00		
Mc2	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00		
Mc3	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00		

Figure 2. Inbreeding coefficients (diagonal terms) and identical-by-descent (IBD) values (upper triangular matrix) among all parents.

The majority of the parents were almost unrelated to each other ($IBD \leq 5\%$ - Figure 2). However, two pairs of parents were related with IBD values consistent with a half-sib relationship (ranging from 0.20 to 0.30). On average, the inbreeding coefficient of the progeny was 0.10. Even for the two-parent pairs that were significantly related, the inbreeding coefficient of offspring was never extreme and did not exceed 0.25 (Table 1).

3.2. Offspring survival rates and parental contributions to the next generation

3.2.1. From fertilization to eyed egg stage (replicate 1 and 2)

From fertilization to the eyed egg stage, significant higher mortality was observed in the second replicate (R2), with 8.99% of dead offspring compared to 6.03% of dead offspring in the first replicate (R1). However, no significant ($p\text{-value} > 0.05$) difference between the offspring survival of each parent between both replicates was observed (Figure 3). Dams with a larger number of dead offspring in the first replicate generally also had a larger number of dead offspring in the second replicate.

In both replicates, disequilibrium in parental contribution, measured as alive eyed egg, was observed from fertilization to the eyed egg stage for dams but not for sires (Table 2).

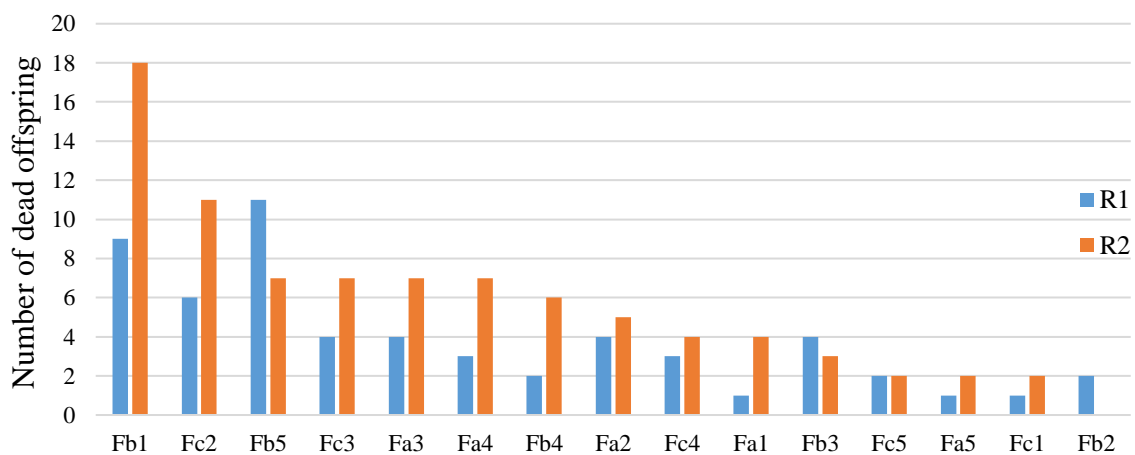


Figure 3. Number of dead offspring per dam in both replicates (R1 and R2) from fertilization to eyed egg stage. Fisher exact test was performed to determine statistical difference in offspring mortality between replicates of each dam.

Table 2. Fisher exact test for offspring survival per dam and sire from fertilization to eyed egg (Se), and for both replicates.

	Replicate 1		Replicate 2	
	Dams	Sires	Dams	Sires
Mean survival (Se)	94.0%	94.0%	91.0%	91.0%
SD	4.7%	3.4%	7.0%	3.7%
Minimum	82.5%	86.7%	71.4%	83.8%
Maximum	98.4%	98.1%	100.0%	95.2%
<i>p-value</i>	0.0059 *	0.0812	0.0005 *	0.1224

3.2.2. From fertilization to three weeks after the first feeding (replicate 2)

In Tables 3 and 4, offspring survival rates at different stages in replicate 2 are presented. Offspring survival rate from fertilization to 3 weeks after the first feeding (Sf) varied from 63.5 % to 95.2% among dams (Table 3) and from 78.1% to 91.4% among sires (Table 4). Larger differences in offspring survival were observed between dams (up to 33.3 %-unit) than between sires (up to 13.3 %-unit) at all stages. Statistically significant unbalanced contributions were only observed for dams (Table 3). Even if an unbalance between dams was seen on Sf, from hatching to three weeks after the first feeding (Shf), no significant effect of the dam was observed. In addition, no significant effect of the dam-sire pairs was observed, but this can be due to the low number (7) of offspring per pair.

Table 3. Statistics of progeny survival (Replicate 2) per dam at different stages and associated Fisher exact tests of dam differences in the numbers of dead and alive offspring.

	Mean	SD	Minimum	Maximum	P-value
Se	91.0%	7.0%	71.4%	100%	0.0005
Sh	87.2%	9.7%	65.1%	98.4%	0.0005
Sf	84.4%	9.3%	63.5%	95.2%	0.0005
Seh	95.6%	5.2%	82.1%	100 %	0.0005
Shf	96.9%	2.7%	92.3%	100%	0.2364

Se: survival from fertilization to eyed egg stage; Sh: from fertilization to hatching; Sf: from fertilization to three weeks after the first feeding; Seh: from eyed egg to hatching; Shf: from hatching to 3 weeks after the first feeding

Table 4. Statistics of progeny survival (Replicate 2) per sire at different stages and associated Fisher exact tests of sire differences in the numbers of dead and alive offspring

	Mean	SD	Minimum	Maximum	p-value
Se	91.0%	3.7%	83.8%	95.2%	0.1224
Sh	87.2%	3.9%	81.9%	93.3%	0.1559
Sf	84.4%	4.6%	78.1%	91.4%	0.1044
Seh	95.8%	2.5%	91.8%	99.0%	0.1584
Shf	96.8%	1.9%	93.3%	100.0%	0.3323

Se: survival from fertilization to eyed egg stage; Sh: from fertilization to hatching; Sf: from fertilization to three weeks after the first feeding; Seh: from eyed egg to hatching; Shf: from hatching to 3 weeks after the first feeding.

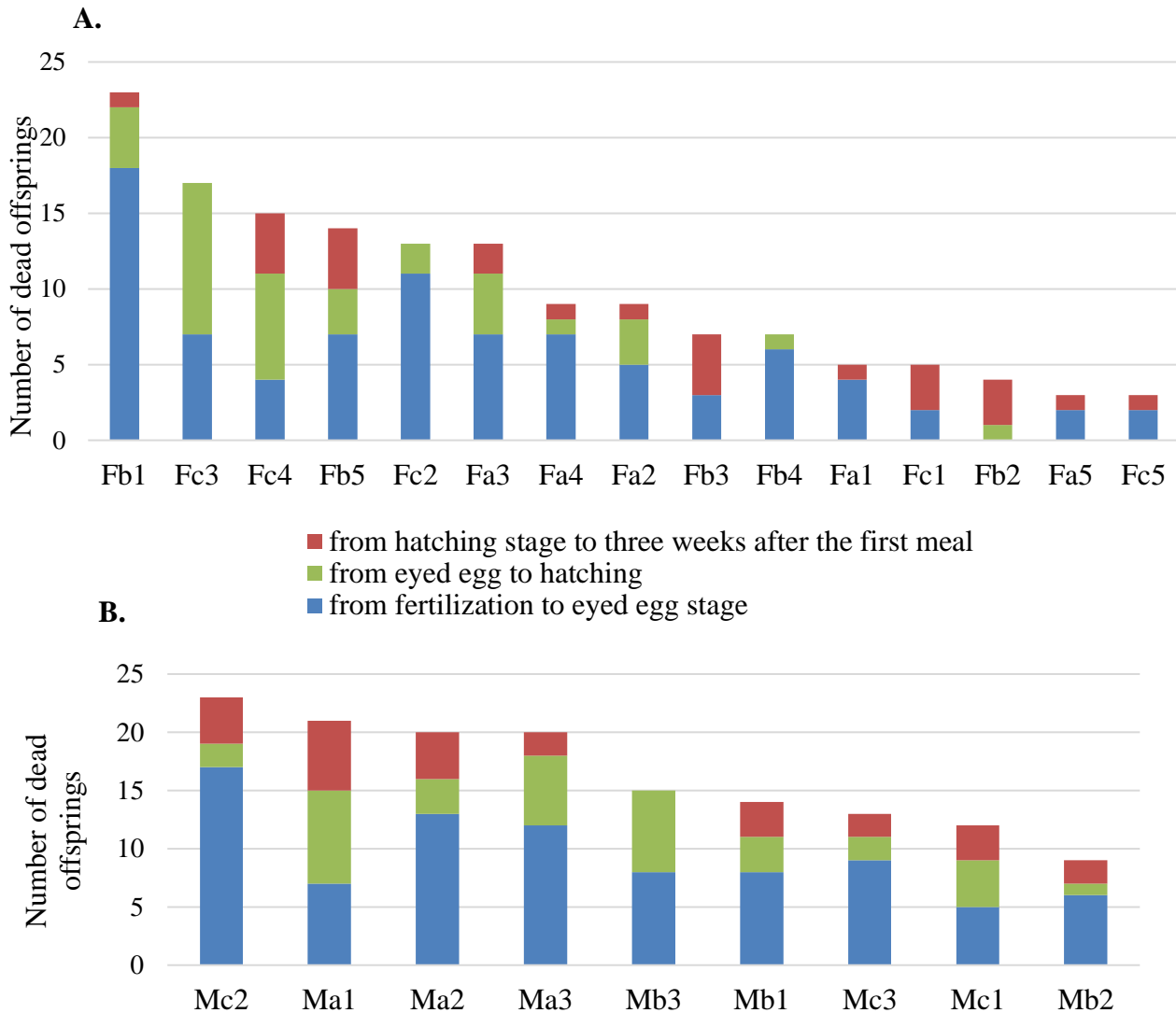


Figure 4. Number of dead offspring (barplot) from fertilization to eyed egg stage, from eyed egg stage to hatching and from hatching to three weeks after the first feeding, for dams (A.) and sires (B). *Fisher exact tests* were performed with FDR correction on offspring survival at Sf in Table A.3.

Figure 4 shows the number of dead offspring from fertilization to eyed egg stage, from eyed egg stage to hatching and from hatching to three weeks after the first feeding for each dam or sire. Dams and sires were classified according to the number of their dead offspring at the end of the experiment, with the highest number of total dead offspring to the left and the lowest to the right. Three dams have significantly higher offspring mortality than the three best dams with the lowest numbers of dead offspring (Table A.3). No significant difference for the number of dead offspring was observed across sires. It appears that both paternal and maternal offspring survival were not significantly correlated between the different stages (Table 5).

Table 5. Correlations between progeny survival rates at different stages (across dams in the upper triangular matrice and across sires in the lower triangular matrice).

	Se	Sh	Sf	Seh	Shf
Se	1	0.89*	0.84*	0.43	-0.27
Sh	0.81*	1	0.97*	0.79*	-0.18
Sf	0.76*	0.94*	1	0.80*	0.06
Seh	-0.15	0.45	0.42	1	-0.01
Shf	0.24	0.29	0.60	0.10	1

* significant correlation (p-value < 0.05).

3.3. Potential sources of imbalance in parental contributions

3.3.1. Quantification of genetic and maternal effects on offspring survival

Early survival variables had low heritability values of 0.196 for Seh and 0.134 for Sef under threshold models. Those values that were not statistically different from zero (Table 6). Slightly higher proportions of phenotypic variance were explained by the common maternal effects (including both genetic and environmental factors) estimated at 0.368 and 0.147, respectively for She and Sef (Table 6).

Table 6. Sources of variation of survival from eyed-egg stage until hatching (Seh) or 3 weeks after first feeding (Sef). Additive genetic, maternal, and residual variances (σ_a^2 , σ_m^2 , σ_r^2), corresponding heritability (h^2) and proportion of phenotypic variance explained by maternal effects (c^2) were estimated under threshold GBLUP models. Standard errors of estimates are given in brackets.

Survival	Seh	Sef
σ_a^2	0.452 (0.332)	0.188 (0.150)
σ_m^2	0.849 (0.759)	0.207 (0.182)
σ_r^2	1.008 (0.090)	1.010 (0.097)
h^2	0.196 (0.116)	0.134 (0.087)
c^2	0.368 (0.156)	0.147 (0.075)

Under linear AM models (Table A.4), early survival variables had very low heritability values of 0.015 and 0.012, respectively for survival from eyed-egg stage to hatching stage and to the end of the experiment. Those values were lower values than those estimated under threshold model and transformed to the observed scale by Dempster and Lerner (1950) formula (0.035 and 0.042 for She and Sef, respectively). However in any of the alternative (observed or underlying) worlds, there was not significant additive genetic variance on early survival traits. Proportions of phenotypic variance explained by the maternal effects were estimated at significant values, 5.8 % and 3.4% under linear models for She and Sef, respectively. Those values were quite consistent with values estimated under threshold model and transformed to the observed scale by Dempster and Lerner (1950) formula (0.066 and 0.046 for She and Sef, respectively).

There was no significant association between early survival traits and neither parental inbreeding coefficients (Table A.5) nor offspring inbreeding (Table A.6).

3.3.2. Maternal phenotypes and their associations with offspring survival rates

Regarding female traits, two size and six reproduction traits were measured, and basic statistics are presented in Table 7. As expected from mathematical constraints for CV (Lande, 1977), the CV of the average egg weight (corresponding to a volume record) was roughly 3 times larger than the CV of egg diameter (corresponding to a length record). However, CV of both egg size traits were low ($CV < 10\%$). Concerning the three other reproductive traits (CF, EN, and SW) and the PW, there were variables between dams with CV of around 20%.

Table 7. Summary of size and reproduction traits of the 15 dams at 3 years of age.

	Mean	SD	Minimum	Maximum	Coefficient of variation (%)
Fork length (FL, mm)	488	24.9	437	522	5.10
Post-spawning weight (PW, g)	2053	342	1382	2753	16.65
Spawn weight (SW, g)	267	55.4	178	406	20.73
Egg number (EN, #)	3765	831	2670	5908	22.06
Relative fecundity (RF, egg number/kg)	1.87	0.45	0.99	2.67	23.88
Weight of Coelomic fluid (CF, g)	27.8	5.2	19.2	35.2	18.55
Egg weight (EW, mg)	71.3	5.1	63.3	80.4	7.15
Egg diameter (ED, mm)	5.34	0.12	5.11	5.52	2.34

Table 8. Phenotypic correlations between size and reproduction traits of the 15 dams at 3 years of age. p-values figure between brackets.

	PW	SW	EN	RF	CF	EW	ED	Fdam
FL	0.689 **	0.516 *	0.362	-0.13	-0.087	0.429	0.511 *	-0.263
PW		0.207	0.047	-0.55 *	-0.012	0.512 *	0.535 *	-0.385
SW			0.952 ***	0.64 **	0.202	-0.022	0.135	-0.198
EN				0.79 ***	0.181	-0.322	-0.139	-0.217
RF					0.14	-0.64 **	-0.49 *	0.03
CF						-0.023	-0.132	0.094
EW							0.910 ***	0.053
ED								-0.118

FL: Female fork length; PW: Post-spawning weight; SW: Spawning weight; EN: egg number; RF: Relative fecundity; CF: Weight of coelomic fluid; EW: Average egg weight; ED: Egg diameter *** p-value < 0.001 ** p-value < 0.01 * p-value < 0.1

In addition, a phenotypic correlation between FL and PW was estimated at 0.69; EW and ED at 0.91; SW and EN at 0.95. The reproductive traits SW, ED and EW were positively correlated with FL or PW (Table 8). Negative but non-significant trends were generally observed for correlations between F_{dam} and maternal traits.

A graphical observation was first performed to ensure that the relationships between dam performances and progeny survival traits were globally linear before calculating the correlations between them.

Correlations of dam phenotypes with their progeny survival are presented in Table 9. Fork length was almost uncorrelated with survival until hatching, while PW was significantly and negatively correlated with survival until the hatching stage. After hatching, correlations between size traits and survival became weakly favorable. Note that the phenotypic correlation between average EW and ED was 0.91. Fecundity traits EN and SW were positively correlated with fry survival traits until hatching but almost uncorrelated with survival from hatching to 3 weeks after the first meal. Weak correlations that were not statistically different from zero were observed between all early survival traits and CF, EW, or ED. However it should be noted that offspring mortality observed from the eyed egg stage to hatching was significantly (p -value=0.003) higher in the dam group with larger eggs (i.e., 8 dams with $EW \geq 73.5$ mg) than in the dam group with lower average EW. In addition, all the 10 abnormal eggs observed at hatching were bred by 5 dams from this dam group with larger eggs.

Table 9. Pearson correlations between dam performances and progeny survival rates from eyed egg stage to 3 weeks after the first feeding.

	Se	Sh	Sf	Seh	Shf
FL	-0.07	-0.12	-0.03	-0.15	0.38
PW	-0.45 *	-0.54 *	-0.46 *	-0.49 *	0.36
SW	0.50 *	0.54 *	0.58 *	0.40	0.11
EN	0.50 *	0.61 *	0.66 **	0.53 *	0.18
RF	0.62 **	0.78 ***	0.78 ***	0.72 ***	-0.05
CF	0.21	0.20	0.11	0.12	-0.34
EW	-0.09	-0.31	-0.38	-0.50 *	-0.27
ED	-0.16	-0.24	-0.29	-0.27	-0.18

FL: Female fork length; PW: Post-spawning weight; SW: Spawning weight; EN: egg number; RF: relative fecundity; CF: Weight of coelomic fluid; EW: Average egg weight; ED: Egg diameter; Se: survival from fertilization to eyed egg stage; Sh: survival from fertilization to hatching; Sf: survival from fertilization to three weeks after first feeding; Seh: survival from eyed egg stage to hatching; Shf: survival from hatching to three weeks after first feeding.

*** p -value < 0.001 ** p -value < 0.01 * p -value < 0.1

4. Discussion

4.1. Evidence for variation in reproductive success in farmed trout

Family variance in reproductive success was identified in our study from fertilization to three weeks after the first feeding. It was mainly due to the stage between fertilization and hatching (Sh) because no family differences in progeny survival were observed from hatching to three weeks after the first feeding (Shf) (Table 3 and 4). Only dam contributions to the next generation were significantly unbalanced. No significant unbalanced contribution was observed for sires, probably because our mating plan was designed to avoid sperm competition, which is known to induce variance in family size (Gage et al., 2004; Wedekind et al., 2007). One-by-one male fecundation of dam egg pools is a usual practice in trout breeding programs. Even though the differences in offspring survival between extreme sires were insignificant, they were 11 and 12% at eyed egg stage in the two replicates. Such differences might have been significant should we have used a higher number of sires and/or more offspring per sire. These results confirm previous studies on early progeny survival in rainbow trout, for which maternal effects are always detected, but sire effects are low (Patton et al., 2007) or absent (Nagler et al., 2000). In the study by Patton et al. (2007), the sire effect may be at least partly confounded with a dominance effect because dams were mated with a single male. In chum salmon, significant sire and dam effects on early survival traits were identified by Smoker (1986). Similar results to ours were observed *in brown trout*, *i.e.*, important maternal effects on offspring survival, but no sire effects when sperm competition was avoided (Régnier et al., 2010; Vandeputte et al., 2002).

As previously found by Rye et al. (1990), we did not observe significant phenotypic correlations between offspring survival at the different stages of larval development (Table 5, Figure 4), because these various traits should be influenced by different determinants.

4.2. Importance of the genetic factors explaining early survival in a cohort of farmed trout fry

We did not observe any significant additive genetic variance on early survival traits. Estimated heritability for survival traits (at hatching or 3 weeks after first feeding) was low (respectively 0.20 and 0.13 under threshold models, but only 0.01 under linear models) in our study. Low to moderate heritability values (ranging from 0 to 0.29 taking into account the binary nature of survival traits) were also reported for early survival in different species of salmonids

(*Oncorhynchus mykiss* and *Salmo salar*: Kanis et al., 1976; Gall and Gross, 1978; Rye et al., 1990; Houde et al., 2013, 2015; Vehviläinen et al., 2008; *Salvelinus fontinalis*: Robison and Luempert, 1984; *Arctic char*: Palaiokostas et al., 2020). However there was high uncertainty on most of these previous estimates and heritability may have been overestimated because maternal and/or dominance effects were not well accounted for in the models. Duenk et al. (2017) have demonstrated that including dominance effects was always equally or more accurate than the simple additive genetic model, even with a small sample size or in the absence of dominance.

In our study, heritability of survival did not significantly change between hatching and 3 weeks after the first feeding. The low mortality levels (4.1 % for Seh and 7.2% for Sef) may limit the detection of genetic variation in early survival and may lead to low estimates of heritability in our study. The additive genetic coefficients of variation of early survival from eyed egg stage to hatching or to three weeks after first feeding, measuring their evolvability (Houle, 1992) were also very low, 0.74 %, and 1.31%, respectively. This means that selection may not be efficient in improving survival from the eyed-egg stage to hatching or 3 weeks after first feeding. However, our experiment was based on a unique cohort with progeny bred from only 15 dams and 9 sires that were randomly sampled from a population with about 10 times more breeders. Uncertainty on our parameter estimates is a consequence of this small sampling set and further analysis on subsequent cohorts should confirmed our first estimates.

Under linear models (Table A.4), extremely high dominance deviation effects were estimated in our study. These results have to be taken with caution as phenotypic variances were low and only a small dataset was used to fit complex genetic models. However, the full AMD model including dominance deviation effects gave the better goodness-of-fit under linear model (minimum AIC value, Table A.4), while the threshold AMD model did not converged. Our high estimates of dominance deviation variance for early survival traits is consistent with the high non-additive genetic variances observed in Atlantic salmon (Houde et al., 2013, 2015) for survival from fertilization to hatching and from hatching to yolk sac resorption. High dominance variance is theoretically linked to inbreeding depression, that has been shown to play a significant role in female reproduction traits in a selected rainbow trout line, even if no genome-wide inbreeding effect was observed (Paul et al., 2021). Indeed, many studies showed inbreeding depression on fitness-related traits due to high levels of parental and own individual inbreeding (*Oncorhynchus mykiss*: Aulstad et al., 1972; Kincaid et al., 1976a; Gjerde et al., 1983; Naish

et al., 2013; *Oreochromis niloticus*: Fessehaye et al., 2009; *Salvelinus alpinus*: Palaiokostas et al., 2021). In our study, we did not estimate any significant effect of dam or sire inbreeding levels on offspring survival neither at hatching nor at the end-point of the experiment. However, Su et al. (1996) revealed that dam inbreeding impacts egg quality and may reduce progeny survival in rainbow trout. Fessehaye et al. (2009) demonstrated that sire inbreeding could influence reproductive success in Nile tilapia, with an inbreeding depression effect observed for the number of progeny when males are in competition (which was not the case in our study). In addition, we did not observe any significant difference in the numbers of dead fry among offspring with inbreeding levels over the average (Table A.6). On the contrary, Gjerde et al. (1983) and Aulstad et al. (1972) estimated inbreeding depression effects around 10% on survival from eyed egg stage to hatching in rainbow trout populations. Gjerde et al. (1983) also indicated inbreeding depression effects ranging from 5 to 11% on offspring survival from hatching to first feeding and from the first feeding to 6 weeks after the first meal. However, these estimations were based on an experimental line with very high inbreeding levels (from 0.25 to 0.50) due to full-sib mating. Kincaid et al. (1976a) showed a high inbreeding depression effect on early survival in a rainbow trout population with high inbreeding levels (from 0.25 to 0.375). In our study, parents were weakly related, and there was no mating between full-sibs. Therefore, the inbreeding levels of the offspring remained moderate, with an average value of 0.10 and a maximum value of 0.25 (Table 1), which may explain why we did not observe any significant inbreeding depression effects compared to previous studies. Even though we did not observe a significant effect of genome-wide inbreeding on offspring survival, it does not mean that inbreeding depression is not occurring in some local regions in the genome, as Paul et al. (2021) revealed for female reproduction traits in rainbow trout.

Similarly, Kincaid (1976b) showed no effect of inbreeding on the egg hatchability stage in a rainbow trout population but a higher inbreeding effect on fry survival in Kincaid (1976a). Houde et al. (2010) also observed no significant differences in offspring survival between inbred and non-inbred lines. In addition, Moss et al. (2007) highlighted that inbreeding has no effect on Pacific white shrimp survival in normal conditions, but a severe inbreeding depression on survival appears after exposure to viral pathogens. Their study proves that inbreeding may play a major role when life conditions change to extreme ones. In consequence, even if we did not identify any inbreeding depression effect on early survival in rainbow trout in our experimental conditions, we cannot exclude that inbreeding effects may reveal in less favourable rearing conditions.

4.3. Importance of maternal phenotypes on fry survival

Compared to our estimates based on animal models with maternal effects, higher proportion of phenotypic variances of egg, alevin or fry survival in salmonids species were explained by maternal effects under sire-dam models (Kanis et al., 1976; Rye et al., 1990; Houde et al., 2013, 2015). These previous results showed that heritability estimates derived from the dam components of variance were substantially higher (between 0.15 to 0.58 from eyed-egg stage to hatching and between 0.07 to 0.86 from hatching to first feeding) than those derived from the sire components, revealing that additive genetic effects were small, but non-additive genetic effects and/or maternal environment effects explained a high proportion of early survival in salmonids. However experimental designs and statistical models used in these previous studies did not allow to disentangle additive genetic, dominance deviation, maternal genetic and maternal environmental effects from the dam variance components.

In an attempt to understand which biological factors may underlie the maternal effects observed on early survival, we estimated Pearson correlations between various dam performances and offspring survival rates (Table 9). Because there is information on only 15 dams, most correlations are not statistically different from zero. This limits the interpretation of the results; however, some challenging results deserve discussion.

Whatever the stage, favourable correlations were estimated between offspring survival and fecundity traits (SW, EN, and RF). This is consistent with previous results on rainbow trout (Estay et al., 2021), showing a significant positive association between relative fecundity (RF) and progeny survival at the eyed egg stage. However, some weak negative trends between fecundity traits and survival stages (at the eyed egg, hatching, and swim up) were identified in another cultured stock rainbow trout (Kanyilmaz et al., 2016). In lake whitefish, survival until hatching and fecundity traits were not correlated (Wedekin and Müller, 2004).

In our study, as in many others in rainbow trout (Su et al., 1996; Gall and Neira, 2004; Solberg et al., 2014; Islam et al., 2021; Estay et al., 2021), results support the evidence that larger females produce larger eggs (higher EW average) than smaller females. Our results, showing no significant association between egg size and survival contradict those of many studies (Bagenal, 1969; Wallace & Aajord, 1984; Beacham & Murray, 1985; Rana, 1985; Marteinsdottir & Able, 1992) that showed that the survival of fry derived from larger eggs was significantly better, probably because these fries would be bigger. Here we even observed an

almost significant negative correlation between egg weight and survival from eyed egg stage to hatching (Table 9, p-value = 0.06). An explanation for this excess in egg mortality may be the over-ripening of the larger eggs as all observed abnormal embryos at hatching came from the dams with the heaviest average egg weights. Indeed, some females may have reached maturity a few days before the others, inducing greater hydration and, therefore, larger eggs without their over-ripening being visible but potentially inducing developmental defects in the offspring (*Salmo salar*: Horreo et al., 2008; *Leuciscus idus*: Nowosad et al., 2018).

Similarly to our observation that dams with heavier eggs had more malformed progenies, Wallace and Aasjord (1984) observed that progeny groups with higher egg size in Arctic charr have higher fry mortality due to malformations. These observations deserve further attention by salmonid breeders as increasing egg size was initially thought to be an interesting way to increase alevin survival. Some studies demonstrate that the quantity and quality (such as lipid, protein composition, and triiodothyronine concentration) of egg reserves are crucial for fish offspring early survival (Brown et al., 1988; Greenblatt et al., 1989; Gisbert et al., 2000; Stuart et al., 2020). According to Stuart et al. (2020), the egg quality (lipid and protein) is increased with egg size because of a higher metabolic reserve (Brooks et al., 1997). Besides, Berg et al. (2001) and Régnier et al. (2013) demonstrated that small eggs are not always those of lesser quality because small eggs allow, for example, to have a better protein content than larger eggs. Even if a larger egg produces a larger fry, this advantage can be masked by other environmental determinants of growth (Bromage et al., 1992), and egg size does not appear to be an essential indicator of egg quality in farmed rainbow trout populations (Brooks et al., 1997). In addition, Einum and Fleming (1999) observed that differences in offspring's survival from small and large eggs decreased with increasing environmental quality. So, in high-quality environments (facilities, farmed environments), no egg size effect should be observed on offspring's survival. To finish, egg size was uncorrelated with progeny survival in various studies and fish species (*Oncorhynchus mykiss*: Springate and Bromage, 1985; Herbinger et al., 1995; Blanc, 2007; *Salmo trutta fario* L.: Vandeputte et al., 2002; *Cyprinus carpio* L.: Vandeputte et al., 2004; *Salvelinus alpinus*: Jónsson, B., & Svavarsson, 2000; Leblanc et al., 2016).

5. Conclusion

This study highlights the biological factors underlying early survival and identifies maternal disequilibrium contributions to the next generation in a domesticated rainbow trout line. Heritability of early survival is low, not statistically different from zero, but maternal effects explain a larger amount of the phenotypic variance of embryo survival. While egg size has no significant effects on early survival, dam weight and fecundity seem to play important roles underlying the maternal effects observed on embryo and fry survival. Further studies are needed to well understand these roles and to confirm the existence of non-additive genetic factors impacting early survival in rainbow trout.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Jerome Bugeon from INRAE (LPGP) for his help in retrieving phenotypic data (egg diameter) and Marc Vandeputte from INRAE (USC Marbec) for very valuable comments to improve the manuscript.

Funding sources

This work was supported by INRAE and the French Ministry of Higher Education, Research and Innovation.

References

- Aulstad, D., Gjedrem, T., and Skjervold, H., 1972. Genetic and environmental sources of variation in length and weight of rainbow trout (*Salmo gairdneri*). *Journal of the Fisheries Board of Canada*, 29(3), 237-241. <https://doi.org/10.1139/f72-042>
- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., & Misztal, I. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution*, 51(1), 1-8. <https://doi.org/10.1186/s12711-019-0469-3>
- Bagenal, T. B., 1969. Relationship between egg size and fry survival in brown trout *Salmo trutta* L. *Journal of Fish Biology*, 1(4), 349-353. <https://doi.org/10.1111/j.1095-8649.1969.tb03882.x>
- Beacham, T. D., & Murray, C. B. (1985). Effect of female size, egg size, and water temperature on developmental biology of chum salmon (*Oncorhynchus keta*) from the Nitinat River, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, 42(11), 1755-1765.
- Benjamini Y, Yekutieli D., 2001. The control of false discovery rate under dependency. *Ann Stat*, 29:1165–1188.
- Berg, O. K., Hendry, A. P., Svendsen, B., Bech, C., Arnekleiv, J. V., and Lohrmann, A., 2001. Maternal provisioning of offspring and the use of those resources during ontogeny: variation within and between Atlantic salmon families. *Functional Ecology*, 15(1), 13-23. <https://doi.org/10.1046/j.1365-2435.2001.00473.x>
- Blanc, J. M., 2007. Effects of egg size differences on juvenile weight between and within lots in rainbow trout *Oncorhynchus mykiss*. *Journal of the World Aquaculture Society*, 33(3), 278-286. <https://doi.org/10.1111/j.1749-7345.2002.tb00504.x>
- Bromage, N., Jones, J., Randall, C., Thrush, M., Davies, B., Springate, J., Duston, J., and Barker, G., 1992. Broodstock management, fecundity, egg quality and the timing of egg production in the rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 100(1-3), 141-166. [https://doi.org/10.1016/0044-8486\(92\)90355-O](https://doi.org/10.1016/0044-8486(92)90355-O)
- Brooks, S., Tyler, C. R., and Sumpter, J. P., 1997. Egg quality in fish: what makes a good egg? *Reviews in Fish Biology and fisheries*, 7(4), 387-416. <https://doi.org/10.1023/A:1018400130692>
- Brown, C. L., Doroshov, S. I., Nunez, J. M., Hadley, C., Vaneennaam, J., Nishioka, R. S., and Bern, H. A., 1988. Maternal triiodothyronine injections cause increases in swimbladder inflation and survival rates in larval striped bass, *Morone saxatilis*. *Journal of Experimental Zoology*, 248(2), 168-176. <https://doi.org/10.1002/jez.1402480207>
- Bry, C., 1981. Temporal aspects of macroscopic changes in rainbow trout (*Salmo gairdneri*) oocytes before ovulation and of ova fertility during the post-ovulation period; effect of treatment with 17 α -hydroxy-20 β -dihydroprogesterone. *Aquaculture*, 24, 153-160. [https://doi.org/10.1016/0044-8486\(81\)90052-1](https://doi.org/10.1016/0044-8486(81)90052-1)
- Cardona, E., Bugeon, J., Segret, E., and Bobe, J., 2020. VisEgg: A robust phenotyping tool to assess rainbow trout egg features and viability. *Fish Physiology and Biochemistry*, 1-9. <https://doi.org/10.1007/s10695-020-00844-2>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742-015. <https://doi.org/10.1186/s13742-015-0047-8>
- Cheverud, J.M., Rutledge, J.J., and Atchley, W.R. 1983. Quantitative genetics of development: genetic

correlations among age-specific trait values and the evolution of ontogeny. *Evolution*, 37(5): 895–905. doi:10.2307/2408405.

Christie, M. R., McNickle, G. G., French, R. A., and Blouin, M. S. 2018. Life history variation is maintained by fitness trade-offs and negative frequency-dependent selection. *Proceedings of the National Academy of Sciences*, 115(17), 4441-4446. <https://doi.org/10.1073/pnas.1801779115>

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., and Dupont-Nivet, M., 2019. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51(1), 1-15. <https://doi.org/10.1186/s12711-019-0468-4>

D'Ambrosio, J., Morvezen, R., Brard-Fudulea, S., Bestin, A., Acin Perez, A., Guéméné, D., ... and

Phocas, F. (2020). Genetic architecture and genomic selection of female reproduction traits in rainbow trout. *BMC genomics*, 21(1), 1-14. <https://doi.org/10.1186/s12864-020-06955-7>

Dempster, E. R., and Lerner, I. M. 1950. Heritability of threshold characters. *Genetics*, 35(2), 212.

Duenk P, Calus MPL, Wientjes YCJ, Bijma P. 2017. Benefits of dominance over additive models for the estimation of average effects in the presence of dominance. *G3* 7:3405–14.

Einum, S., and Fleming, I. A. 1999. Maternal effects of egg size in brown trout (*Salmo trutta*): norms of reaction to environmental quality. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1433), 2095-2100. <https://doi.org/10.1098/rspb.1999.0893>

Estay, F. J., Colihueque, N., and Yáñez, M., 2021. Reproductive performance assessed during three spawning seasons in a naturalized rainbow trout population from southern Chile. *Fisheries Research*, 244, 106107. <https://doi.org/10.1016/j.fishres.2021.106107>

Fessehaye, Y., Bovenhuis, H., Rezk, M. A., Crooijmans, R., van Arendonk, J. A., and Komen, H., 2009. Effects of relatedness and inbreeding on reproductive success of Nile tilapia (*Oreochromis niloticus*). *Aquaculture*, 294(3-4), 180-186. <https://doi.org/10.1016/j.aquaculture.2009.06.001>

Fong, Y., Huang, Y., Gilbert, P. B., and Permar, S. R. 2017. chngpt: Threshold regression model estimation and inference. *BMC bioinformatics*, 18(1), 1-7. <https://doi.org/10.1186/s12859-017-1863-x>

Gage, M. J., Macfarlane, C. P., Yeates, S., Ward, R. G., Searle, J. B., and Parker, G. A., 2004. Spermatozoal traits and sperm competition in Atlantic salmon: relative sperm velocity is the primary determinant of fertilization success. *Current biology*, 14(1), 44-47. <https://doi.org/10.1016/j.cub.2003.12.028>

Gall, G.A.E., & Gross, S.J., 1978. A genetic analysis of the performance of three rainbow trout broodstocks. *Aquaculture*, 15(2), 113-127. [https://doi.org/10.1016/0044-8486\(78\)90057-1](https://doi.org/10.1016/0044-8486(78)90057-1)

Gall, G. A., & Neira, R., 2004. Genetic analysis of female reproduction traits of farmed coho salmon (*Oncorhynchus kisutch*). *Aquaculture*, 234(1-4), 143-154. <https://doi.org/10.1016/j.aquaculture.2004.01.029>

Gharrett, A. J., & Shirley, S. M. 1985. A genetic examination of spawning methodology in a salmon hatchery. *Aquaculture*, 47(2-3), 245-256. [https://doi.org/10.1016/0044-8486\(85\)90071-7](https://doi.org/10.1016/0044-8486(85)90071-7)

Gisbert, E., Williot, P., and Castello-Orvay, F., 2000. Influence of egg size on growth and survival of early stages of Siberian sturgeon (*Acipenser baeri*) under small scale hatchery conditions. *Aquaculture*, 183(1-2), 83-94. [https://doi.org/10.1016/S0044-8486\(99\)00287-2](https://doi.org/10.1016/S0044-8486(99)00287-2)

Gjerde, B., Gunnes, K., and Gjedrem, T., 1983. Effect of inbreeding on survival and growth in rainbow trout. *Aquaculture*, 34(3-4), 327-332. [https://doi.org/10.1016/0044-8486\(83\)90212-0](https://doi.org/10.1016/0044-8486(83)90212-0)

Greenblatt, M., Brown, C. L., Lee, M., Dauder, S., and Bern, H. A., 1989. Changes in thyroid hormone levels in eggs and larvae and in iodide uptake by eggs of coho and chinook salmon, *Oncorhynchus kisutch* and *O. tshawytscha*. *Fish Physiology and Biochemistry*, 6(5), 261-278.

Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., and Vandeputte, M., 2020. APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Molecular ecology resources*, 20(2), 579-590.

Herbinger, C. M., Doyle, R. W., Pitman, E. R., Paquet, D., Mesa, K. A., Morris, D. B., ... and Cook, D., 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture*, 137(1-4), 245-256. [https://doi.org/10.1016/0044-8486\(95\)01109-9](https://doi.org/10.1016/0044-8486(95)01109-9)

Houde, A. L. S., Fraser, D. J., O'Reilly, P., and Hutchings, J. A. 2011. Maternal and paternal effects on fitness correlates in outbred and inbred Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, 68(3), 534-549. <https://doi.org/10.1139/f2011-00>

Houde, A. L., Wilson, C. C., and Neff, B. D. 2013. Genetic architecture of survival and fitness-related traits in two populations of Atlantic salmon. *Heredity*, 111(6), 513-519. <https://doi.org/10.1038/hdy.2013.74>

Houde, A. L. S., Black, C. A., Wilson, C. C., Pitcher, T. E., and Neff, B. D. 2015. Genetic and maternal effects on juvenile survival and fitness-related traits in three populations of Atlantic salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(5), 751-758. <https://doi.org/10.1139/cjfas-2014-0472>

Houle, D., 1992. Comparing evolvability of quantitative traits. *Genetics*, 130, 195-204.

Islam, S. S., Wringe, B. F., Bøe, K., Bradbury, I. R., and Fleming, I. A. 2021. Early-life fitness trait variation among divergent European and North American farmed and Newfoundland wild Atlantic salmon populations. *Aquaculture Environment Interactions*, 13, 323-337. <https://doi.org/10.3354/aei00412>

Jónsson, B., & Svavarsson, E. (2000). Connection between egg size and early mortality in arctic charr, *Salvelinus alpinus*. *Aquaculture*, 187(3-4), 315-317.

Kanis, E., Refstie, T., and Gjedrem, T., 1976. A genetic analysis of egg, alevin and fry mortality in salmon (*Salmo salar*), sea trout (*Salmo trutta*) and rainbow trout (*Salmo gairdneri*). *Aquaculture*, 8(3), 259-268. [https://doi.org/10.1016/0044-8486\(76\)90088-0](https://doi.org/10.1016/0044-8486(76)90088-0)

Kanyılmaz, M., Yılayaz, A., Uysal, G., Sevgili, H., and Tabakoğlu, Ş. S., 2016. Hatchery performance in a major rainbow trout (*Oncorhynchus mykiss*, Walbaum, 1792) seed production area of Turkey. *Journal of Applied Ichthyology*, 32(4), 661-668. <https://doi.org/10.1111/jai.13063>

Kaspar, V., Vandeputte, M., Kohlmann, K., Hulak, M., Rodina, M., Gela, D., ... & Linhart, O. 2008. A proposal and case study towards a conceptual approach of validating sperm competition in common carp (*Cyprinus carpio* L.), with practical implications for hatchery procedures. *Journal of Applied Ichthyology*, 24(4), 406-409. <https://doi.org/10.1111/j.1439-0426.2008.01148.x>

Kincaid, H. L., 1976a. Inbreeding in rainbow trout (*Salmo gairdneri*). *Journal of the Fisheries Board of Canada*, 33(11), 2420-2426. <https://doi.org/10.1139/f76-288>

Kincaid, H. L., 1976b. Effects of inbreeding on rainbow trout populations. *Transactions of the American Fisheries Society*, 105(2), 273-280. <https://doi.org/10.1577/1548-8659>

Leblanc, C. A. L., Kristjánsson, B. K., and Skúlason, S., 2016. The importance of egg size and egg energy density for early size patterns and performance of Arctic charr *Salvelinus alpinus*. *Aquaculture Research*, 47(4), 1100-1111. <https://doi.org/10.1111/are.12566>

Lande, R., 1977. On comparing coefficients of variation. *Systematic Zoology*, 26(2), 214-217. <https://doi.org/10.2307/2412845>

Marteinsdottir, G., & Able, K. W. (1992). Influence of egg size on embryos and larvae of *Fundulus heteroclitus* (L.). *Journal of fish biology*, 41(6), 883-896.

Merilä, J., and Sheldon, B. C., 1999. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity*, 83(2), 103-109. <https://doi.org/10.1046/j.1365-2540.1999.00585.x>

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUP F90 family of programs. Available from http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Moss, D. R., Arce, S. M., Ootshi, C. A., Doyle, R. W., and Moss, S. M., 2007. Effects of inbreeding on survival and growth of Pacific white shrimp *Penaeus* (*Litopenaeus*) *vannamei*. *Aquaculture*, 272, S30-S37. <https://doi.org/10.1016/j.aquaculture.2007.08.014>

Nagler, J. J., Parsons, J. E., and Cloud, J. G., 2000. Single pair mating indicates maternal effects on embryo survival in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 184(1-2), 177-183. [https://doi.org/10.1016/S0044-8486\(99\)00309-9](https://doi.org/10.1016/S0044-8486(99)00309-9)

Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., and Quinn, T. P., 2013. Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. *Molecular Ecology*, 22(5), 1295-1309. <https://doi.org/10.1111/mec.12185>

Palaiokostas, C., Jeuthe, H., and De Koning, D. J., 2020. Assessing the potential of improving growth and survival to the eyed stage in selectively bred Arctic charr (*Salvelinus alpinus*). *Journal of Animal Breeding and Genetics*, 138(3), 326-337. <https://doi.org/10.1111/jbg.12509>

Palaiokostas, C., Anjum, A., Jeuthe, H., Kurta, K., Lopes Pinto, F., and de Koning, D. J., 2021. A genomic-based vision on the genetic diversity and key performance traits in selectively bred Arctic charr (*Salvelinus alpinus*). *Evolutionary Applications* <https://doi.org/10.1111/eva.13261>

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., Rexroad III, C. E., Moen, T., 2015. The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. *Molecular ecology resources*, 15(3), 662-672. <https://doi.org/10.1111/1755-0998.12337>

Patton, S. J., Kane, S. L., Wheeler, P. A., and Thorgaard, G. H., 2007. Maternal and paternal influence on early embryonic survival of androgenetic rainbow trout (*Oncorhynchus mykiss*): Implications for measuring egg quality. *Aquaculture*, 263(1-4), 26-34. <https://doi.org/10.1016/j.aquaculture.2006.12.005>

Paul, K., D'Ambrosio, J. and Phocas, F., 2021. Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout. *Evolutionary Applications*. <https://doi.org/10.1111/eva.13308>

Pilastro, A., Scaggiante, M., and Rasotto, M. B., 2002. Individual adjustment of sperm expenditure accords with sperm competition theory. *Proceedings of the National Academy of Sciences*, 99(15), 9913-9915.

<https://doi.org/10.1073/pnas.152133499>

Phocas, F., Belloc, C., Bidanel, J., Delaby, L., Dourmad, J. Y., Dumont, B., ... and Brochard, M., 2016. Towards the agroecological management of ruminants, pigs and poultry through the development of sustainable breeding programs: I-selection goals and criteria. *Animal*, 10(11), 1749-1759. <https://doi.org/10.1017/S1751731116000926>

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.

Rana, K. J. (1985). Influence of egg size on the growth, onset of feeding, point-of-no-return, and survival of unfed *Oreochromis mossambicus* fry. *Aquaculture*, 46(2), 119-131.

Régnier, T., Bolliet, V., Labonne, J., and Gaudin, P., 2010. Assessing maternal effects on metabolic rate dynamics along early development in brown trout (*Salmo trutta*): an individual-based approach. *Journal of Comparative Physiology B*, 180(1), 25-31. <https://doi.org/10.1007/s00360-009-0385-x>

Régnier, T., Bolliet, V., Gaudin, P., and Labonne, J., 2013. Bigger is not always better: egg size influences survival throughout incubation in brown trout (*Salmo trutta*). *Ecology of Freshwater Fish*, 22(2), 169-177. <https://doi.org/10.1111/eff.12018>

Robison, O. W., and Luempert III, L. G., 1984. Genetic variation in weight and survival of brook trout (*Salvelinus fontinalis*). *Aquaculture*, 38(2), 155-170. [https://doi.org/10.1016/0044-8486\(84\)90227-8](https://doi.org/10.1016/0044-8486(84)90227-8)

Rye, M., Lillevik, K. M., and Gjerde, B., 1990. Survival in early life of Atlantic salmon and rainbow trout: estimates of heritabilities and genetic correlations. *Aquaculture*, 89(3-4), 209-216. [https://doi.org/10.1016/0044-8486\(90\)90126-8](https://doi.org/10.1016/0044-8486(90)90126-8)

Sakai, K., Nomura, M., Takasima, F., and Oto, H., 1975. The over-ripening phenomenon of rainbow trout, 2: Changes in the percentage of eyed eggs, hatching rate and incidence of abnormal alevins during the process of over-ripening. *Bulletin of the Japanese Society of Scientific Fisheries*.

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC genomics*, 15(1), 478. <https://doi.org/10.1186/1471-2164-15-478>

Smoker, W. W., 1986. Variability of embryo development rate, fry growth, and disease susceptibility in hatchery stocks of chum salmon. *Aquaculture*, 57(1-4), 219-226. [https://doi.org/10.1016/0044-8486\(86\)90200-0](https://doi.org/10.1016/0044-8486(86)90200-0)

Solberg, M. F., Fjellidal, P. G., Nilsen, F., and Glover, K. A. 2014. Hatching time and alevin growth prior to the onset of exogenous feeding in farmed, wild and hybrid Norwegian Atlantic salmon. *PLoS One*, 9(12), e113697. <https://doi.org/10.1371/journal.pone.0113697>

Springate, J. R. C., and Bromage, N. R., 1985. Effects of egg size on early growth and survival in rainbow trout (*Salmo gairdneri* Richardson). *Aquaculture*, 47(2-3), 163-172. [https://doi.org/10.1016/0044-8486\(85\)90062-6](https://doi.org/10.1016/0044-8486(85)90062-6)

Stuart, K. R., Armbruster, L., Johnson, R., and Drawbridge, M. A., 2020. Egg diameter as a predictor for egg quality of California yellowtail (*Seriola dorsalis*). *Aquaculture*, 522, 735154. <https://doi.org/10.1016/j.aquaculture.2020.735154>

Su, G. S., Liljedahl, L. E., and Gall, G. A., 1996. Effects of inbreeding on growth and reproductive traits

in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 142(3-4), 139-148. [https://doi.org/10.1016/0044-8486\(96\)01255-0](https://doi.org/10.1016/0044-8486(96)01255-0)

Thompson, R., Brotherstone, S., and White, I. M., 2005. Estimation of quantitative genetic parameters. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459), 1469-1477. <https://doi.org/10.1098/rstb.2005.1676>

Tsuruta, S., & Misztal, I. (2006). THRGIBBS1F90 for estimation of variance components with threshold and linear models. *Threshold*, 3(4).

Vandeputte, M., Quillet, E., and Chevassus, B., 2002. Early development and survival in brown trout (*Salmo trutta fario* L.): indirect effects of selection for growth rate and estimation of genetic parameters. *Aquaculture*, 204(3-4), 435-445. [https://doi.org/10.1016/S0044-8486\(01\)00829-8](https://doi.org/10.1016/S0044-8486(01)00829-8)

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Rodina, M., ... & Linhart, O. 2004. Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture*, 235(1-4), 223-236. <https://doi.org/10.1016/j.aquaculture.2003.12.019>

Vehviläinen, H., Kause, A., Quinton, C., Koskinen, H., and Paananen, T., 2008. Survival of the currently fittest: genetics of rainbow trout survival across time and space. *Genetics*, 180(1), 507-516. <https://doi.org/10.1534/genetics.108.089896>

Vehviläinen, H., Kause, A., Koskinen, H., and Paananen, T., 2010. Genetic architecture of rainbow trout survival from egg to adult. *Genetics research*, 92(1), 1-11. <https://doi.org/10.1017/S0016672310000017>

Vehviläinen, H., Kause, A., Kuukka-Anttila, H., Koskinen, H., and Paananen, T., 2012. Untangling the positive genetic correlation between rainbow trout growth and survival. *Evolutionary applications*, 5(7), 732-745. <https://doi.org/10.1111/j.1752-4571.2012.00251.x>

Vitezica Z G, Varona L, and Legarra A, 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195, 1223–1230.

Wallace, J. C., & Aasjord, D. (1984). An investigation of the consequences of egg size for the culture of Arctic charr, *Salvelinus alpinus* (L.). *Journal of Fish Biology*, 24(4), 427-435.

Wedekind, C., Rudolfsen, G., Jacob, A., Urbach, D., and Müller, R., 2007. The genetic consequences of hatchery-induced sperm competition in a salmonid. *Biological conservation*, 137(2), 180-188. <https://doi.org/10.1016/j.biocon.2007.01.025>

Wedell, N., Gage, M. J., and Parker, G. A., 2002. Sperm competition, male prudence and sperm-limited females. *Trends in ecology and evolution*, 17(7), 313-320. [https://doi.org/10.1016/S0169-5347\(02\)02533-8](https://doi.org/10.1016/S0169-5347(02)02533-8)

Supplementary files

Table A.1. Preparation of ActiFish solution

<i>Actifish</i>	
5 (L)	demineralized water
30.0 (g)	NaCl
9.5 (g)	Glycine or Gycocolle
5 (g)	Tris (hydroxymethyl) methylamine
0.75 (g)	CaCl
0.1 (g)	Caffeine

Figure A.1. Incubators position in the trough.

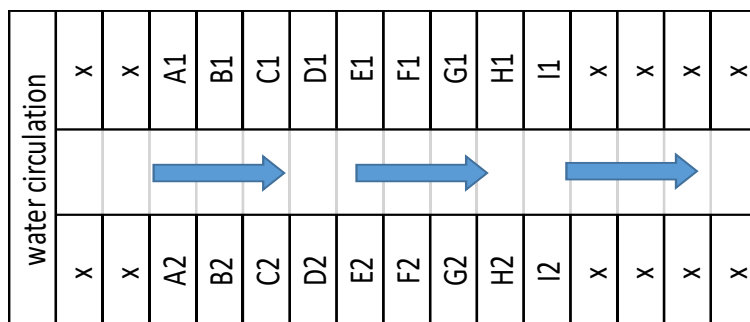


Table A2. 96 SNP-panel used for offsprings assignation. Two SNPs were eliminated due to their lower genotyping quality : call rate = 0% (SNP number 47) and lack of heterozygotes (SNP number 96).

SNP_number	ID-57K-AX	ID-57K-Affx	CHR_SW	POS_SW	Upstream sequence	SNP	Downstream sequence
1	AX-89965686	Affx-88948630	1	27969361	GAGCTTTCATTTTCAGTTTTCCACACATAAATTTGCCACTCCCTTAGGC	[T/C]	TTGCAAGAACAGTTGACATCGTCTGGTGCAGAGAGATTGGACATCTCATCGA
2	AX-89967592	Affx-88910086	1	45426369	GATGGGTGACACATCTATTTAATCCCAATTTGAATTCAGGTGTAACACA	[G/A]	TAAGTCAAAAGTATGAGTACTCTCTCTCAGGACTGTACTGATCACTTG
3	AX-89919880	Affx-88933991	1	55822124	CGAACTTGGAGCGTTATTAACCGCTTCGTTTCATGATCGACACTAGACT	[A/C]	TTTTCTGTCAACATGCATTTTGCTCTCGACAAACGTTCCITACACAACAT
4	AX-89952191	Affx-88916911	1	59254579	AAGGCAAATAACCATATTGCACAGTAGGAAAGTAAACCCGCAGGTAGCTTG	[T/C]	AAATGGCAGATCTAGCGAGTGTGGAAATCGACTACACTGCAGTAGGACT
5	AX-89933495	Affx-88950822	1	64632546	TCGTGGTCCACAGCCGGAAACCCCTCTATGACTCCCTGTAACCTCAACT	[C/T]	CTGTCCCTACCCAGAACCCCTCTCACATTTGACCAGTGTCTGCACC
6	AX-89930502	Affx-88960123	10	13938884	TCTTGACAGATTCTGACATGGCAAATTTGGTGCTACATTTTAGATGATA	[A/G]	TGAGTTCGTTGCAGATGTGTACTGGTAGCCGAAGTCAAGTGCAGGAGAGCA
7	AX-89943723	Affx-88920298	10	31045123	GTCATTTGACTGACAGAAAGGGCTATGTGTGGACTTTGCATGATTAAGG	[A/G]	GAACCCCTATTGGAACATAATCTGAGCAGATGTATGCTTATGATAGT
8	AX-89965431	Affx-88940627	10	40065914	AAGAAAAAGAGACCGGAGTCAAGTTAAGAAAGTTAATGTTTTGCTTATGT	[A/G]	GCAATGTATACAAACCATACATCAGGGTACTTGAACCTGACCAATTAAGT
9	AX-89938481	Affx-88935080	10	48846652	TGATGTTCTTGTTCCTGTAACAGAGAAACAGACCCCTACACCAAGCTCC	[A/C]	ATAATGTACTGAGAGAGACCTCCAGCTTCTTACAGCTGGCCAGGGCTG
10	AX-89940691	Affx-88926753	11	25220069	GCATATGCCTCCCTTAAGGTAGCATGTCACCTCTTTGGCGCCAAAAGA	[C/T]	GATGTTGTATTGTTACGTTGTTGAGATAACCAGAGCCTCCGTAGGGA
11	AX-89935160	Affx-88911487	11	48406831	AATATCAGGAGTGCACGTCAAGCCCGAACCCGCATTAAGCAGAAGCA	[C/T]	CCTTATGTGTAACTGGTTAATGTTCTGATTACACAGAAGTGTCCAAAA
12	AX-89950043	Affx-88920824	11	76201460	GAAGCCGATATGCACAGTGTATCAAGAAGATGGCGTCCCTCTCAGTGC	[A/G]	GATCATATGTTATCCAGCTTGAGTGAACATGAGAAAAAACAAGTTGTT
13	AX-89921652	Affx-88930319	12	17467798	AGCTAATTTAATCTCACACACACTCTCCTAGGATAGGATCATAATGCCT	[C/A]	ATTTAATCTAAAGGTTATATATGTTAGGATACCTCTGCAGTGTACTGAC
14	AX-89925972	Affx-88923291	12	28539219	TCACTCCAGATGACATAAATCCGATTTGATTTAGCTCAGGAAAATAAA	[C/A]	AAGTTTTATGAGAATGAAATCCAAAGAAACCAACCAATGCCATATCAGAAA
15	AX-89935184	Affx-88960741	12	42975846	CAACAATTTGTAGAATAACATTTAGACTTTAGAGACTTCACAGCAGA	[C/A]	ATAGTTTTCTGTACATTTAGAGAAGCAGTATACACGCAACAGAACAAAGT
16	AX-89922610	Affx-88957579	12	58724140	ATCCCACTTTTACCAGAAAAAGTATCCATGCTCTTTCATATGTTTTT	[C/A]	GCTAGCTTATATAAAATACATGTCTGTTTATATAATAATTAAGCTGC
17	AX-89941956	Affx-88925294	13	19254574	AGGATCAATTTGAAGTCTACTATTGTCGATTGCTCGGATCTCAGATG	[C/T]	GTAACATGCTTTTCTCTCTGTACTGACAAAATAGAGAAAGTAGTCG
18	AX-89940707	Affx-88943743	13	26266252	TCAAAATCATTGTAATTGGGTCACCTTAATAATTCTAAATGAAATTAAT	[C/A]	ATATCTAAAAGATGCATTCCAACCAATTCACCAAGAGCACCCTTAGATC
19	AX-89971103	Affx-88930116	13	38098191	GGACCCTGCTTTCAGACAGAAAGATCAATCATCAGAAAGTGACTTAAT	[A/C]	TTGGCTGCAGAAAGTCCAGGGAGAGGAGAGAGCAGAGATAAGTTGGAGA
20	AX-89948958	Affx-88931303	14	31988819	ATGTAATGTCAGTTTCTTATGTTAACAATTTTCGTCAGGTAATATCTTCA	[G/T]	AATTCAGAAAATAAAAAACAGAGAGCTTGGCACGCTGAGACAGACAGGAAG
21	AX-89958404	Affx-88941212	14	59909868	TGCCTTTAATATTGACAGTTGTTCCCATGGAAATTTTCTCATGGATGCC	[C/T]	TTAACTCTAATTTGAAAGCCTCGGGGAGAAATAGATTGAGCTCAGTGC
22	AX-89976637	Affx-88908613	14	75812323	AAGGAAAGCAATTCATTAGGTAAGAGAAGGCATTTGCACAAAGTTGACC	[A/G]	CTGGTATTGTTGAGCATGTTTCTTACCTTTCCTTACTCATTCTATTTTGT
23	AX-89961791	Affx-88935816	15	20504984	ACCCTCTGCAGCAAGCCGAGCTAGCTCAGTAATGGTAGCAGCAACAA	[T/C]	CGAAATAGAGCTTGAATGGTTTCTGATAGAATGATAAATTAATTTGTT
24	AX-89965325	Affx-88933791	15	40947776	TACCAAATGTACTGTTAAACAGCTAATCTTCTGTAATCTCTTCTGTT	[C/A]	TATTATTACTCTGTAATTTCCAGCATAAAATAAATGTAATCTTAAATGTA
25	AX-89956807	Affx-88907954	15	49466284	GAGAAGACCTCTGCTCGGTTAAAACGGCTCATTCTTGAACATGTAGGAT	[C/T]	GGCAAGTCTAAAAGTCTTACTATTCTGATTTTGTTTAAAGTCTCCCTGAC
26	AX-89960748	Affx-88921885	16	30115455	TCCAACCAGGAGGAGATGGACCTGTGGTTTGGCCTGCTCAAAGAACACAT	[C/T]	GAGATCAACGGGGCACCCCATTCGCACCTCTCGAGACCTACACCAGGTT
27	AX-89974922	Affx-88908638	16	57091655	GCACTTCATAAATACATATGAGCAGTGTCTGTAGCGAGTGTAGTGTG	[T/C]	GTTCTGTGTTCTCATTGAGCAATAAATACGCGGAAGCTTTATAAGAGA
28	AX-89948491	Affx-88904297	16	67381621	GAGTTGGGAACTGACAGATAAGTACACAATTAATTAATCCTTCCCTTT	[C/T]	ATTTCACTCCACAGTCAACCAATAATAGTACAGCATAATGTTTAAAGAAA
29	AX-89958544	Affx-88957749	17	19684392	AGCCAGACTAACGACTGTGGCTTAAACCCTGGCAAAATGCTAACATCCTTT	[T/G]	TGATGAGATAGTTTCTATGTGCTTACTTACTTGCAGTACACAATCACC
30	AX-89950127	Affx-88935977	17	31843473	GTGAACTCTCATGTTAGTTGGCAGTGTGCTGGTTATCATTCACCAAGTGG	[C/T]	GGAGAAATTTCCACAGGAAATCTGCTGAATACCAGCCATTCACAAATAA
31	AX-89968537	Affx-88933710	17	56405768	CTTCTGTGCTTTCATCTTAAAGTAAGTATGACAGTTTTCAGAACACAT	[G/A]	CAGTGATGACAATAGGATGTTGGGTCAAGGGATTGAGACTGTCAATA
32	AX-89961099	Affx-88919822	18	31109343	TGTCCTCATAGCAGTGTGTTTGTTGTTAGGTTCTTGAACATCTCACATT	[A/G]	AGTCATTGTTATTTTTCAGTTTGAAGGTGTTCTGTACCTTTAGCTGAGGA
33	AX-89963875	Affx-88929164	18	46441071	TCCATGAGGTCACATTTGCTCCTCCAGAGAAGGGAGAGTTTCTCAGTGTAT	[G/T]	AAITFAAGCCCTTCTCCTCATGAACTGCTAATATATATCTAAGAGAC
34	AX-89971700	Affx-88904708	18	60697513	GAAACGGTGTATCCCCTGCAGGCCAACCTGGTTTAAAAAACATAACTTAAT	[A/G]	TACTGTGATATAAATCTGGTCCCGGACTGTCTGGGCTGTCTGGGCTACG
35	AX-89952960	Affx-88958176	19	20300272	CCATTTGAACCTGAAACCAATACTACTGTGCAGAGGAAGGTAAGGTTGCA	[T/C]	TATTTTGCATATTTTAAACAGCAGGCCAGTGCAGGTTACCTTATTTGTAA
36	AX-89929300	Affx-88932061	19	30536367	AGGTAGAGGTTCCCGTGGTATGGATGATCCAGGCTGTGTTCTCACTAAG	[C/T]	GCCACAAAGGCAGAATCCTCCAGGCTTTGTACCAGTGCCTCTTGCCTTT
37	AX-89954513	Affx-88913761	19	42148220	TTAGGCAAGAGTACACATATAGCACTATGCATTTTGGAGACTGCACAACT	[T/G]	TAGATAGCAATACAATGTCTTATGTTATGTACATAATATGATTTGGC
38	AX-89975694	Affx-88919574	2	34500400	TTAAATGCAATGTTTGTAGGCTACCATGAAATCCATGTTTACATAT	[T/C]	TATTAAGTAGGCTGCAGGCTGGTCTAGTCTGCATAGACATGGGAG
39	AX-89955615	Affx-88940420	2	41265731	GGTAAACCTGCAGTCCAGGTCATCTTTATAAACCTTGTCTCATGCT	[A/C]	TTTTCTATGGAAGCAAGCAATAATACATTTCTCAAAGAGCTCCATTTGGG
40	AX-89920606	Affx-88937886	2	54132800	CCTGTAACCTTGACCAGAGAATATTGCGGTTGGTGGCTAAAATATCTCAGAA	[G/T]	AAGAAGATATAACAGTGGACTACAAGTCAATTAAGTGCAGGCACATACAGT
41	AX-89969417	Affx-88929144	2	81832941	GACGAGGCTGATAGGACTGTAGAAAATAGGTTCTGATCATTTAGGGCA	[T/C]	GCAACAGGAAAACAGAAAGTGAGCTTTTCTACTGTGCAAGTCTAGATAGT
42	AX-89959835	Affx-88943568	20	20652042	ACAAGTTTCATATTAAACTAGCTTCTAAACTTAGTCAGCAATGTTGTA	[A/C]	TGTAACAGAGTATGGTCAAGGGTAGCTGTGTAAGCACTTAAAGCAAA
43	AX-89940779	Affx-88945100	20	29041675	GATGTTGATAATGAATGTACCCTATTTTAGAAATGGGGCATTAGAGGCT	[G/A]	GTGCTCGGGCAAAAGGTTAAGCCTACCTCTGGACTAAAAGCGTGTCCGAT
44	AX-89965084	Affx-88913592	20	40461619	GCCCAGGCTGCGGGGAGGTCGCATAAATAACAGGACTGTCTACGAGA	[G/A]	GCAGCGCAACCCGAGCAGCAAGGCTTACCTTACCAGCAAAACGGGAG
45	AX-89930493	Affx-88911021	21	24841633	CGGATGTTTTCATAGTGGTATGTTACAGGGGGAACCACTTGTCTGCG	[T/C]	GTAACGTAAATAGAGCAGGTAGAGAAATATGTTGTCTGTTCTGCACAAC
46	AX-89967805	Affx-88919223	21	33849268	GGGGTCAAGTTTCAATGCTATTCAATTCAACTCAGAACTTAAAGCAAATTC	[A/C]	AAATCCAAATGAAAAGCATTGGAATATAGTGTACTCTCTGAAATTGATG
47	AX-89919557	AX-89919557	21	44890174	CGTAATGATAGTCAATTTTATTCTGTTGTAATGAAAAGAAATGGGT	[C/T]	TTCATTTCTAGGTCGGGTCCTCATCTTACTGGTTCGTAAGATGCCATT
48	AX-89926113	Affx-88910182	22	2527230	GGAGAACGGTATTTGCTAAAACATGTGGTTAGCATGGTGCATGTTCCG	[C/T]	TTTGTCTGATTTTGGCCGTCATGACTTTCAGTGGGCTGCATGATG
49	AX-89934313	Affx-88944137	22	15276760	CCACGCCACAGAGTCCGCAAAAGCAAACACTTGTTCATGAACATTCTCAG	[G/A]	GCAGTTGGGATTGACATATTGTTAAGCAGTACCAATTTGCTCTGCGCT
50	AX-89936199	Affx-88957045	22	25067842	ACAAGAGGGCTCCACATGTAGTATACTCAGTAAAGAAAATATGTTCTGCAC	[G/A]	AGCTTTATTGTGGAATAAAGGATAATACATTGGTCAAACTTCCCTGCACAG

Table A2. (continued)

51	AX-89976251	Affx-88956849	22	34499516	CAAATAAAATATTTTGTGGGTGTTTTCTCCTCTTCAGTTTTAAATATAA	[A/C]	ATTCCGTAGTATCCAAAGGGAGCTGACTGGTAAAAAATAATAATAATCA
52	AX-89939024	Affx-88941198	23	5005294	CTGGACGTGAGCGATGTATCCTCCAAGTGGCCGAGTATATCAGGATGT	[C/T]	GTCAATATACACAACCACCTGACACCTCGTTACGAATGCCTGGAACA
53	AX-89967838	Affx-88954880	23	22877723	CATTGTCCCTGCCCGCGTAACCTGCAGATTAAGTTAATACTGCTGGAGG	[T/C]	AGTCTGACCAAGTGGTACATGCGGAGGGGGCTTGTGACCCGACCCTGAGGTA
54	AX-89939948	Affx-88938715	23	31554543	TAAATATATTAATAATCTGCAGTGTGAAAAGTACAGCCATTCCACAAA	[G/A]	AATAGTTGATGTTCTCCAACATGTAATGACTATCAATGATGGTCTCTCTT
55	AX-89964627	Affx-88918711	24	11413051	TACATTTCTCTGCAGTCTGATCCAGGATTCACGCGTAGTAACAACCCC	[G/A]	TTATTAAGATTTAAGAGCACTAGTCCAGAAATCAG
56	AX-89955558	Affx-88910480	24	21775168	TGGGATATTTCTGTCTTTTTTCTTGCCTGTGAAAAAATAGACCAATG	[G/A]	AAATAATGATTCCTGAATAACAATTAATTTATTTGGGATTAATGTCTTTGT
57	AX-89961464	Affx-88942163	24	32122911	CAGCTTGACCAGTCTGCAGTCACTGTTTCTTCATAGTTTACCCCAAGCA	[A/G]	CAAGTGATGGATGATGTTGACGACTCCATCCAGGCTTCAATCATCTCTG
58	AX-89917006	Affx-88916475	25	27755396	TTATATACATGTATATGTAGTAATCTAATCGTGCTTAGAATGATTTGTG	[G/A]	GTTTTGCTAATATATTTGGCAAATTCATGTAACCTTTGTGTGTAGTCCATCG
59	AX-89933102	Affx-88905156	25	49243042	GAGGAAGAAGAGGGAGGAGTCTTAATACTGGGCCAGTAACTCAGCAGAT	[C/T]	ATGGGCGGGTTCTCTGTCTCAGGCATTGCTTACAGTCCACACACTGGA
60	AX-89969648	Affx-88944911	25	75579174	ATGGCCAGGTAGGTGTGCCAGGTATGTAATGTACCTTCGTTCTCCTCCA	[G/A]	CCTTCGGACTTTGTTGAGAACAGGCTGCAGGTTTCAATGATAGAGGCGATGGC
61	AX-89926530	Affx-88921539	26	6956186	TTTAAACTAAATTTACCAGTTTAACTTTCATGTATTCATTTTTCAATCG	[G/A]	CTCAATCTGGTCCCGTGACTTCGAATTAAGAGAGAGAGAAGCTCGTGT
62	AX-89929326	Affx-88934622	26	10750755	CTGCAGTTACTTGCTGTAAATGGTAGTATTTAGTATTTCGGATAGTA	[T/G]	ATCTAATGGACACTATTAGAGTCTACTGGTGTTTTTCTCTTTTGCTT
63	AX-89968211	Affx-88914406	26	16466139	GTCCAATATAAATCCAGGACAAGTGTCCATCATATAACTAAATCTTCT	[T/C]	ACACAAATATATTCTGCAGTAATCTGTTCTTCAATGAAAAGGCAAGGATT
64	AX-89957732	Affx-88919119	27	4377318	TAAGGCATTCTCCTCTCACATGTGAGAGTTGGATCTGTTGAGAACCAC	[T/C]	GTGTCATGAATTTGGTGCCAACTCCAAGAAGCTTGAGGCAAAGCACTTC
65	AX-89917563	Affx-88908551	27	20183064	AGTCAAACCTCTGAGAACTAATATCTTTGTAAGTGAAGTAGAAGAGAAA	[A/G]	CTCCCAAGTACTAACAGACTAGTGTTCGATTTTTAAATACCTTCAAATA
66	AX-89945678	Affx-88918968	27	31272568	AGTGTCACTTGATTTCCACAGCAAGGGAAGGAGGATTCAACAAATGAG	[T/C]	GAGAGCTGCGTCACTTTTTCTGATTCATGTATTTTGGGCTGCACCTCC
67	AX-89926812	Affx-88927877	28	2208195	ATAAAACGTAGTCTCCATGTTAATGTATTTTCTCTCTCACTATA	[T/C]	ATACAGGCACGCCAGACGCTCTTTCCAGTCTGACCTGCCAATACCTGC
68	AX-89941147	Affx-88933367	28	14750831	GAGTGTGTCTCAATGTGTGCTTAGATATTCATCATGTTGGCGACACAAG	[C/T]	CTCCACAACAGAACAGGGGTCTGCCACAAGTGCACCAGAGGGCTCACTG
69	AX-89952141	Affx-88948120	28	29339648	CAGCATCCAATTTCCATGCCTTTTGCATTCATTAAGCATAACAGTAAACCA	[T/G]	CTATCATACCTACAGAGTCTCGAGTCAAGATAATATAACCAATGAG
70	AX-89918953	Affx-88942472	29	2989045	GTTGCCCAACAAAACACATCAAGGTAAGAGAAACAAAGCCTTCTACT	[A/G]	CAACCAACGCTGAGACTTTACAGTGTCTACTACTGGGAGAACTTGTCT
71	AX-89974397	Affx-88957563	29	5462777	TCCATTTTTACTCATGTTTTCATGCTTTTACATGAGTTCTACTGTCC	[T/C]	CAGTTGTAGTCTACCGAGTAATACTGTAGCACAGCACTAATGCTCA
72	AX-89930351	Affx-88925262	29	17427735	ACCAAGTGTCTCAGTCAATTCAAATTCACAAGTTCAAATAAAAATATT	[C/A]	TTCTATTCGCATACATTTATTCAGTACTGACATTCACATAGCTCATAT
73	AX-89923922	Affx-88935389	29	35732238	CAGTTGTTTTCTGTAGCATCAATTTGACACATTTAAAAATACCATTTCT	[C/T]	TTTATCTAGTCTTTTACGTGACAAAACCTCAGAAATGTTCAATATGAAT
74	AX-89939700	Affx-88932008	3	24650556	TCCATCTAAAAGTTAATATTTCCATCAAGGCAACAGAAAAACAAATGGGA	[G/T]	AAATGCTTGCTGCCCTTAACATAGTGCAGTACAGAAAAAACGCACACA
75	AX-89950885	Affx-88938229	3	43704944	GCCATCAAAGAGAACAGCGATAAGATGAACCTGCAGGTCAGACTTCTT	[T/C]	GCAGAGAAATGCTGAGATCTTTGAGATGATGATGTCGGCCACGGCTT
76	AX-89959983	Affx-88941602	3	78281519	ATTCTGTTTTAGTGTGTGCAACTTGTCTACTAACTGTAATAAGTCCCG	[C/T]	AGCGCCTGCTCTGAGATCACTTGGAAAGTCACTCCACCCACCGGCT
77	AX-89972102	Affx-88933505	3	82697693	AAAAGAGAAGAGGAAGTTGATAGAAGATCTGCAGGTGCTAATTTGTGTATA	[C/T]	GATTAAGAGAGACACAGATGATGTTGTTCAAGTAAAAAGCTAAAAGAGG
78	AX-89961701	Affx-88927585	4	32219851	GGTGTGACTCCCTTACAACACACATCTACAGCATACAATCTGTGTAG	[T/C]	CTCTCAGTGGTCCAGTGTACGTTCCCAATGGCCTGGTTAAGTTCACT
79	AX-89971402	Affx-88935025	4	44594130	TGCTACATCTGCATTTCTGCAGCTTGGTGAATACTCCTAACCTCTCATTT	[T/G]	TATTTTAGACATACTGTAGATTACATCTGGAGAAATGAAACAATAGTCTT
80	AX-89955511	Affx-88947077	4	73965149	TCCTCTACCAAGTCTCATTCTGGTGCATCTGTAGGAAGATGTCACAGCC	[G/A]	CCCACAACTCTCAATTTAAGAATATCTGTGGGATCGTGGCCAGTTGGA
81	AX-89946502	Affx-88940792	5	19183576	CAGAGATATTTGGTCTGTGTGTTGCTCTGCTACAGTGCTGCAGTGAAC	[C/T]	CAACCCAGGTTTCCCTCAGGTGTGTGTTGTCTGTGATGCTGCAGGCTCTA
82	AX-89954518	Affx-88928878	5	44591149	TCATACTAAGTTAGCCATATCATGGTAGGAAAAAGTAAGTAAACATGAAAC	[C/T]	GAAGTGTATTTTTGACGACAAAAGCCTTTTTCAATGTCTGTTTTTGTGAA
83	AX-89937071	Affx-88922459	5	55579593	AACAAGGTACTTTCTTCCATCTGAAAATGTTAATTAATCTTAATCAATT	[A/C]	ATCCCAAAAATGGAATGTACATTAATCAATGGGATACAATAATCATGGCAAAA
84	AX-89923577	Affx-88958005	5	68929157	GGTCTCCATAATGATGTCATGACTCAGTCAAGGCTATTATTAAGAT	[C/T]	GTAAAGGTTCTGCTCTGACAGCAAGTGTCTGTAATCTGTTCCATGACC
85	AX-89918398	Affx-88949230	6	27351700	TGAAGCACAAAAACTCAAAATACACATTTAAAAATGCGCAAGGGG	[A/C]	TAGCAACAGTTGACTGACGTGCAGGTTCCAGATTGGGACTTACATAACA
86	AX-89931324	Affx-88951526	6	35261851	GCAGTGCCATCTGATGATCAACTGATCTGTACATGATCTGAGGTAATTA	[C/T]	CTTAGAAAATAAATCACTAGAACAGGCTCGAGGAACTCAGACTTGGC
87	AX-89931135	Affx-88935793	6	47365165	GAGTTAGGCGCTGGGCAGAGTGTTTTCTCTCTGCTCAGAACCCATCAT	[C/T]	TCTCTGGCTCCTGCTGCGAGATTATCTCATACTCCAGTCTCTCTGCAGT
88	AX-89958110	Affx-88961206	7	2366487	AATAAATACCAAGTTCTAGTAAATGGGTAATGAGGACAACAGGAAATGTT	[C/T]	GTGGAACGAGAGACGCTCTGGATTGCATGCTGACTGAAATCTACCAAGTG
89	AX-89951181	Affx-88914755	7	21103558	CTCAATGGCCTGTGTTCCGGCTCGTGAATCAAAGATGTAAATATAGTT	[C/A]	ATCCAGATAGTTACAATGGTTTGCTCTATAAAGATATATGGAATCACAGC
90	AX-89969081	Affx-88944605	7	44439207	AGTTGACTCACTACTCTCATGGTGTATTTATGCTGTCTATGAACA	[C/T]	AGGTGCTGTGACTATTGGACATGAGGAACACTCACTGATTTAGGTCTGC
91	AX-89946845	Affx-88927443	8	1506527	TTTCTACAGAGATACCCAAAATGCTCTGATTTGGTAAACTGAGGCAA	[C/T]	GACATTAGATTAGCATGTTATTTAAACATAAATGTCACCCCTCTACTC
92	AX-89929725	Affx-88913106	8	52683012	TGCCTTCCAGTGCACAAACCCCTCCCTCCAGGGGATCTTTAGCTCCAC	[A/G]	TTTAAAGGTAATGTAACACTAAATAACCAAGTACATCGCTTTTAAA
93	AX-89938504	Affx-88907941	8	79885643	CTGATAAATCCCAAAAATAGGATGACCTGTCCCAACGAGACGCGTAG	[G/T]	TGTACCAAGATTTTTATAATTAACTCAAGATAGACCACAACCTTAGTCA
94	AX-89946241	Affx-88923924	9	18273610	TATAACCCCTCCTCTAAATCCACTCACTGCTCCTGTTCCCTCAGAAAAGT	[C/A]	TATTTTTGCCCTCTGAGAGTGCACCCGAACTTGTTCCCATAGCCAAATAC
95	AX-89941429	Affx-88928670	9	37294044	TAGTTCTGATTATGTACATTAATACAAACAGTATAAAACAAATGAATCTT	[C/T]	ACATACATTTGATAACAACTGACATTTGTTTTTAAACACTGCCGTACTA
96	AX-89946316	Affx-88921797	9	67682705	CTACAAAAACCTCACACACACACACACTGTGTCTGTAAACGACGAC	[C/T]	ATTACAGTCTGTGATTTACAGGAGATGTCAGTGGTCTCTGCTGTAC

Table A.3. Progeny overall survival for each parents (A. Dams and B. Sires) at Sf stage (*Sf*: from fertilization to three weeks after the first feeding). Fisher exact tests were performed with FDR correction on alive/dead number of offspring per parents (Letters (a, b) : p-value < 0.05).

Dam	Progeny survival (%)
Fb1	63.5 ^{a, b}
Fc3	73.0 ^a
Fc4	76.2 ^a
Fb5	77.8
Fc2	79.4
Fa3	79.4
Fa4	85.7
Fa2	85.7
Fb3	88.9 ^b
Fb4	88.9 ^b
Fa1	92.1 ^b
Fc1	92.1 ^b
Fb2	93.6 ^a
Fa5	95.2 ^a
Fc5	95.2 ^a

Sires	Progeny survival (%)
Mc2	78.1
Ma1	80.0
Ma2	80.9
Ma3	80.9
Mb3	85.7
Mb1	86.7
Mc3	87.6
Mc1	88.6
Mb2	91.4

Table A.4. Estimates of genetic parameters and Akaike information criterion (AIC) for two GBLUP linear models (AM and AMD models) of survival from eyed-egg to hatching stage (She, dead=0 / alive =100) and from eyed-egg to three weeks after first feeding (Sef, dead=0 / alive =100) considering correction for embryo inbreeding coefficients and accounting either only for additive genetic (A), and maternal (M) effects, or for additional dominance deviation effects (D). Heritability (h^2), ratios of dominance variance (d^2) and maternal variance (c^2) over phenotypic variance (σ_p^2) are presented with their standard errors given in brackets.

GBLUP model	<i>She</i>		<i>Sef</i>	
	AM	AMD	AM	AMD
AIC	7487.55	7477.15	7951.50	7945.35
σ_p^2	403.23 (22.65)	409.65 (23.73)	666.88 (33.96)	671.63 (35.12)
h^2	0.015 (0.02)	0.013 (0.02)	0.012 (0.02)	0.010 (0.02)
d^2	NA	0.249 (0.08)	NA	0.284 (0.09)
c^2	0.058 (0.03)	0.063 (0.03)	0.034 (0.019)	0.029 (0.02)

Table A.5. Correlations between parental inbreeding coefficients and offspring survival at any stages (p-values figure between brackets).

	Se	Sh	Sf	She	Shf
F_{dam}	0.22 (0.418)	0.19 (0.485)	0.17 (0.522)	0.08 (0.769)	-0.11 (0.702)
F_{sire}	-0.64 (0.06)	-0.49 (0.171)	-0.65 (0.095)	0.16 (0.685)	-0.34 (0.346)

Table A.6. Statistical test on survival among individuals splitted into two groups according to their degree of inbreeding (F_p).

	Moderate inbreeding threshold		High inbreeding threshold	
	$F_p \geq 0.10$	$F_p < 0.10$	$F_p \geq 0.15$	$F_p < 0.15$
Number of dead individuals	18	18	0	36
Number of alive individual	403	419	49	773
Fisher exact test	p-value = 1		p-value = 0.258	

Chapitre 3

Effets de la consanguinité génomique locale, récente ou ancienne, sur les caractères de taille et de reproduction des femelles de truite arc-en-ciel

Depuis plus de trente ans, l'amélioration génétique de la truite arc-en-ciel a permis d'accroître l'efficacité de la production truticole. Mais cette sélection a aussi contribué à une augmentation de la consanguinité et une perte de diversité génétique (D'Ambrosio et al. 2019), mettant en péril le progrès génétique à long terme et la capacité d'adaptation des populations aux changements environnementaux. Afin de mieux gérer la consanguinité et ses conséquences au sein des populations de truites arc-en-ciel, il est nécessaire de mieux comprendre l'architecture des effets de la consanguinité sur des caractères d'intérêt aquacole.

A cette fin, trois questions importantes ont été traitées dans ce chapitre. La première question était de **quantifier les effets globaux de la consanguinité sur les performances de taille et de reproduction des truites**. La seconde question était **d'identifier les zones du génome associées à ces variations de performance : des régions particulières sont-elles plus souvent impliquées ? Sont-elles liées à des événements récents d'accouplements entre reproducteurs apparentés ?** La dernière question était de savoir si **nos observations corroboraient l'hypothèse privilégiée pour expliquer le phénomène de dépression de consanguinité**, à savoir le rôle majeur d'allèles récessifs à effet délétère.

Cette étude est la première qui ait visé à estimer les effets précis de la consanguinité accumulée localement et dans le temps au sein d'un génome de poisson. Pour ce faire, les performances de croissance et de reproduction de plus de 1300 truites d'une population commerciale sélectionnée depuis 9 générations ont été analysées.

Conclusion du Chapitre 3

Nous avons montré que la consanguinité localisée dans certaines parties du génome peut avoir des effets notables, négatifs comme positifs, sur les performances de la truite arc-en-ciel, sans que l'on observe forcément un effet de la consanguinité globale à l'échelle de tout le génome. Comme largement rapporté dans la littérature, le principal phénomène observé est un impact négatif de la consanguinité récente, expliqué par l'accumulation d'allèles récessifs délétères à l'état homozygote. Cependant la consanguinité, en particulier celle accumulée dans les générations passées, engendre aussi des effets favorables sur les performances de reproduction des truites.

Cette étude est une première étape vers une gestion précise des accouplements entre reproducteurs afin de favoriser ou, au contraire, limiter la consanguinité locale selon que ses effets sont positifs ou négatifs sur les performances. L'utilisation d'indicateurs génomiques locaux de la consanguinité permettra ainsi aux sélectionneurs de mieux gérer les compromis entre valeurs génétiques des reproducteurs et effets indésirables de la consanguinité de leur progéniture.

Les différents résultats obtenus dans ce chapitre ont été présentés sous forme de communications orales aux *Journées Techniques du Sysaaf* en 2020, ainsi qu'aux congrès *International symposium on Genomics In Aquaculture (GIA)* et *European Aquaculture Society (EAS)* en 2021. Ils ont également fait l'objet d'un article « ***Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout*** » publié dans le Special Issue 'Genomics for Aquaculture and Livestock breeding improvement' d'*Evolutionary Applications* en 2022.

Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout

Katy Paul¹ | Jonathan D'Ambrosio^{1,2} | Florence Phocas¹ 

¹Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France

²SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, Rennes, France

Correspondence

Florence Phocas, Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas 78350, France.

Email: florence.phocas@inrae.fr

Funding information

European Maritime and Fisheries Fund and FranceAgrimer, Grant/Award Number: RFEA47 0016 FA 1000016

Abstract

Recent studies have shown that current levels of inbreeding, estimated by runs of homozygosity (ROH), are moderate to high in farmed rainbow trout lines. Based on ROH metrics, the aims of our study were to (i) quantify inbreeding effects on female size (postspawning body weight, fork length) and reproduction traits (spawning date, coelomic fluid weight, spawn weight, egg number, average egg weight) in rainbow trout, and (ii) identify both the genomic regions and inbreeding events affecting performance. We analysed the performance of 1346 females under linear animal models including random additive and dominance genetics effects, with fixed covariates accounting for inbreeding effects at different temporal and genomic scales. A significant effect of genome-wide inbreeding (F) was only observed for spawning date and egg weight, with performance variations of +12.3% and -3.8%, respectively, for 0.1 unit increase in F level. At different local genomic scales, we observed highly variable inbreeding effects on the seven traits under study, ranging from increasing to decreasing trait values. As widely reported in the literature, the main scenario observed during this study was a negative impact of recent inbreeding. However, other scenarios such as positive effects of recent inbreeding or negative impacts of old inbreeding were also observed. Although partial dominance appeared to be the main hypothesis explaining inbreeding depression for all the traits studied, the overdominance hypothesis might also play a significant role in inbreeding depression affecting fecundity (egg number and mass) traits in rainbow trout. These findings suggest that region-specific inbreeding can strongly impact performance without necessarily observing genome-wide inbreeding effects. They shed light on the genetic architecture of inbreeding depression and its evolution along the genome over time. The use of region-specific metrics may enable breeders to more accurately manage the trade-off between genetic merit and the undesirable side effects associated with inbreeding.

KEYWORDS

dominance, fish, inbreeding depression, overdominance, run of homozygosity, selection

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Inbred offspring tend to have a greater number of abnormalities and poorer survival, growth and fertility compared to the progeny of unrelated parents (Fessehaye et al., 2007; Mrakovčić, & Haley, 1979; Thrower & Hard, 2009). This decrease in fitness is referred to as inbreeding depression and is caused by the increased homozygosity of individuals (Aulstad & Kittelsen, 1971; Charlesworth & Charlesworth, 1999; Landweber & Dobson, 1999). Inbreeding depression is currently explained by two main hypotheses. The partial dominance hypothesis assumes that inbreeding depression results from the expression of deleterious recessive alleles in homozygous individuals, that is the so-called genetic load. These deleterious alleles are present at low frequencies in populations. As inbreeding increases in a population, the frequency of deleterious recessive homozygotes, which were initially hidden in heterozygotes, increases and exposes the deleterious effects (Fu & Ritland, 1994). Alternatively, the overdominance hypothesis refers to dominance effects displaying the heterozygote advantage and, as inbreeding increases, the number of heterozygous genotypes is reduced and the superior heterozygote genotypes are less frequent (Wright, 1984). A final hypothesis is that epistasis between dominance effects across loci can also generate inbreeding depression, such that as inbreeding accrues, favourable gene combinations among heterozygous genotypes decrease in frequency (Jain & Allard, 1966). Distinguishing this 'pseudo-overdominance' hypothesis from true overdominance is hampered by the difficulty in distinguishing between linked deleterious mutations in a genome region and a single locus with heterozygote advantage (Ohta & Kimura, 1969). If deleterious mutations are common, genome regions may often carry mutations in different genes in repulsion. The region would therefore display a heterozygote advantage even though no overdominant gene is present. The partial dominance hypothesis is widely accepted as the most common mechanism causing inbreeding depression and has received the strongest empirical support until now (Ceballos et al., 2018; Charlesworth & Willis, 2009; Fu & Ritland, 1994). All hypotheses concern the existence of genetic effects that are not additive and need to be considered when describing performance. Fitness-related traits are more affected by inbreeding than traits under less severe directional selection (DeRose & Roff, 1999). Inbreeding depression has been documented in many different plants and animals, both in the wild (Chapman et al., 2009; Reed & Frankham, 2003) and among farmed livestock (Leroy, 2014). In populations under artificial selection, an accumulation of inbreeding is unavoidable because of the use of a limited number of breeding individuals and their intense directional selection. These procedures lead to the widespread use of related individuals as parents of the next generation, thus reducing the effective population size and hence genetic diversity over the generations. In livestock species, inbreeding depression has a direct impact on the income of breeders (Leroy, 2014) and maintaining genetic diversity is a challenge.

Before the massive increase in genomic information, pedigree relationships were used to estimate inbreeding effects and manage inbreeding within a population. The effect of inbreeding on

economically important traits was mainly investigated by regressing phenotypes of interest on the inbreeding coefficient value using pedigree (Wang et al., 2002), which severely limited the understanding of inbreeding effects. Pedigree information is based on the expected proportion of the genome that is identical by descent (IBD) between two parents, so it does not capture variations due to Mendelian sampling and linkage during gamete formation (Kardos et al., 2015; Wang, 2016). Genomic information enables an estimate of the realized proportion of the genome shared by two individuals either genome-wide or in specific regions (Hill & Weir, 2011). The use of genomics rather than pedigrees to measure inbreeding has therefore been promoted (Kardos et al., 2015; Wang, 2016) as enabling major advances in our understanding of inbreeding and identifying the regions and genes that result in inbreeding depression along the genome (Kardos et al., 2016; Pryce et al., 2014). It has also led to renewed interest in estimates of dominance effects that can improve our understanding of the genetic architecture of inbreeding depression and enable more accurate estimates of inbreeding effects (Toro & Varona, 2010; Vitezica et al., 2016).

Although access to local inbreeding is now greatly facilitated due to the development of genomics, the number of studies focusing only on global inbreeding and local inbreeding remains quite low (Botero-Delgado et al., 2020; Judson et al., 2018; Lieutenant-Gosselin & Bernatchez, 2006). The use of region-specific metrics to identify areas of low genetic diversity may enable breeders to more accurately manage the trade-off between genetic merit and the undesirable side effects associated with inbreeding (Howard et al., 2017).

Among the various molecular-based inbreeding indicators available (Ferencakovic et al., 2013; Zhang et al., 2015), runs of homozygosity (ROH) have been revealed as being the most accurate measure of inbreeding (Howrigan et al., 2011; Zhang et al., 2015). ROH are defined as contiguous homozygous stretches of the genome assumed to be inherited from a common ancestor and thus considered as IBD segments (McQuillan et al., 2008). ROH detection has been used to identify genetic anomalies based on homozygosity mapping (Keller & Wallers, 2002; Li et al., 2011). In both humans and cattle, it has been found that ROH are enriched with deleterious variants (Szpiech et al., 2013; Zhang et al., 2015) and they can be used to directly estimate inbreeding depression (Silió et al., 2013). Keller et al. (2011) showed by simulation that ROH-based inbreeding measure is the most powerful method to detect inbreeding depression (compared to pedigree or marker-by-marker-based inbreeding metrics) and that the statistical power of ROH-based inbreeding metrics was good (80%) in a randomly breeding population of moderate effective population size (i.e. 100) with only a small sample size of 700 individuals.

As well as its ability to measure region-specific inbreeding, ROH is able to date inbreeding events through the length of ROH segments (Ceballos et al., 2018; Gomez-Raya et al., 2015). A large ROH is likely to result from a recent inbreeding event because only a few recombinations have occurred since generation of the most recent common ancestor of the parents, while a small ROH is in favour of an older inbreeding event as several recombination events are likely to have occurred, leading to reduced IBD segments

(Ceballos et al., 2018; Purfield et al., 2012). In addition, when N_e is small, rare mutant alleles are tagged by long haplotypes as a result of relatively recent inbreeding. However, when N_e is large, these long haplotypes are broken down by recombination and new ones are not created rapidly because new inbreeding occurs slowly (Keller et al., 2011; Thompson, 2013).

In the particular case of *Oncorhynchus mykiss*, D'Ambrosio et al. (2019) showed that effective population sizes of French farmed rainbow trout lines have been steadily decreasing during the past 10 generations and inbreeding appears to be high not only due to selection, but also because of founder effects and sweepstakes reproductive success at the start of breeding programmes. The impact of these significant levels of inbreeding on French rainbow trout performance needs to be quantified in order to assess potential inbreeding depression phenomena and risk of limiting future genetic gains due to a loss of genetic diversity (Jannink, 2010; Muller & Pearson, 1979). Estimates of inbreeding effects on important traits in fish have been limited to a few experiments on salmonids, as reviewed by Wang et al. (1999, 2002). These previous studies revealed significant but generally moderate inbreeding depression effects on weight at harvest (Kincaid, 1976; Pante et al., 2001; Rye & Mao, 1998), body weight and length (Naish et al., 2013), egg mass, egg hatchability and fry survival (as reviewed by Kincaid, 1983), on egg numbers and the spawning age of rainbow trout (Su et al., 1996) as well as on spawning date for wild populations (Waters et al., 2020).

Therefore, based on ROH metrics, the objectives of our study were to quantify the effects of dominance and inbreeding on female size and reproduction traits in rainbow trout and identify both the genomic regions involved and the time of inbreeding events with effects on these performance traits. The traits under study were therefore female fork length, postspawning weight, spawning date, coelomic fluid weight, spawn weight, egg numbers and egg size. The first question we addressed was whether genome-wide inbreeding and dominance effects could explain significant proportions of phenotypic variance in female size and reproduction performance. The second issue was to assess whether the effects of inbreeding were mainly due to recent inbreeding events and only observed in specific regions of the genome. The last question was whether our results could corroborate the partial dominance hypothesis, that is the major role of deleterious recessive alleles as the underlying mechanism explaining inbreeding depression.

As far as we know, this is the first study to have reported on genome-wide and region-specific inbreeding effects based on ROH metrics, as well as their temporal variations for fish traits. Our results may help rainbow trout breeders to design and implement new genomic selection methods that take account of local genetic diversity along the genome to manage inbreeding at the genome level. In addition, and more globally, they will perhaps shed new light on the genetic architecture of inbreeding depression and its evolution over time.

2 | MATERIALS AND METHODS

2.1 | Population

The phenotyped and genotyped rainbow trout population comprised 1346 females from two successive cohorts: C1 (726 individuals) and C2 (620 individuals), which were produced in 2014 and 2015, respectively. After hatching, the juveniles were raised to about 20 g (C1) in size or 58 g (C2) at the Escort site on the plain (spring water temperature within the range 14–16°C) in 17 m³ breeding tanks at densities increasing from 2 to 25 kg/m³ (C1) or 36 kg/m³ (C2). They were then moved from the plain to the mountains at the Sarrance site (spring water temperature within the range 8–9°C) and raised in 126 m³ tanks until they reached weights of 450 g (C1) or 580 g (C2). Finally, the trout were transferred to 70 m³ tanks with rearing densities ranging from 40 to 100 kg/m³ until spawning. Both juveniles and trout were fed with commercial standard diets, although some specific modifications were requested by the breeder regarding the diet given to trout from 1 kg until spawning.

The two cohorts were part of the 9th generation of selection by the 'Viviers de Sarrance' breeding company and were produced from 71 fathers and 83 mothers. This 9th generation was the first to have been produced using an optimum contribution method to select and mate the parents for a targeted genetic gain while minimizing the increase in inbreeding through a 'Minimum Parentage Selection' procedure (Chapuis et al., 2016). There were around 4 (± 2) full-sibs and 45 (± 20) half-sibs per parent in the phenotyped population under study. Details of the population structure can be found in the article by D'Ambrosio et al. (2020).

TABLE 1 Summary statistics of female reproduction and weight traits at 2 years of age

Traits	Number of individuals	Mean	Standard deviation	Median	Min	Max
Postspawning body weight (g)	1346	1867.0	410.4	1840.0	840.0	3116.0
Female fork length (mm)	1346	525.1	34.9	527.0	430.0	660.0
Spawning date (week rank)	1346	2.6	1.5	2.0	1.0	5.0
Coelomic fluid (g)	1187	43.2	57.3	32.0	1.0	455.0
Spawn weight (g)	1346	189.6	70.9	182.0	20.0	398.0
Egg number (#)	1346	4802.0	1765.4	4730.0	619	10,434.0
Average egg weight (mg)	1346	39.8	6.4	39.8	20.0	68.1

2.2 | Phenotypes

Raw phenotypes were collected from 2-year-old females. As for size traits, the ready-to-spawn weight (FW in g), postspawning weight (PW in g) and fork length (FL in mm) of the females were recorded (Table 1).

Regarding reproduction traits, the weight of the total egg mass (here in after referred to as the spawn weight (SW)), the weight (EPW) and number (EPN) of eggs pooled in a 2.5 ml sampling spoon and the spawning week number in the calendar year were all recorded. The spawning week number enabled us to calculate the spawning date (SD) corresponding to the rank of the week number within the spawning period, with discrete values ranging from 1 (for the first week) to 5 (for the 5th and subsequent weeks) within the cohort. The presence of overmature eggs in the spawn was also reported as a potential factor to explain certain reproduction traits. This is an important phenomenon in salmonids that do not spawn naturally under farmed conditions (Escaffre & Billard, 1979). Overmature eggs change in terms of their morphology (McEvoy, 1984) and composition (Craik & Harvey, 1984; Springate et al., 1984), which implies a significant decrease in egg quality.

The weight of coelomic fluid (CF) was determined by subtracting the spawn weight (SW) from the difference between the female weight before (FW) and after (PW) spawning:

$$CF = (FW - PW) - SW.$$

Average egg weight (EW) was the ratio between the weight of eggs (EPW) and the number of eggs (EPN) contained in the sampling spoon:

$$EW = EPW/EPN.$$

The egg numbers in the spawn (EN) were calculated as

$$EN = SW/EW.$$

To summarize, the traits analysed were three raw phenotypes, FL, PW and SW, and four derived phenotypes, average egg weight (EW), egg numbers in the spawn (EN), the weight of coelomic fluid (CF) and the spawning date (SD). If any two records were more than four standard deviations from the mean in absolute values, they were considered as outliers and discarded from the study. Thus, the phenotypes of 1346 fish were considered during the study (Table 1).

2.3 | Genotypes

One thousand three hundred and forty-six fish (726 and 620 individuals from the C1 and C2 cohorts, respectively) were genotyped for 57,501 SNPs (single nucleotide polymorphism markers) using the Axiom™ Trout Genotyping array (Palti et al., 2015) at the INRAE Gentyane Genotyping Platform. The quality control of genotyped SNPs was performed as described by D'Ambrosio et al. (2019), with

particular focus on removing SNPs with probe polymorphism and multiple locations on the genome. Only the 29,799 SNPs with a call rate higher than 0.97, a test of deviation from the Hardy-Weinberg equilibrium with a p -value >0.0001 and a minor allele frequency higher than 0.01 were retained for the analysis. All missing genotypes for the 29,799 SNPs were imputed using family information with Flmpute software and default values for all Flmpute parameters (Sargolzaei et al., 2014).

2.4 | Runs of homozygosity

Runs of homozygosity were identified for each fish using the PLINKv1.9 *homozyg* function (Chang et al., 2015) with the options `--homozyg-kb 1000 --homozyg-window-snp 30 --homozyg-snp 30 --homozyg-gap 1000 --homozyg-density 100 --homozyg-het 1`. ROH were thus defined by sliding windows with a minimum length of 1 Mb containing at least 30 homozygous SNPs as defined by D'Ambrosio et al. (2019). This minimum number of homozygous SNP was chosen using the formula described by Purfield et al. (2012) in order to limit the number of ROH that might only occur by chance. The maximum gap allowed between two consecutive homozygous SNPs in a run was kept at the default value of 1 Mb. A minimum density of one SNP every 100 kb was considered not to overestimate ROH length and up to one possible heterozygous genotype was permitted for each ROH. All values for parameters used to define ROH were tuned by D'Ambrosio et al. (2019) according to the marker density and the level of recombination along the genome to well estimate the number and size of ROH segments for French rainbow trout populations.

2.5 | Estimation of inbreeding coefficients

The total inbreeding coefficient (F_i) was calculated as the sum of ROH lengths in an individual i ($\sum \text{Length}(\text{ROH}_i)$) divided by the total length of the autosomal genome covered by SNPs (LGenome):

$$F_i = \frac{\sum \text{Length}(\text{ROH}_i)}{\text{LGenome}}.$$

The total size of the autosomal genome covered by SNPs (= 1.788 Gb) was calculated as the length of the autosomal genome, removing gaps of more than 1 Mb without any SNP from the total size.

We also derived local inbreeding coefficients, at either the chromosome Omyk level or a smaller region r scale (20 Mb segment) in order to accurately localize inbreeding events in the genome:

$$F_{\text{Omyk},i} = \frac{\sum_{k,r} \text{Length}(\text{ROH}_i)}{\text{LRegion}_{k,r}},$$

where $\sum_{k,r} \text{Length}(\text{ROH}_i)$ is the sum of ROH lengths of an individual i on chromosome Omyk (or the 20 Mb region k_r on Omyk), and $\text{LRegion}_{k,r}$

is the total length covered by SNPs on the Omyk chromosome or on the r region on Omyk.

Furthermore, we derived three other inbreeding coefficients (F_{Gt}) depending on the expected generation Gt (ancient, middle or recent) in which inbreeding started to accumulate.

$$F_{Gt} = \frac{\sum \text{Length}(\text{ROH}_i > n)}{L_{\text{Genome}}},$$

where t is the number of generations and n the minimal size of a ROH.

The correspondence between the number t of past generations and ROH length was derived using the average recombination rate c (in Morgan) between markers distant from the threshold value of n Mb (with 1 cM corresponding to 600 kb) and the formula $t = 1/2c$ (D'Ambrosio et al., 2019). Thus, recent inbreeding (F_{G3}) accumulated on average during the last three generations was calculated as the sum of ROH lengths longer than 10 Mb divided by the total length of the genome covered by SNPs. This '3 generation'/10 Mb' ROH length time point was chosen because it is well known that recent inbreeding events occurring during the last three generations play an important role in inbreeding depression effects (Makanjuola et al., 2020). We were also interested by the '9 generations' time point because it corresponded to the start of the breeding programme. On average, this earlier time point corresponds to an ROH size of 3 Mb. In addition, we chose an intermediate time point at 6 generations to better describe the dynamics of inbreeding events throughout the course of the breeding programme. The ROH length corresponding to this intermediate inbreeding (F_{G6}) was 5 Mb. It should be noted that any one of these temporal inbreeding coefficients could account for all the inbreeding events accumulated since the threshold generation considered, and therefore always contains the most recent inbreeding events. In addition, because the recombination rate varies along the genome, the average dating of size-specific ROH segments in terms of the number of generations overestimates the age of the most recent ancestor in genomic regions with lower than average recombination rates, but underestimates this age in genomic regions with higher recombination rates. Dating is therefore a very rough proxy used to describe global dynamics at the genome scale without any consideration of absolute dating values at region-specific scales.

2.6 | Mixed linear BLUP in animal models

In order to estimate the effects of inbreeding on female size and reproduction traits, two different genetic models were considered: (1) GBLUP— a genomic animal BLUP model with only additive genetic effects, and (2) D_GBLUP— a GBLUP animal model with both additive and dominance genetic effects.

For each of the seven traits studied, the following statistical models were considered to describe the vector of performance \mathbf{y} of the 1346 phenotyped fish. Genomic BLUP (GBLUP) considering a genomic relatedness matrix \mathbf{G} (VanRaden, 2007, 2008):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \sum_r b_r \mathbf{F}_r + \mathbf{e} \quad (1)$$

where $\boldsymbol{\beta}$, \mathbf{a} and \mathbf{e} are the vectors of, fixed environmental effects, random genetic additive effects and random residual effects, respectively, explaining the performance of all phenotyped animals. \mathbf{X} and \mathbf{Z} are the incidence matrices for $\boldsymbol{\beta}$ and \mathbf{a} , respectively.

For all traits, the cohort fixed effect was considered (two levels). In addition, the spawning week number was introduced as a covariate nested within the cohort for the SW, EN, EW and PW traits. For SW and EN, the presence of overmature eggs was an additional significant fixed effect considered.

The vector \mathbf{a} includes the breeding values of 1346 phenotyped and genotyped individuals related through the genomic relationship matrix \mathbf{G} .

The regression coefficients b_r of \mathbf{y} on the vectors \mathbf{F}_r account for inbreeding effects in different ways depending on the model considered. First of all, a single regression on the total inbreeding vector \mathbf{F} was considered ($r = 1$ in that case). A multiple regression model was then studied and accounted for $r = 30$ vectors of chromosomal inbreeding coefficients F_{omy} . To finish, a series of nk multiple regression analyses was considered to assess the local inbreeding effects for a specific chromosome k ; in that case, r equals the total number nk of 20 Mb-window regions of chromosome k in addition to the 29 remaining chromosomes. Depending on the chromosome size, nk ranged from 2 (e.g. Omy28) to 4 (e.g. Omy1).

The model accounting for dominance effects (D_GBLUP) was previously described by Vitezica et al. (2016):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \sum_r b_r \mathbf{F}_r + \mathbf{e} \quad (2)$$

where \mathbf{d} corresponds to the dominance breeding values of 1346 phenotyped and genotyped individuals related through the dominance relationship matrix \mathbf{D} . \mathbf{W} is the incidence matrix for \mathbf{d} . The additive \mathbf{G} and dominance \mathbf{D} genomic relationship matrices were built using the *parallelf90* program (Vitezica et al., 2016).

All traits were analysed separately to estimate inbreeding regression coefficients, breeding values and the genetic additive and dominance variances based on the BLUPf90 package (Misztal et al., 2002) using the AIREMLF90 program (Thompson et al., 2005).

3 | RESULTS

3.1 | Evaluation of total inbreeding, recent inbreeding and local inbreeding along the rainbow trout genome

Inbreeding coefficients were based on the number and size of ROHs derived for each individual in the population. The average number of ROHs per individual was 61, with an average ROH size of 5.07 Mb. The smallest number of ROHs observed in an individual

Population	Inbreeding	Mean (%)	SD (%)	Min (%)	Max (%)
Full population (C1 + C2)	F	17.27	3.20	7.92	30.05
	F_{G9}	14.10	3.24	2.56	27.00
	F_{G6}	11.16	3.24	0.96	24.00
	F_{G3}	6.56	2.96	0.00	18.51
Cohort C1	F	16.76	2.99	8.59	26.55
	F_{G9}	13.61	3.01	6.06	23.81
	F_{G6}	10.73	3.05	1.88	21.77
	F_{G3}	6.13	2.72	0.00	15.14
Cohort C2	F	17.87	3.34	7.92	30.05
	F_{G9}	14.67	3.41	2.56	27.00
	F_{G6}	11.65	3.39	0.96	24.00
	F_{G3}	7.07	3.14	0.00	18.51

TABLE 2 Summary statistics of total (F), old (F_{G9}), middle (F_{G6}) and recent (F_{G3}) inbreeding coefficients in the full population and its two cohorts, C1 and C2

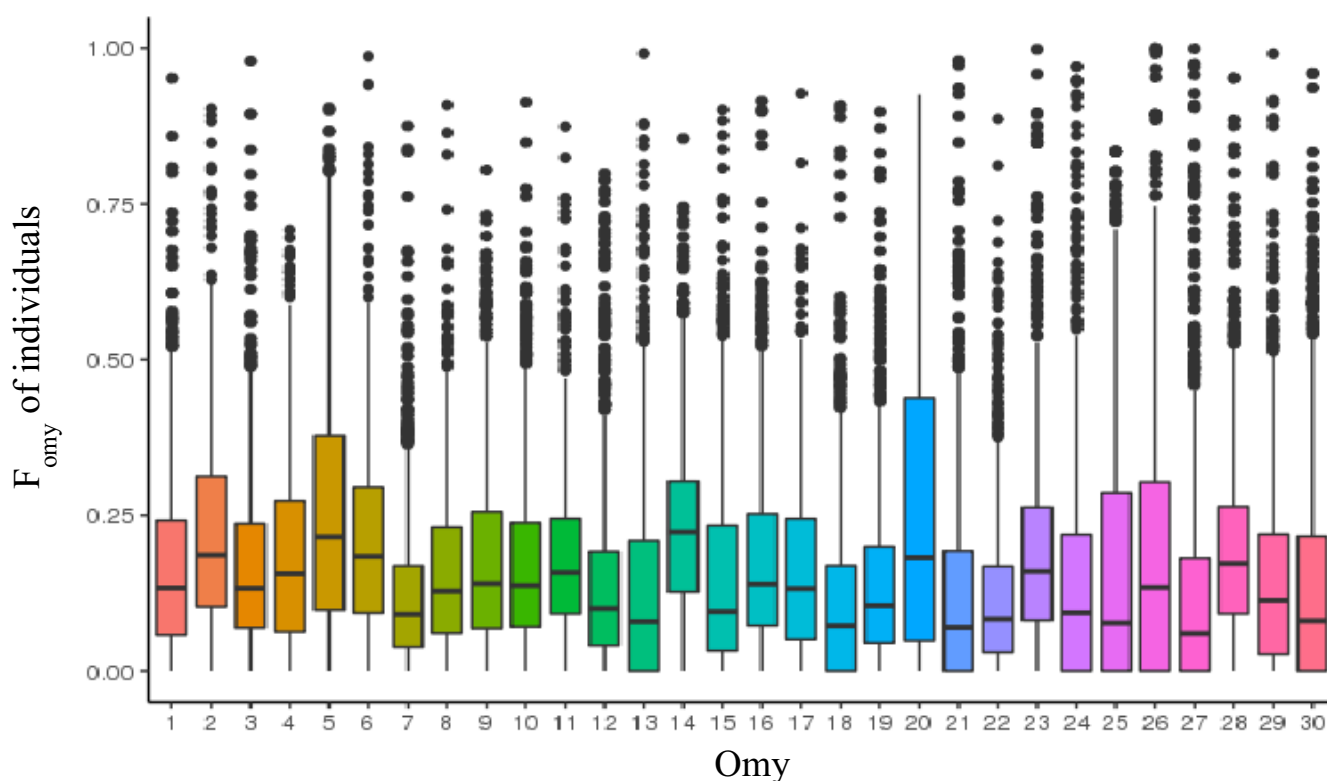


FIGURE 1 Boxplot of inbreeding coefficients for each of the 30 chromosomes in the 1346 female rainbow trout genotyped

was 24 and the largest 113. The proportion of large ROH segments (ROH > 10 Mb) corresponding to recent inbreeding events in the population was 11.5%. Inbreeding coefficient values for the complete genome are shown in Table 2 for the full population and its two cohorts according to the number of ancestral generations included.

The average inbreeding level in the population was about 17%, with individual values ranging from 8% to 30%. We observed that 38% of the genome-wide inbreeding came from the last three generations. In addition, nearly 82% of all inbreeding observed in the C1 and C2 cohorts had been produced since the start of the breeding programme (nine generations earlier). On average, individuals in

cohort C2 were more inbred than those in cohort C1 (+6.6% average inbreeding coefficient), this being linked to a few individuals with extreme inbreeding coefficient values ($F > 25\%$).

As for inbreeding across chromosomes (Figure 1), we noted that chromosomal inbreeding levels (F_{omy}) varied considerably along the genome. While the average F_{omy} was 17.0%, values ranged from 11.6% for F_{18} on Omy18 to 25.1% for F_{20} on Omy20. Some individuals had very high F_{omy} for a particular chromosome, with values reaching almost 100%. However, because the total F of any individual did not exceed 30% (Table 2), an individual with a very high F_{omy} on a specific chromosome would have a low F_{omy} on the other chromosomes. Details on F_{omy} distributions are given in Table S1.

TABLE 3 Phenotypic variance (V_p), heritability (h^2) and dominance ratio (d^2) of female size and reproduction traits estimated under GBLUP and D_GBLUP models accounting for either the genome-wide F coefficient or the 30 chromosomal inbreeding coefficients F_{omy} as covariates (standard errors in brackets). The minimum Akaike information criterion (AIC) value indicates the preferred model for each trait

Traits	GBLUP				D_GBLUP with F as covariate				D_GBLUP with 30 F_{omy} as covariates			
	V_p	h^2	AIC	V_p	h^2	d^2	AIC	Var	h^2	d^2	AIC	
PW	107,000 (5033)	0.34 (0.04)	19,196.4	106,740 (5009)	0.33 (0.04)	0.05 (0.04)	19,196.3	105,440 (4977)	0.31 (0.04)	0.07 (0.05)	18,871.9	
FL	734.36 (34.9)	0.36 (0.04)	12,502.6	733.9 (34.8)	0.35 (0.04)	0.03 (0.04)	12,503.9	722.9 (34.4)	0.33 (0.045)	0.05 (0.05)	12,324.1	
SD	2.074 (0.1)	0.26 (0.04)	11,528.8	2.09 (0.1)	0.26 (0.03)	0.05 (0.01)	11,502.1	2.1 (0.1)	0.26 (0.03)	0.07 (0.01)	10,701.8	
CF	1909.9 (80.2)	0.08 (0.03)	12,296.5	1910.0 (77.1)	0.08 (0.03)	0.00 (0.00)	12,270.2	1919.9 (82.9)	0.09 (0.07)	0.04 (0.06)	12,100.6	
SW	2746.6 (123.8)	0.28 (0.04)	14,321.3	2756.0 (125.1)	0.27 (0.04)	0.13 (0.05)	14,313.3	2766.0 (128.8)	0.26 (0.04)	0.19 (0.06)	14,092.6	
EN	1,760,000 (79,195)	0.27 (0.04)	22,986.1	1,761,000 (79,891)	0.26 (0.04)	0.08 (0.05)	22,984.0	1,775,400 (82,247)	0.25 (0.04)	0.14 (0.06)	22,588.3	
EW	37.54 (1.6)	0.26 (0.04)	8606.0	37.6 (1.6)	0.22 (0.04)	0.08 (0.05)	8600.1	38.1 (1.7)	0.23 (0.04)	0.11 (0.06)	8523.4	

Abbreviations: CF, coelomic fluid weight; EN, egg number; EW, average egg weight; FL, fork length; PW, postspawning body weight; SD, spawning date; SW, spawn weight.

With the sole exception of the moderate correlation estimated at 0.61 between F_{25} and F_{30} , the chromosomal inbreeding coefficients (F_{omy}) were almost uncorrelated (Figure S1).

3.2 | Heritability and dominance ratios for female size and reproduction traits

With the sole exception of coelomic fluid weight with low heritability value, all traits had intermediate heritability values with estimates higher than 30% for female size traits and between 20% and 30% for reproduction traits, for all genetic models (Table 3). Adding a dominance effect into the genomic model (D_GBLUP) resulted in slight reductions in estimates of genetic variance and heritability when compared to GBLUP. However, the correlations between the additive estimated breeding values from G_BLUP and D_GBLUP models were higher than 0.99 for all traits.

Considering the genome-wide inbreeding coefficient as the only covariate in the D_GBLUP model, dominance variance explained about 0%–13% of the phenotypic variance of each trait, the greatest effect being estimated for spawning weight. However, the dominance ratio increased (+25%–107% depending on the trait) when all 30 chromosomal inbreeding coefficients F_{omy} were considered as covariates in the D_GBLUP model. In the same model, dominance variance explained about 4% to almost 19% of the phenotypic variance of each trait, the greatest effect still being estimated for spawning weight. The lowest d^2 estimates concerned female size traits and coelomic fluid weight; they did not differ significantly from 0. Although the d^2 estimates were moderate for spawning date (5%–7% depending on the D_GBLUP model), they differed significantly from 0. Intermediate dominance ratios (ranging from 8% to 14% depending on the D_GBLUP model) were estimated for egg number and average weight.

Because of their better goodness-of-fit (minimum AIC values), the D_GBLUP models were considered to be the reference models to study the impact of inbreeding on performance for all traits, as dominance effects explained at least 5% of phenotypic variance in six out of the seven traits analysed.

3.3 | Estimating the total inbreeding effect on female size and reproduction performance

Regression coefficients b of performance on the total F estimated under the D_GBLUP model are presented in Table 4. Significant effects of total inbreeding were only observed for SD and EW. Increased inbreeding raised SD values, that is delayed spawning. A significant reduction of 3.8% and an increase of 12.3% for EW and SD performance were observed with a 0.1 unit increase in the F coefficient. Very minor effects (not differing significantly from 0) of total inbreeding were estimated on all other traits, but slight negative trends were observed in all cases.

Please note that we use the terms 'positive' or 'negative' to simply qualify increased or decreased trait values because of an increase in

F. When we wish to interpret these trends in terms of fitness, the terms 'favourable' or 'unfavourable' are used.

3.4 | Local variations in inbreeding effects regarding female size and reproduction traits

When considering the chromosome scale (Figure 2), we observed highly variable effects of local inbreeding on all reproduction traits, even those where total inbreeding had no significant effects. Details of the chromosomal inbreeding effects are presented for all traits in Table S2. For a given chromosome, an increase of 0.1 unit in the

TABLE 4 Regression coefficients b for trait performance on the total inbreeding coefficient F in the D_GBLUP model

Trait	b on F (SE)	B^* (%)
PW	-175.2 (317.1)	-0.9
FL	-32.0 (25.3)	-0.6
SD	3.15 (1.4)	12.3
CF	-1.52 (44.8)	-0.4
SW	-67.1 (54.7)	-3.5
EN	-491.4 (1336)	-1.0
EW	-15.1 (6.2)	-3.8

Note: B^* corresponds to the effect for 0.1 unit increase in the inbreeding coefficient and is expressed as a proportion of mean performance. Traits where inbreeding had significant effects are in bold.

Abbreviations: CF, coelomic fluid weight; EN, egg number; EW, average egg weight; FL, fork length; PW, postspawning body weight; SD, spawning date; SW, spawn weight.

inbreeding coefficient corresponded to variations in performance ranging from -2.9% to +3.9% of the trait mean. For some chromosomes, we observed similar negative (e.g. on Omy1, except on coelomic fluid weight) or positive (e.g. Omy20) trends of inbreeding effects on female reproduction traits, while for most of the chromosomes, opposite inbreeding effects were observed on the different traits. Chromosomal inbreeding effects were more variable on coelomic fluid weight (-2.9 to +3.9% of the trait mean, depending on the chromosomes) than on other traits. In particular, the chromosomal inbreeding effects on egg number had a very small amplitude (ranging from -1.1% to 0.9% of the trait mean) and none of them were significant (Figure 2; Table S2). Moreover, we saw a significant and high total inbreeding effect on average egg weight, but with very small chromosomal inbreeding effects ranging from -0.7% to 0.3% of the trait mean. Whatever scale considered (genome-wide, chromosomal or local scale), very small inbreeding effects were seen to affect female fork length (ranging from -0.2% to 0.3%). Despite these very weak inbreeding effects on this trait, some significant effects were observed on certain chromosomes (see Omy17 and Omy28 in Figure 2) as well as more region-specific inbreeding effects (see Omy28_r1 in Figure S2).

A few significant chromosomal inbreeding effects (one to four chromosomes, depending on the traits) were estimated for SW, EW, PW and FL. Although we estimated some significant and positive effects of chromosomal inbreeding on the size traits PW and FL, a majority of negative effects were observed for the reproductive traits SW and EW. None of the chromosomal inbreeding effects estimated for SD, EN and CF were statistically different from zero.

Moreover, we did not observe any strong association between chromosomal inbreeding coefficients (F_{omy}) and the corresponding

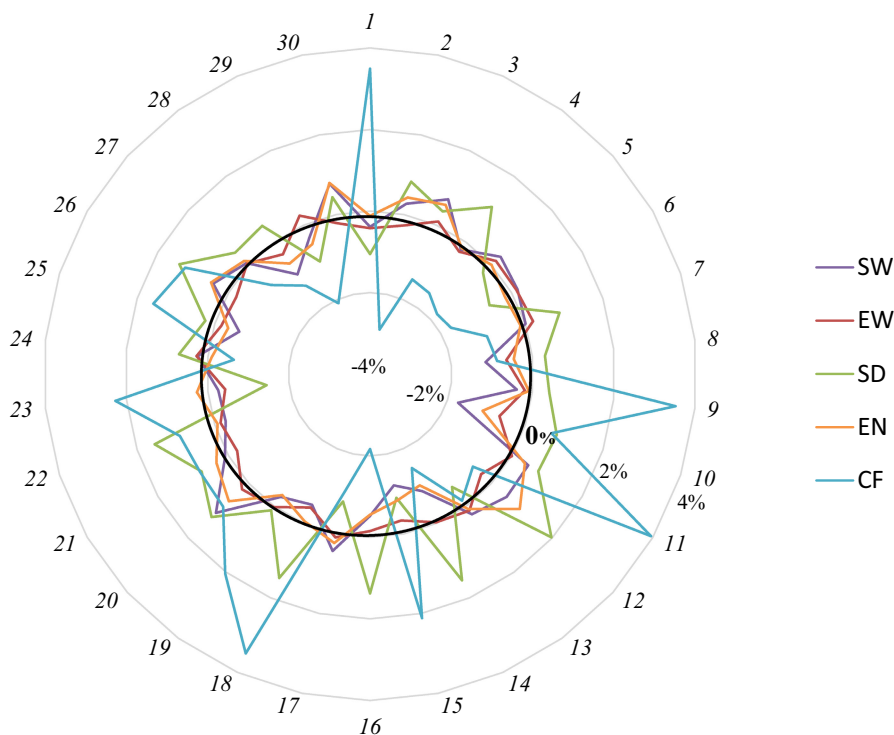


FIGURE 2 Variations across chromosomes in chromosomal inbreeding effects on reproduction traits. Each point outside the circle represents the 30 chromosomes. The axis of graduation corresponds to B^* , that is the effect on trait performance of a variation of +0.1 unit in the chromosomal inbreeding coefficient (expressed as a proportion of mean performance)

regression coefficients on performance (Table 5); except for CF, the correlations were not statistically different from zero. Thus, we did not even observe a clear trend towards stronger inbreeding depression with higher F_{omy} values. Nevertheless, recent chromosomal inbreeding levels ($F_{G3,omy}$) were more markedly negatively associated with the corresponding regression coefficients on performance, particularly for the CF, SW and EN traits.

Zooming at the intra-chromosomal scale, we also observed highly variable effects of local inbreeding on trait performance whatever the chromosome ($F_{omy,r}$ values are presented in Table S3). We focused our analysis on two chromosomes (Figure 3): Omy1 containing several QTLs on reproduction traits (D'Ambrosio et al., 2020) and Omy10 containing an ROH segment (in the 10_r3 region) shared by diverse rainbow trout populations (D'Ambrosio et al., 2019). Along Omy1 (Figure 3a), highly variable local inbreeding effects were observed, in particular for coelomic fluid weight, with positive effects of the first and last 20-Mb regions of the chromosome. A unit increase of 0.1 in the local inbreeding coefficient corresponded to performance variations ranging from -1.2% to +3.8% of the trait means.

TABLE 5 Pearson correlations between chromosomal inbreeding levels F_{omy} and the corresponding regression coefficients b_{omy} for each trait, considering either all generations or the three most recent generations

Traits	$r(F_{omy}, b_{omy})$	$r(F_{G3,omy}, b_{omy,G3})$
PW	-0.098	-0.056
FL	-0.076	-0.097
SD	-0.045	-0.097
CF	-0.217	-0.179
SW	0.113	-0.273
EN	-0.029	-0.313
EW	0.075	-0.037

Note: Correlations in bold differed significantly from 0.

Abbreviations: CF, coelomic fluid weight; EN, egg number; EW, average egg weight; FL, fork length; PW, postspawning body weight; SD, spawning date; SW, spawn weight.

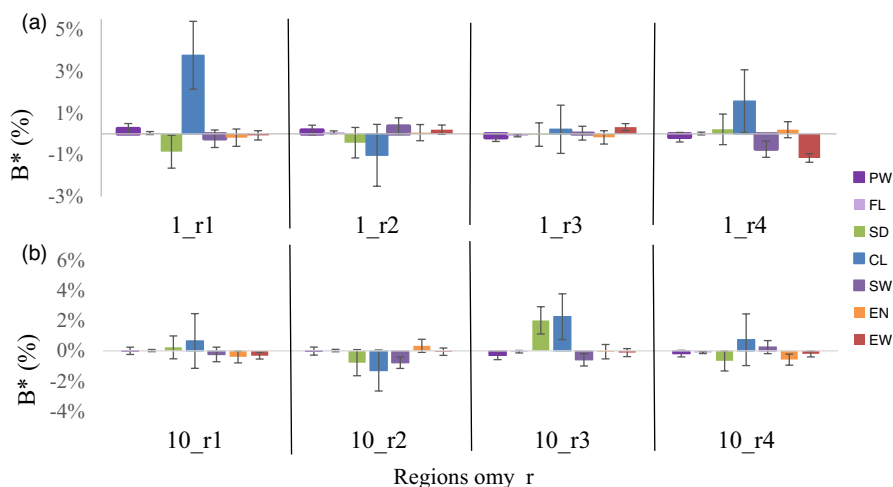
Except for average egg weight, which had a significant positive inbreeding effect on region 1_r3, local inbreeding coefficients exerted no significant effects on the two intermediate regions of Omy1. In addition to the positive inbreeding effect of the 1_r4 region on CF, significant negative effects of inbreeding in this 1_r4 region were observed for EW and SW. We also observed highly variable effects of local inbreeding along the Omy10 chromosome (Figure 3b). The effects were small and not statistically different from 0 in the first and last regions of Omy10. However, local inbreeding in the second and third regions of Omy10 exerted significant effects, but only on spawn weight (decreased performance) and spawning date (increased performance).

3.5 | Variations over generations in inbreeding effects on female size and reproduction traits

Genome-wide inbreeding effects over generations were generally very small, except for SD, SW and EW (Figure 4). We analysed inbreeding effects due to total inbreeding (F) accumulated in the population, but also due to inbreeding events that had accumulated over the past nine generations that we qualified as old inbreeding (F_{G9}), or over just the past three generations that we qualified as recent inbreeding (F_{G3}). Comparing this recent time point to the intermediate point of six generations (F_{G6}), we are able to assess inbreeding events that had occurred between generations 4 and 6. Regarding average egg weight, the effects of recent inbreeding, as well as those accumulated since older generations, were significantly negative. Similar trends were observed for spawn weight although the inbreeding effects did not differ significantly from 0. All temporal inbreeding effects showed clear positive trend on spawning date, but they did not differ significantly from 0, except for the total inbreeding effect. Inbreeding effects mainly appeared to be due to recent inbreeding (F_{G3}) for EW and SW, whereas they appeared to be due to both recent and older inbreeding events for SD.

As for both temporal and chromosomal inbreeding effects along the genome, we observed highly variable effects for all traits (Table S2). We focused our analysis on postspawning body weight and

FIGURE 3 Variations along chromosomes 1 (a) and 10 (b) in local inbreeding effects on female size and reproduction traits. B^* corresponds to the effect of a variation of +0.1 unit in the local inbreeding coefficient (expressed as a proportion of trait means)



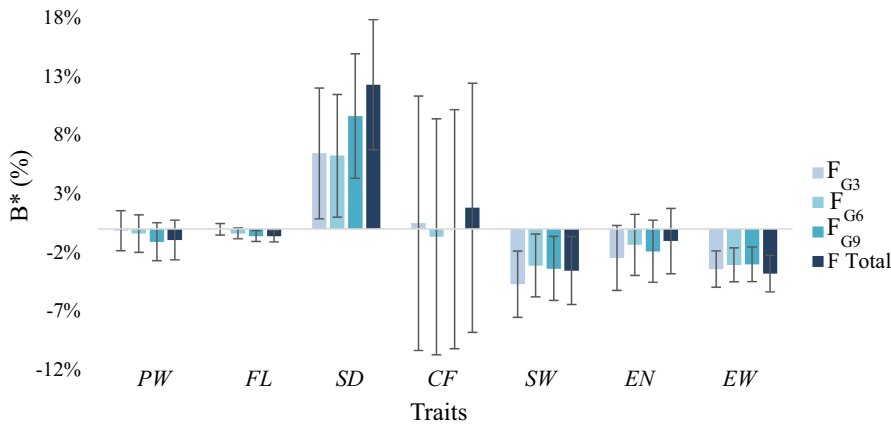


FIGURE 4 Variations over cumulated generations of ancestral inbreeding effects on female size and reproduction traits. B^* corresponds to the effect of a variation of +0.1 unit in the inbreeding coefficient (expressed as a proportion of trait means)

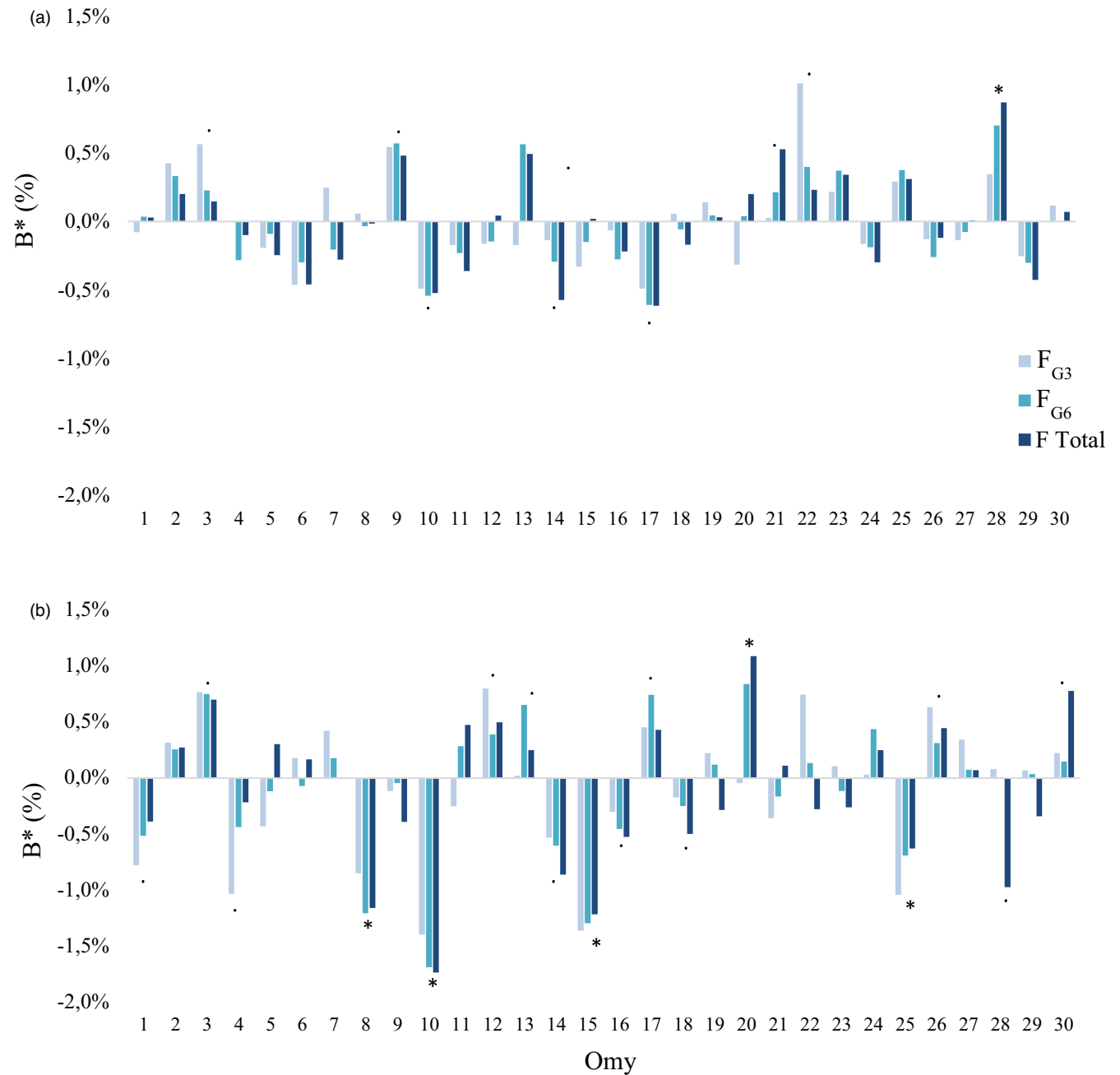


FIGURE 5 Variations in inbreeding effects on traits postspawning weight (a) and spawn weight (b) along chromosomes and over generations. * indicates a significant inbreeding effect; . indicates a inbreeding effect trend

spawn weight (Figure 5) in order to highlight the point that effects could be revealed at temporal and/or chromosomal scales even when total inbreeding effects were null at the genome scale. Indeed, a variation of 0.1 unit in the inbreeding coefficient corresponded to performance variations ranging from -3.9% to +4.6%, depending on the trait, the chromosome and the number of generations used to evaluate the inbreeding effect.

An impact of recent inbreeding (F_{G3}) was considered to be important when, in absolute values, the recent inbreeding effect was equal to or higher than the other temporal inbreeding effects. Important impacts of older inbreeding events were considered when nonzero values of total (F) and old (F_{G9}) inbreeding effects were estimated but the recent inbreeding effect was estimated to be null. Applying these definitions, chromosomal inbreeding effects appeared to be largely explained by recent inbreeding. Indeed, 50% (all traits gathered together, otherwise 40%–67% depending on the trait) of chromosomes with nonzero B^* estimates (above the threshold of 0.5% of the trait mean) indicated an important impact of recent inbreeding, while 33% of chromosomes (21%–50% depending on the trait) displayed important effects of older inbreeding events.

Six scenarios of inbreeding evolution over generations were observed depending on the trait and chromosome. We present them below, from the most common to the least frequent.

The first scenario corresponded to negative effects of recent inbreeding (as shown in Figure 5 for PW on Omy6; SW on Omy4). The second scenario consisted in positive effects of recent inbreeding (e.g. Figure 5 for PW on Omy22; SW on Omy12). The third scenario was associated with negative effects of older inbreeding events (as shown in Figure 5 for SW on Omy28). The fourth scenario was related to positive inbreeding effects in old generations (e.g. Figure 5 for PW on Omy13; SW on Omy30), while the fifth scenario corresponded to observations of negative inbreeding effects accumulated from old to recent generations (as shown in Figure 5 for PW on Omy11; SW on Omy16). The last scenario was related to positive inbreeding effects accumulated from old to recent generations (as shown in Figure 5 for PW on Omy28).

4 | DISCUSSION

Because inbreeding effects are mainly due to dominance genetic effects, we first of all quantify and discuss the importance of dominance effects on performance.

4.1 | Genome-wide dominance effect on female size and reproduction traits

Dominance variance arises from heterozygotes deviating from the mean of the two homozygotes. Loci with overdominant alleles (i.e. inducing a heterozygote advantage) cause high dominance genetic variance in populations (Haldane, 1947), whereas the partial

dominance hypothesis predicts that most genetic variations will be additive when the mutations are incompletely recessive (Houle et al., 1992). Therefore, the partitioning of genetic variance into components of additive variance and dominance variance may help to assess the relative contributions of genetic load and heterozygote advantage to genetic variations in performance.

At least 4% of the phenotypic variances of the female size and reproduction traits were explained by dominance effects in our study. The high dominance ratios (up to 19%) we estimated for reproduction traits (particularly spawn weight and egg number) when compared to size traits may have corresponded to a larger number of loci with overdominant alleles acting on female fecundity.

As far as we know, the literature is very scarce regarding estimates of variance dominance in fish, except for growth and size traits. Based on pedigree estimates and large number of families composed of about 100 full-sibs in three populations of rainbow trout, the d^2 for rainbow trout body weight (Pante et al., 2002) ranged from 0% to 21% depending on the populations. With a similar family structure in four populations of Atlantic salmon, Rye and Mao (1998) indicated that dominance and additive-by-additive epistasis variances were equal to or greater than the additive variance for body weight, with ratios ranging from 2% to 9% and 13% to 16% of phenotypic variance, respectively.

In a recent study based on genomic information on Nile tilapia, Joshi et al. (2020) showed that nonadditive genetic effects were negligible regarding body length and were explained by additive-by-additive epistasis rather than dominance for body weight at harvest. They reported that under pedigree analysis, the additive-by-additive epistasis component was confounded with the dominance component. Our low estimates of dominance ratios for postspawning body weight and female fork length therefore appear to be in relatively good agreement with these previous findings in different fish species.

Regarding d^2 estimates for reproduction traits, we could only compare our estimates to previous studies in terrestrial livestock species where dominance variance ratios were estimated at around 2% for pig litter size (Vitezica et al., 2016, 2018) and 0%–10% for layers on reproduction traits (i.e. egg production, egg colour, egg weight and yolk weight) (Heidaritabar et al., 2016; Misztal & Besbes, 2000).

To the best of our knowledge, our study is the first to have looked at the partitioning of phenotypic variance when performance is corrected for chromosomal inbreeding effects rather than a genome-wide inbreeding effect. The estimates of dominance variance were markedly higher when applying a D-GBLUP model with multiple corrections for all the 30 chromosomal inbreeding coefficients (F_{omy}) compared to simple regression on the genome-wide inbreeding coefficient (Table 3). While this important increase in dominance variance captured by the model using the combined effects of all F_{omy} was mainly associated with a reduction in residual variances for reproduction traits, it was generally associated with a reduction in additive genetic variance among female size traits. The reduction in residual variances

corresponded to a better fit of the model to the data (minimum AIC value) but did not provide any clues regarding the mechanisms that underlie dominance for reproduction traits. The reduction in genetic variance for size traits could perhaps be explained by additive-by-additive epistasis effects captured in the dominance effect when regressing performance on all F_{omy} , which could confirm the pseudo-overdominance hypothesis affecting some loci (Ohta & Kimura, 1969).

4.2 | Genome-wide inbreeding effect on female size and reproduction traits

The mean levels of recent (6.6%) and total (17.3%) inbreeding in the ninth generation of selection of the population under study were slightly higher than previous estimates (5.8% and 16.6%, respectively) in the eight generation (D'Ambrosio et al., 2019). This corresponds to an increase in inbreeding rate ΔF of 0.7% in one generation, which is below the 1% limit targeted under the optimal contribution selection procedure recommended for fish breeding programmes (Skaarud et al., 2014).

Nevertheless, inbreeding levels were quite high when compared to those that can be observed in terrestrial livestock (D'Ambrosio et al., 2019). The main objective of our study was therefore to quantify inbreeding effects at different genomic scales (all genome, chromosome or region-wide), as well as at different time periods. Considering that the positive effects of inbreeding we observed on spawning date were in fact unfavourable because they delayed the moment of spawning, we can sum up our results by saying that unfavourable or null effects of genome-wide inbreeding were estimated on all female size and reproduction traits.

These results corroborate widespread observations of inbreeding depression on fitness-related traits. First of all, the significant negative inbreeding effect observed on average egg weight was undoubtedly unfavourable as large egg size is advantageous in terms of fitness (Einum & Fleming, 1999; Hutchings, 1991); egg size has a direct effect on the body size of offspring at emergence and hence on fry survival. As for spawning date, delayed spawning in a wild context is considered to be favourable to fitness when predators and competition exist between either females for spawning space (Essington et al., 1998) or offspring for territory (Brännäs, 1995). Without predators, however, early spawning may be advantageous because offspring can grow faster than progeny born later because of their access to the best available feeding habitat (Brännäs, 1995; Einum & Fleming, 2000). We can hypothesize that spawning date is always under stabilizing selection when environmental conditions remain unchanged, as it has been suggested by Ford et al. (2006) evaluating long-term changes in a naturally spawning coho salmon population after several decades of intensive hatchery supplementation. Their study showed an optimum run timing observed with fish that returned to their native creek at either end of the run timing distribution producing fewer offspring compared with fish that returned in the middle of the distribution. Under farming and selection

conditions, early spawning is expected to be favourable to fitness because offspring are then larger and have a greater chance of being selected (Chevassus et al., 2004). In addition, delayed spawning leads to larval rearing at warmer temperatures, which is unfavourable for some species of salmonids and may increase fry mortality (Crozier & Zabel, 2006; Smith et al., 2003).

It therefore makes sense to interpret the significant positive inbreeding effect observed on spawning date in our commercial selected line as being unfavourable in terms of fitness.

It should be noted that we did not observe any significant genome-wide effects of inbreeding for size traits, although such effects have generally been observed in studies on salmonids and particularly on rainbow trout. In terms of postspawning weight in rainbow trout, Su et al. (1996) estimated an inbreeding depression of 3.9% per 0.1 unit increase in inbreeding. Pante et al. (2001) estimated a reduction in body weight at harvest that ranged from -1.7% to -5.0% per 0.1 unit increase in inbreeding, depending on the population, the highest values being very similar to earlier estimates for adult body weight (Gjerde et al., 1983; Kincaid, 1983). All these estimations were higher than ours with respect to female postspawning weight. Our estimate (-0.9% per 0.1 unit increase in F) was closer to an estimate of 2-year weight in four Atlantic salmon populations, ranging from -0.6% to -2.6% per 0.1 unit increase in F (Rye & Mao, 1998). Based on genomic information, Waters et al. (2020) found that inbreeding did not affect female weight and fork length in two hatchery lines of Chinook salmon derived from the same source. Inbreeding and its potential effects on growth depend on a variety of factors, including environmental conditions (Armbruster & Reed, 2005), that may explain the variable effects observed across salmonid populations.

Concerning inbreeding effects on reproduction traits, Waters et al. (2020) showed in Chinook salmon that inbreeding did not affect fecundity but delayed spawn timing by 1.75 days per one standard deviation increase in F . While we also observed a significant delay in spawn timing (with an increase of +12% in SD per +0.1 unit increase in F in our trout population), we detected some inbreeding depression effects on spawn weight (-3.5% per +0.1 unit in F) and EN (-1.0% per +0.1 unit in F), although these effects were not statistically significant. The inbreeding effect on egg number was modest when compared to the estimate of -6.1% for EN per +0.1 unit increase unit in F derived from an experimental line of rainbow trout (Su et al., 1996). This negative inbreeding effect is likely to be unfavourable in terms of fitness, because a large number of eggs is expected to produce a large number of juveniles, at least if egg quality is not altered when egg production is high. While a slight negative phenotypic correlation was observed in our study population between average egg weight and egg number, a null genetic correlation had been estimated between them (D'Ambrosio et al., 2020), which corroborates earlier results in the rainbow trout (Su et al., 1996). Assuming that egg weight is a good predictor of egg quality, we can hypothesize that the fall in EN and EW values observed with increased inbreeding levels would be unfavourable to fitness in rainbow trout populations.

All these varying results across studies could be due to species differences, environmental differences (Armbruster & Reed, 2005), different degrees of inbreeding and performance or population management factors. For example, the effect of inbreeding may be undetectable on fitness traits if the rate of inbreeding is slow (Wang et al., 1999).

4.3 | Local inbreeding effects on female size and reproduction performance

Because of the marked variations in inbreeding levels we observed along the genome, it was indeed intuitive to expect stronger and more variable effects of local inbreeding than those of genome-wide inbreeding. The first question we tried to address was therefore whether a higher chromosomal inbreeding coefficient corresponded to greater inbreeding effects on performance? The general answer was that no strong associations were observed between the chromosomal inbreeding level and the corresponding regression coefficient impacting performance. However, for all traits, higher recent chromosomal inbreeding levels ($F_{G3,omy}$) have been associated with negative effects on performance (Table 5), although these trends were only clear for CF, SW and EN. This observation is consistent with the literature, where it has been said that recent inbreeding has more deleterious effects than older inbreeding (Doekes et al., 2019).

When focusing on chromosomal and region-specific inbreeding effects on performance, we observed a mixture of negative and positive effects on each trait. This phenomenon had been also shown in a wild bird population by Botero-Delgadillo et al. (2020) through a correlation study between heterozygosity and fitness traits. Because region-specific inbreeding effects are not exclusively unfavourable, the global unfavourable or null effects observed for genome-wide inbreeding should however correspond to an accumulation of larger numbers and/or stronger impacts of unfavourable local inbreeding effects. For example, F_{10} had a significant negative effect on spawn and average egg weight (Figure 3), but to the best of our knowledge, no QTL has been found to be related to female reproduction traits on Omy10. As for the favourable effects of inbreeding, we should underline the significant positive effect of F_{28} on female length and weight (Figure S2), which may be linked to a putative QTL on Omy28 and which explained about 1.5% of genetic variance in body weight at 18 months in a Chilean rainbow trout line (Neto et al., 2019). Two candidate genes have been identified in this QTL region (located between 20.6 and 21.3 Mb): G protein-coupled receptor-54 like 1 (GPR54L1) and early growth response 1 (EGR1), which plays an important role in growth processes (Aljada et al., 2002; McKee et al., 1997). These results may explain the positive effect of inbreeding on the second region r_2 of Omy28 with respect to female size traits in our French line. Increased size performance (PW and FL) has undoubtedly favourable effects in terms of fitness among farmed populations. Firstly, large body weight is one of the main selection goals for breeders, and secondly, large body size has been positively correlated with fish survival and reproduction in several

studies (Foote, 1990; Huang & Gall, 1990; Pollock et al., 2007; Quinn & Peterson, 1996).

When focusing on local windows of 20 Mb within a chromosome, we also observed strong variations in inbreeding effects for all traits. For example, and as shown in Figure 3, for average egg weight and Omy1 we observed a small but significant favourable effect on region r3 (from 40 to 60 Mb), but a clearly unfavourable inbreeding effect on region r4 (60 to 80 Mb). Regarding the latter case, several QTLs for egg weight have been detected on the same dataset (D'Ambrosio et al., 2020) and deleterious haplotypes have been observed in another French line (Fraslin et al., 2020) that might explain inbreeding depression.

These observations convinced us that the study of local inbreeding is of prime importance to identifying genomic regions and genes with a major impact on inbreeding depression or, on the contrary, regions where inbreeding should be favoured. Further study of inbred regions will help us to highlight selective sweeps and genomic signatures of selection (Aramburu et al., 2020), thus retracing the evolutionary history of the population. Although we have considered the identified QTLs in splitting chromosomes in 20-Mb regions, we probably cut some ROHs into two segments and the QTLs may not locate in the same region as the majority of the ROHs. Therefore, the search for regions where increased inbreeding is associated with impaired phenotypic performance is not trivial. The use of sliding windows of small size rather than fixed 20-Mb windows may overcome this issue and help to better estimate local inbreeding effects.

4.4 | Origin and evolution of inbreeding along the genome

The second objective of our study was to propose underlying mechanisms that might explain our observations regarding temporal and spatial inbreeding effects. All hypotheses relate to underlying mechanisms associated to nonadditive genetic effects, but their long-term implications are not the same. For the overdominance hypothesis, selection would favour heterozygote states at multiple loci, so that mutations would be maintained by mechanisms related to balancing selection. Under the partial dominance hypothesis, the selection of inbred individuals with good performance to become reproducers would purge any deleterious alleles generated by mutations (Kristensen & Sørensen, 2005).

We have identified six scenarios that describe the evolution of local inbreeding effects over time. Here, we will connect these scenarios with some hypotheses including the three that underlie inbreeding depression, that is partial dominance, overdominance and pseudo-overdominance.

Scenarios with inbreeding effects due to recent inbreeding are the most common, particularly when the effects are negative, as has also been shown in cattle (Doekes et al., 2019; Mekanjuola et al., 2020). Our first scenario, corresponding to the unfavourable effects of recent inbreeding, might be explained by any of the three hypotheses underlying inbreeding depression. Under the partial dominance

hypothesis, this means that recent deleterious alleles have not yet been purged. Under the overdominance and pseudo-overdominance hypotheses, inbreeding depression is not subject to any allelic purge (Charlesworth & Willis, 2009). The second scenario corresponds to a favourable effect of recent inbreeding. This can be explained by one favourable allele or underdominance, which corresponds to the disadvantages of a heterozygous genotype regarding reproductive success (the advantage of both homozygous genotypes) that increased the frequency of homozygotes (Reed et al., 2013).

Scenarios 3 and 4 concern the regions where we observed older inbreeding effects on performance which were either unfavourable (scenario 3) or favourable (scenario 4).

Scenario 4 corresponds to either an underdominance phenomenon or to the purging of deleterious alleles (under the partial dominance hypothesis), leading to the fixation of favourable alleles and improved fitness associated with domestication and/or selection phenomena. Purging tends to regress inbreeding depression towards zero because homozygosity in some genomic regions is no longer unfavourable (Boakes et al., 2007). Scenario 3 relates to a lack of purging and local recombination events under the three hypotheses of inbreeding depression, leading to the maintenance of old inbreeding with unfavourable effects on performance (Hedrick, 1994). This lack of effective purging can be explained by alleles, which have only minor unfavourable effects on performance. Indeed, the purging mechanisms are effective in the case of alleles with severely deleterious or even lethal effects (Charlesworth et al., 1990; Wang et al., 1999), which are purged at a faster rate than mildly deleterious alleles. In the case of rare alleles with minor effects, deleterious alleles can be eliminated over hundreds or even thousands of generations (Hedrick, 1994; Hendry et al., 2011; Lande & Schemske, 1985; Larsen et al., 2011). As a direct consequence of purging, populations with a long history of inbreeding are likely to be less affected by inbreeding depression than others with similar levels of inbreeding because they have had more opportunity to purge deleterious alleles than those with recent inbreeding history (Day et al., 2003; Ehiobu et al., 1989). As we observed quite frequently under scenario 3, we can suppose that the bottleneck associated with initial steps in rainbow trout domestication and selection have increased the mutational load of the population and caused an accumulation of deleterious alleles without it being possible to purge them all (Bosse et al., 2019).

The least frequent scenarios with unfavourable (scenario 5) or favourable (scenario 6) effects from old to recent inbreeding correspond to genomic regions subjected to a combination of events observed in the previous scenarios. Recent inbreeding effects are observed, but a lack of purging of deleterious alleles or a low recombination rate in those specific regions may maintain unfavourable old inbreeding effects.

For all the scenarios revealing unfavourable effects of inbreeding, it is not possible to differentiate the three hypotheses underlying inbreeding depression. Nevertheless, the high estimates of dominance variance for some traits may indicate that these traits might be more affected by overdominant loci.

4.5 | Variable evolution of inbreeding effects depending on traits

Inbreeding effects vary considerably along the genome but also from one trait to another because of different histories during domestication and selection processes.

Size traits have been strongly selected since species domestication. The purging of deleterious alleles has already occurred and some favourable alleles for growth function have been selected, thus explaining the positive effects of old inbreeding such as those observed on Omy28. Favourable effects of recent inbreeding are also observed on Omy22, although no QTL seems to have been identified yet for body weight or size on this chromosome, apart one linked to head yield in a different French rainbow trout population line (Blay et al., 2021).

Globally, we found that all chromosomal inbreeding effects were low for female length and weight, as were the estimates of dominance variance which did not exceed 7%. We therefore believe that the partial dominance hypothesis is the best to explain inbreeding depression affecting female size traits.

By contrast, the quantity of coelomic fluid (CF) is not a trait under direct selection. We observed the highest chromosomal inbreeding effects for this trait and a general trend towards an accumulation of recent and older chromosomal inbreeding effects. We can therefore assume that inbreeding depression for this trait is due to an accumulation of deleterious mutations. There are no mentions in the literature regarding the quantity of coelomic fluid, but its composition appears to be linked to progeny survival (Inanan, 2020; Kobayashi et al., 2001). Nevertheless, our CF results need to be taken with caution as CF was a trait derived indirectly through the use of three weight measurements.

For spawning date, we observed both a significant genome-wide inbreeding effect and marked chromosomal inbreeding effects. As a result of the low estimate for the dominance variance ratio (5%), the overdominance hypothesis could not explain a large part of inbreeding effects and the hypothesis of partial dominance should be preferred to explain inbreeding depression.

A significant genome-wide inbreeding effect was also observed for average egg weight, while chromosomal inbreeding effects were null and the dominance ratio was moderate (8%). We therefore postulate an accumulation of small negative inbreeding effects due to slightly unfavourable alleles, which is consistent with the partial dominance hypothesis underlying inbreeding depression for egg weight.

As for the fecundity traits spawn weight and egg number, we observed quite similar profiles of temporal and local inbreeding effects, consistent with their high genetic correlation (D'Ambrosio et al., 2020). Although we did not estimate any significant genome-wide inbreeding effect for these two traits, numerous chromosomes displayed significant and high inbreeding effects. In addition, the dominance variance ratios were quite high (19% for SW and 14% for EN). We can therefore assume that the inbreeding depression observed for fecundity traits could be explained by both the partial dominance and overdominance hypotheses.

5 | CONCLUSION AND IMPLICATIONS

To conclude, we observed very variable inbreeding effects on female size and reproduction traits along the genome and over generations. These results suggest that local inbreeding can strongly impact performance without observing any effects of genome-wide inbreeding. In general, and as might usually be expected, we observed either null or unfavourable effects of genome-wide inbreeding on female size and reproduction traits. However, some favourable effects of local inbreeding were also observed and could be explained by either the selection of favourable homozygotes in the population or an underdominance effect of the heterozygotes. Partial dominance appeared to be the main hypothesis explaining inbreeding depression for all the traits studied, although regarding fecundity traits (spawn weight and egg number), the overdominance hypothesis could also play a significant role in inbreeding depression. Future studies need to focus on the evolution of inbreeding effects near regions under intense directional selection. The limitation of using ROH-based metrics to identify a region associated with inbreeding depression is that a region with long stretches of homozygosity may contain multiple ROH genotypes with variable effects on the phenotype of interest. Further investigations are therefore necessary to determine which specific ROH genotypes result in unfavourable performance.

These new findings offer important keys to the future management of breeding programmes by considering the local genomic scale and not just the genome-wide scale to allow inbreeding in areas where its impact is favourable, and limit it when its effect is deleterious on traits of interest. To achieve this, two options have been advocated to reduce inbreeding depression; purging induced by deliberate inbreeding and genetic rescue.

ACKNOWLEDGEMENTS

The authors warmly thank Zulma Vitezica for giving them access and advice to use her Fortran program to compute the dominance variance matrix. Ana Acin Perez's implication in data acquisition is also heartfully acknowledged. Biological samples and phenotypes were provided by the breeding company Viviers de Sarrance. Their acquisition and the genotyping were funded by the European Maritime and Fisheries Fund and FranceAgrimer (SG-Truite project, no. RFEA47 0016 FA 1000016).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the breeding company 'Viviers de Sarrance' but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The data can be made available for reproduction of the results from Florence Phocas (florence.phocas@inrae.fr) and Ana Acin-Perez (ana@sarrance.com) on request via a material transfer agreement and with permission of the breeding company 'Viviers de Sarrance'.

ORCID

Florence Phocas  <https://orcid.org/0000-0003-1161-3665>

REFERENCES

- Aljada, A., Ghanim, H., Mohanty, P., Kapur, N., & Dandona, P. (2002). Insulin inhibits the pro-inflammatory transcription factor early growth response gene-1 (Egr)-1 expression in mononuclear cells (MNC) and reduces plasma tissue factor (TF) and plasminogen activator inhibitor-1 (PAI-1) concentrations. *The Journal of Clinical Endocrinology & Metabolism*, 87(3), 1419–1422. <https://doi.org/10.1210/jcem.87.3.8462>
- Aramburu, O., Ceballos, F., Casanova, A., Le Moan, A., Hemmer-Hansen, J., Bekkevold, D., Bouza, C., & Martínez, P. (2020). Genomic signatures after five generations of intensive selective breeding: Runs of homozygosity and genetic diversity in representative domestic and wild populations of turbot (*Scophthalmus maximus*). *Frontiers in Genetics*, 11, 296. <https://doi.org/10.3389/fgene.2020.00296>
- Armbruster, P., & Reed, D. H. (2005). Inbreeding depression in benign and stressful environments. *Heredity*, 95(3), 235–242. <https://doi.org/10.1038/sj.hdy.6800721>
- Aulstad, D., & Kittelsen, A. (1971). Abnormal body curvatures of rainbow trout (*Salmo gairdneri*) inbred fry. *Journal of the Fisheries Board of Canada*, 28(12), 1918–1920. <https://doi.org/10.1139/f71-290>
- Blay, C., Haffray, P., Bugeon, J., D'Ambrosio, J., Dechamp, N., Collewet, G., Enez, F., Petit, V., Cousin, X., Corraze, G., Phocas, F., & Dupont-Nivet, M. (2021). Genetic parameters and genome-wide association studies of quality traits characterized using three imaging technologies in rainbow trout, *Oncorhynchus mykiss*. *Frontiers in Genetics*, 12, 219. <https://doi.org/10.3389/fgene.2021.639223>
- Boakes, E., Wang, J., & Amos, W. (2007). An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity*, 98(3), 172–182. <https://doi.org/10.1038/sj.hdy.6800923>
- Bosse, M., Megens, H. J., Derks, M. F., de Cara, Á. M., & Groenen, M. A. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary Applications*, 12(1), 6–17. <https://doi.org/10.1111/eva.12691>
- Botero-Delgado, E., Gilson, C., Mueller, J. C., & Kempnaers, B. (2020). Negative effects of individual heterozygosity on reproductive success in a wild bird population. *Molecular Ecology*, 29(17), 3196–3216. <https://doi.org/10.1111/mec.15553>
- Brännäs, E. (1995). First access to territorial space and exposure to strong predation pressure: A conflict in early emerging Atlantic salmon (*Salmo salar* L.) fry. *Evolutionary Ecology*, 9(4), 411–420. <https://doi.org/10.1007/BF01237763>
- Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, 19(4), 220. <https://doi.org/10.1038/nrg.2017.109>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J., & Sheldon, B. C. (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology*, 18(13), 2746–2765. <https://doi.org/10.1111/j.1365-294X.2009.04247.x>
- Chapuis, H., Pincet, C., & Colleau, J. J. (2016). Optimizing selection with several constraints in poultry breeding. *Journal of Animal Breeding and Genetics*, 133(1), 3–12. <https://doi.org/10.1111/jbg.12178>
- Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetics Research*, 74(3), 329–340. <https://doi.org/10.1017/S0016672399004152>
- Charlesworth, D., Morgan, M. T., & Charlesworth, B. (1990). Inbreeding depression, genetic load and the evolution of outcrossing rates in

- a multi-locus system with no linkage. *Evolution*, 44, 1469–1489. <https://doi.org/10.1111/j.1558-5646.1990.tb03839.x>
- Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), 783–796. <https://doi.org/10.1038/nrg2664>
- Chevassus, B., Quillet, E., Krieg, F., Hollebecq, M.-G., Mambrini, M., Fauré, A., Labbé, L., Hiseux, J.-P., & Vandeputte, M. (2004). Enhanced individual selection for selecting fast growing fish: The “PROSPER” method, with application on brown trout (*Salmo trutta fario*). *Genetics Selection Evolution*, 36(6), 1–19. <https://doi.org/10.1186/1297-9686-36-6-643>
- Craik, J. C. A., & Harvey, S. M. (1984). Egg quality in rainbow trout: The relation between egg viability, selected aspects of egg composition, and time of stripping. *Aquaculture*, 40(2), 115–134. [https://doi.org/10.1016/0044-8486\(84\)90350-8](https://doi.org/10.1016/0044-8486(84)90350-8)
- Crozier, L. G., & Zabel, R. W. (2006). Climate impacts at multiple scales: Evidence for differential population responses in juvenile Chinook salmon. *Journal of Animal Ecology*, 75(5), 1100–1109. <https://doi.org/10.1111/j.1365-2656.2006.01130.x>
- D'Ambrosio, J., Morvezen, R., Brard-Fudulea, S., Bestin, A., Poncet, C., Haffray, P., Dupont-Nivet, M., & Phocas, F. (2020). Genetic architecture and genomic selection of female reproduction traits in rainbow trout. *BMC Genomics*, 21(1), 1–14. <https://doi.org/10.21203/rs.3.rs-21280/v1>
- D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., & Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51(1), 1–15. <https://doi.org/10.1186/s12711-019-0468-4>
- Day, S. B., Bryant, E. H., & Meffert, L. M. (2003). The influence of variable rates of inbreeding on fitness, environmental responsiveness, and evolutionary potential. *Evolution*, 57(6), 1314–1324. <https://doi.org/10.1111/j.0014-3820.2003.tb00339.x>
- DeRose, M. A., & Roff, D. A. (1999). A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution*, 53(4), 1288–1292. <https://doi.org/10.1111/j.1558-5646.1999.tb04541.x>
- Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., & Windig, J. J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. *Genetics Selection Evolution*, 51(1), 54. <https://doi.org/10.1186/s12711-019-0497-z>
- Ehiobu, N. G., Goddard, M. E., & Taylor, J. F. (1989). Effect of rate of inbreeding on inbreeding depression in *Drosophila melanogaster*. *Theoretical and Applied Genetics*, 77(1), 123–127. <https://doi.org/10.1007/BF00292326>
- Einum, S., & Fleming, I. A. (1999). Maternal effects of egg size in brown trout (*Salmo trutta*): Norms of reaction to environmental quality. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1433), 2095–2100. <https://doi.org/10.1098/rspb.1999.0893>
- Einum, S., & Fleming, I. A. (2000). Selection against late emergence and small offspring in Atlantic salmon (*Salmo salar*). *Evolution*, 54(2), 628–639. <https://doi.org/10.1111/j.0014-3820.2000.tb00064.x>
- Escaffre, A. M., & Billard, R. (1979). Évolution de la fécondabilité des ovules de truite arc-en-ciel (*Salmo gairdneri*) laissés dans la cavité abdominale au cours de la période post-ovulatoire. *Bulletin Français de Pisciculture*, 272, 56–70. <https://doi.org/10.1051/kmae:1979012>
- Essington, T. E., Sorensen, P. W., & Paron, D. G. (1998). High rate of redd superimposition by brook trout (*Salvelinus fontinalis*) and brown trout (*Salmo trutta*) in a Minnesota stream cannot be explained by habitat availability alone. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(10), 2310–2316. <https://doi.org/10.1139/f98-109>
- Ferencakovic, M., Solkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution*, 45, 42. <https://doi.org/10.1111/jbg.12012>
- Fessehaye, Y., Komen, H., Rezk, M. A., van Arendonk, J. A., & Bovenhuis, H. (2007). Effects of inbreeding on survival, body weight and fluctuating asymmetry (FA) in Nile tilapia, *Oreochromis niloticus*. *Aquaculture*, 264(1–4), 27–35. <https://doi.org/10.1016/j.aquaculture.2006.12.038>
- Foot, C. J. (1990). An experimental comparison of male and female spawning territoriality in a Pacific salmon. *Behaviour*, 115(3–4), 283–314. <https://doi.org/10.1163/156853990X00617>
- Ford, M. J., Fuss, H., Boelts, B., LaHood, E., Hard, J., & Miller, J. (2006). Changes in run timing and natural smolt production in a naturally spawning coho salmon (*Oncorhynchus kisutch*) population after 60 years of intensive hatchery supplementation. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(10), 2343–2355. <https://doi.org/10.1139/f06-119>
- Fraslin, C., Phocas, F., Bestin, A., Charles, M., Bernard, M., Krieg, F., Dechamp, N., Ciobotaru, C., Hozé, C., Petitprez, F., Milhes, M., Lluch, J., Bouchez, O., Poncet, C., Hocdé, P., Haffray, P., Guiguen, Y., & Quillet, E. (2020). Genetic determinism of spontaneous masculinisation in XX female rainbow trout: New insights using medium throughput genotyping and whole-genome sequencing. *Scientific Reports*, 10, 17693. <https://doi.org/10.1038/s41598-020-74757-8>
- Fu, Y. B., & Ritland, K. (1994). Evidence for the partial dominance of viability genes contributing to inbreeding depression in *Mimulus guttatus*. *Genetics*, 136(1), 323–331. <https://doi.org/10.1093/genetics/136.1.323>
- Gjerde, B., Gunnes, K., & Gjedrem, T. (1983). Effect of inbreeding on survival and growth in rainbow trout. *Aquaculture*, 34, 327–332. [https://doi.org/10.1016/0044-8486\(83\)90212-0](https://doi.org/10.1016/0044-8486(83)90212-0)
- Gomez-Raya, L., Rodríguez, C., Barragán, C., & Silió, L. (2015). Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. *Genetics Selection Evolution*, 47(1), 1–15. <https://doi.org/10.1186/s12711-015-0153-1>
- Haldane, J. B. S. (1947). Parental and fraternal correlations for fitness. *Annals of Eugenics*, 14(1), 288–292. <https://doi.org/10.1111/j.1469-1809.1947.tb02405.x>
- Hedrick, P. W. (1994). Purging inbreeding depression and the probability of extinction: Full-sib mating. *Heredity*, 73(4), 363–372. <https://doi.org/10.1038/hdy.1994.183>
- Heidaritabar, M., Wolc, A., Arango, J., Zeng, J., Settari, P., Fulton, J. E., O'Sullivan, N. P., Bastiaansen, J. W. M., Fernando, R. L., Garrick, D. J., & Dekkers, J. C. (2016). Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. *Journal of Animal Breeding and Genetics*, 133(5), 334–346. <https://doi.org/10.1111/jbg.12225>
- Hendry, A. P., Kinnison, M. T., Heino, M., Day, T., Smith, T. B., Fitt, G., Bergstrom, C. T., Oakeshott, J., Jørgensen, P. S., Zalucki, M. P., Gilchrist, G., Southerton, S., Sih, A., Strauss, S., Denison, R. F., & Carroll, S. P. (2011). Evolutionary principles and their practical application. *Evolutionary Applications*, 4(2), 159–183. <https://doi.org/10.1111/j.1752-4571.2010.00165.x>
- Hill, W. G., & Weir, B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, 93(1), 47–64. <https://doi.org/10.1017/S0016672310000480>
- Houle, D., Hoffmaster, D. K., Assimacopoulos, S., & Charlesworth, B. (1992). The genomic mutation rate for fitness in *Drosophila*. *Nature*, 359(6390), 58–60. <https://doi.org/10.1038/359058a0>
- Howard, J. T., Pryce, J. E., Baes, C., & Maltecca, C. (2017). Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science*, 100(8), 6009–6024. <https://doi.org/10.3168/jds.2017-12787>
- Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three

- autozygosity detection algorithms. *BMC Genomics*, 12(1), 1–15. <https://doi.org/10.1186/1471-2164-12-460>
- Huang, N., & Gall, G. A. E. (1990). Correlation of body weight and reproductive characteristics in rainbow trout. *Aquaculture*, 86(2–3), 191–200. [https://doi.org/10.1016/0044-8486\(90\)90112-Z](https://doi.org/10.1016/0044-8486(90)90112-Z)
- Hutchings, J. A. (1991). Fitness consequences of variation in egg size and food abundance in brook trout *Salvelinus fontinalis*. *Evolution*, 45(5), 1162–1168. <https://doi.org/10.1111/j.1558-5646.1991.tb04382.x>
- Inanan, B. E. (2020). Fertilization rate, motility, lipid peroxidation and pH changes after chilled storage of rainbow trout (*Oncorhynchus mykiss*) eggs and spermatozoa by a RMPI medium. *Aquaculture Research*, 51(1), 222–231. <https://doi.org/10.1111/are.14368>
- Jain, S. K., & Allard, R. W. (1966). The effects of linkage, epistasis, and inbreeding on population changes under selection. *Genetics*, 53(4), 633.
- Jannink, J. L. (2010). Dynamics of long-term genomic selection. *Genetics Selection Evolution*, 42(1), 1–11. <https://doi.org/10.1186/1297-9686-42-35>
- Joshi, R., Meuwissen, T. H., Woolliams, J. A., & Gjøen, H. M. (2020). Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia. *Genetics Selection Evolution*, 52(1), 1. <https://doi.org/10.1186/s12711-019-0522-2>
- Judson, J. L. M., Knapp, C. R., & Welch, M. E. (2018). Age-dependent, negative heterozygosity–fitness correlations and local effects in an endangered Caribbean reptile, *Iguana delicatissima*. *Ecology and Evolution*, 8(4), 2088–2096. <https://doi.org/10.1002/ece3.3826>
- Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity*, 115(1), 63–72. <https://doi.org/10.1038/hdy.2015.17>
- Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9(10), 1205–1218. <https://doi.org/10.1111/eva.12414>
- Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5), 230–241. [https://doi.org/10.1016/S0169-5347\(02\)02489-8](https://doi.org/10.1016/S0169-5347(02)02489-8)
- Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189(1), 237–249. <https://doi.org/10.1534/genetics.111.130922>
- Kincaid, H. L. (1976). Effects of inbreeding on rainbow trout populations. *Transactions of the American Fisheries Society*, 105(2), 273–280. [https://doi.org/10.1577/1548-8659\(1976\)105<273:EOIOR T>2.0.CO;2](https://doi.org/10.1577/1548-8659(1976)105<273:EOIOR T>2.0.CO;2)
- Kincaid, H. L. (1983). Inbreeding in fish populations used for aquaculture. *Aquaculture*, 33(1–4), 215–227. [https://doi.org/10.1016/0044-8486\(83\)90402-7](https://doi.org/10.1016/0044-8486(83)90402-7)
- Kobayashi, H., Ohtomi, M., Sekizawa, Y., & Ohta, N. (2001). Toxicity of coelomic fluid of the earthworm *Eisenia foetida* to vertebrates but not invertebrates: Probable role of sphingomyelin. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 128(3), 401–411. [https://doi.org/10.1016/S1532-0456\(00\)00213-1](https://doi.org/10.1016/S1532-0456(00)00213-1)
- Kristensen, T. N., & Sørensen, A. C. (2005). Inbreeding—lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science*, 80(2), 121–133. <https://doi.org/10.1079/ASC41960121>
- Lande, R., & Schemske, D. W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution*, 39(1), 24–40. <https://doi.org/10.1111/j.1558-5646.1985.tb04077.x>
- Landweber, L., & Dobson, A. (1999). *Genetics and the extinction of species: DNA and the conservation of biodiversity*. Princeton University Press.
- Larsen, L. K., Pelabon, C., Bolstad, G. H., Viken, Å., Fleming, I. A., & Rosenqvist, G. (2011). Temporal change in inbreeding depression in life-history traits in captive populations of guppy (*Poecilia reticulata*): Evidence for purging? *Journal of Evolutionary Biology*, 24(4), 823–834. <https://doi.org/10.1111/j.1420-9101.2010.02224.x>
- Leroy, G. (2014). Inbreeding depression in livestock species: Review and meta-analysis. *Animal Genetics*, 45(5), 618–628. <https://doi.org/10.1111/age.12178>
- Li, M. H., Strandén, I., Tiirikka, T., Sevón-Aimonen, M. L., & Kantanen, J. (2011). A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS One*, 6(11), e26256. <https://doi.org/10.1371/journal.pone.0026256>
- Lieutenant-Gosselin, M., & Bernatchez, L. (2006). Local heterozygosity–fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution*, 60(8), 1658–1668. <https://doi.org/10.1111/j.0014-3820>
- Makanjuola, B. O., Maltecca, C., Miglior, F., Schenkel, F. S., & Baes, C. F. (2020). Effect of recent and ancient inbreeding on production and fertility traits in Canadian Holsteins. *BMC Genomics*, 21(1), 1–15. <https://doi.org/10.1186/s12864-020-07031-w>
- McEvoy, L. A. (1984). Ovulatory rhythms and over-ripening of eggs in cultivated turbot, *Scophthalmus maximus* L. *Journal of Fish Biology*, 24(4), 437–448. <https://doi.org/10.1111/j.1095-8649.1984.tb04814.x>
- McKee, K. K., Tan, C. P., Palyha, O. C., Liu, J., Feighner, S. D., Hreniuk, D. L., Smith, R. G., Howard, A. D., & Van der Ploeg, L. H. T. (1997). Cloning and characterization of two human G protein-coupled receptor genes (GPR38 and GPR39) related to the growth hormone secretagogue and neurotensin receptors. *Genomics*, 46(3), 426–434. <https://doi.org/10.1006/geno.1997.5069>
- McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., ... Wilson, J. F. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83(3), 359–372. <https://doi.org/10.1016/j.ajhg.2008.08.007>
- Misztal, I., & Besbes, B. (2000). Estimates of parental-dominance and full-sib permanent environment variances in laying hens. *Animal Science*, 71(3), 421–426. <https://doi.org/10.1017/S1357729800055326>
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., & Lee, D. H. (2002). BLUPF90 and related programs (BGF90). In Proceedings of the 7th world congress on genetics applied to livestock production (August 19–23, 2002) (Vol. 28, No. 07).
- Mrakovčić, M., & Haley, L. E. (1979). Inbreeding depression in the Zebra fish *Brachydanio rerio* (Hamilton Buchanan). *Journal of Fish Biology*, 15(3), 323–327. <https://doi.org/10.1111/j.1095-8649.1979.tb03612.x>
- Muller, R. H., & Pearson, R. E. (1979). Economic aspect of selection. *Animal Breeding Abstracts*, 47.
- Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., & Quinn, T. P. (2013). Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. *Molecular Ecology*, 22(5), 1295–1309. <https://doi.org/10.1111/mec.12185>
- Neto, R. V. R., Yoshida, G. M., Lhorente, J. P., & Yáñez, J. M. (2019). Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics*, 294(3), 563–571. <https://doi.org/10.1007/s00438-018-1518-2>
- Ohta, T., & Kimura, M. (1969). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics*, 63(1), 229.
- Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., Rexroad, C. E. III, & Moen, T. (2015). The development and characterization of a 57 K single nucleotide polymorphism array for rainbow

- trout. *Molecular Ecology Resources*, 15(3), 662–672. <https://doi.org/10.1111/1755-0998.12337>
- Pante, M. J. R., Gjerde, B., & McMillan, I. (2001). Effect of inbreeding on body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 192(2–4), 201–211. [https://doi.org/10.1016/S0044-8486\(00\)00467-1](https://doi.org/10.1016/S0044-8486(00)00467-1)
- Pante, M. J. R., Gjerde, B., McMillan, I., & Misztal, I. (2002). Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 204(3–4), 383–392. [https://doi.org/10.1016/S0044-8486\(01\)00825-0](https://doi.org/10.1016/S0044-8486(01)00825-0)
- Pollock, K. H., Yoshizaki, J., Fabrizio, M. C., & Schram, S. T. (2007). Factors affecting survival rates of a recovering lake trout population estimated by mark-recapture in Lake Superior, 1969–1996. *Transactions of the American Fisheries Society*, 136(1), 185–194. <https://doi.org/10.1577/T05-317.1>
- Pryce, J. E., Haile-Mariam, M., Goddard, M. E., & Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution*, 46(1), 71. <https://doi.org/10.1186/s12711-014-0071-7>
- Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics*, 13(1), 1–11. <https://doi.org/10.1186/1471-2156-13-70>
- Quinn, T. P., & Peterson, N. P. (1996). The influence of habitat complexity and fish size on over-winter survival and growth of individually marked juvenile coho salmon (*Oncorhynchus kisutch*) in Big Beef Creek, Washington. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(7), 1555–1564. <https://doi.org/10.1139/f96-092>
- Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, 17(1), 230–237. <https://doi.org/10.1046/j.1523-1739.2003.01236.x>
- Reed, F. A., Traulsen, A., & Altrock, P. M. (2013). Underdominance. In S. Maloy & K. Hughes (Eds.), *Brenner's encyclopedia of genetics* (Vol. 2, pp. 247–249). Elsevier.
- Rye, M. A., & Mao, I. L. (1998). Non additive genetic effects and inbreeding depression for body weight in Atlantic salmon (*Salmo salar* L.). *Livestock Production Science*, 57(1), 15–22. [https://doi.org/10.1016/S0301-6226\(98\)00165-1](https://doi.org/10.1016/S0301-6226(98)00165-1)
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15(1), 478. <https://doi.org/10.1186/1471-2164-15-478>
- Sillió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., Óvilo, C., & Fernández, A. I. (2013). Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding and Genetics*, 130(5), 349–360. <https://doi.org/10.1111/jbg.12031>
- Skaarud, A., Woolliams, J. A., & Gjøen, H. M. (2014). Optimising resources and management of genetic variation in fish-breeding schemes with multiple traits. *Aquaculture*, 420, 133–138. <https://doi.org/10.1016/j.aquaculture.2013.10.033>
- Smith, S. G., Muir, W. D., Hockersmith, E. E., Zabel, R. W., Graves, R. J., Ross, C. V., Connor, W. P., & Arnsberg, B. D. (2003). Influence of river conditions on survival and travel time of Snake River subyearling fall Chinook salmon. *North American Journal of Fisheries Management*, 23(3), 939–961. <https://doi.org/10.1577/M02-039>
- Springate, J. R. C., Bromage, N. R., Elliott, J. A. K., & Hudson, D. L. (1984). The timing of ovulation and stripping and their effects on the rates of fertilization and survival to eying, hatch and swim-up in the rainbow trout (*Salmo gairdneri* R.). *Aquaculture*, 43(1–3), 313–322. [https://doi.org/10.1016/0044-8486\(84\)90032-2](https://doi.org/10.1016/0044-8486(84)90032-2)
- Su, G. S., Liljedahl, L. E., & Gall, G. A. E. (1996). Effects of inbreeding on growth and reproductive traits in rainbow trout *Oncorhynchus mykiss*. *Aquaculture*, 142, 139–148. [https://doi.org/10.1016/0044-8486\(96\)01255-0](https://doi.org/10.1016/0044-8486(96)01255-0)
- Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *The American Journal of Human Genetics*, 93(1), 90–102. <https://doi.org/10.1016/j.ajhg.2013.05.003>
- Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2), 301–326. <https://doi.org/10.1534/genetics.112.148825>
- Thompson, R., Brotherstone, S., & White, I. M. (2005). Estimation of quantitative genetic parameters. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459), 1469–1477. <https://doi.org/10.1098/rstb.2005.1676>
- Thrower, F. P., & Hard, J. J. (2009). Effects of a single event of close inbreeding on growth and survival in steelhead. *Conservation Genetics*, 10(5), 1299–1307. <https://doi.org/10.1007/s10592-008-9709-8>
- Toro, M. A., & Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genetics Selection Evolution*, 42(1), 33. <https://doi.org/10.1186/1297-9686-42-33>
- VanRaden, P. M. (2007). Genomic measures of relationship and inbreeding. *Interbull Bulletin*, 37, 33.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Vitezica, Z. G., Reverter, A., Herring, W., & Legarra, A. (2018). Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genetics Selection Evolution*, 50(1), 1–8. <https://doi.org/10.1186/s12711-018-0437-3>
- Vitezica, Z. G., Varona, L., Elsen, J. M., Misztal, I., Herring, W., & Legarra, A. (2016). Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genetics Selection Evolution*, 48(1), 1–8. <https://doi.org/10.1186/s12711-016-0185-1>
- Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, 107, 4–13. <https://doi.org/10.1016/j.tpb.2015.08.006>
- Wang, J., Hill, W. G., Charlesworth, D., & Charlesworth, B. (1999). Dynamics of inbreeding depression due to deleterious mutations in small populations: Mutation parameters and inbreeding rate. *Genetics Research*, 74(2), 165–178. <https://doi.org/10.1017/S0016672399003900>
- Wang, S., Hard, J. J., & Utter, F. (2002). Salmonid inbreeding: A review. *Reviews in Fish Biology and Fisheries*, 11(4), 301–319. <https://doi.org/10.1023/A:1021330500365>
- Waters, C. D., Hard, J. J., Fast, D. E., Knudsen, C. M., Bosch, W. J., & Naish, K. A. (2020). Genomic and phenotypic effects of inbreeding across two different hatchery management regimes in Chinook salmon. *Molecular Ecology*, 29(4), 658–672. <https://doi.org/10.1111/mec.15356>
- Wright, S. (1984). *Evolution and the genetics of populations, volume 3: Experimental results and evolutionary deductions* (Vol. 3). University of Chicago Press.
- Zhang, Q., Calus, M. P., Gulbrandtsen, B., Lund, M. S., & Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics*, 16(1), 88. <https://doi.org/10.1186/s12863-015-0227-7>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Paul, K., D'Ambrosio, J., & Phocas, F. (2022). Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout. *Evolutionary Applications*, 15, 645–662. <https://doi.org/10.1111/eva.13308>

Supplementary Informations

Supplementary tables :

TABLE S1 Summary statistics of chromosomal inbreeding coefficients for F_{omy} values in the full population of rainbow trout (1346 individuals).

F_{omy}	Mean (%)	Q1 (%)	Median (%)	Q3 (%)
F ₁	16.5	5.78	13.3	24.2
F ₂	22.2	10.3	18.6	31.2
F ₃	16.8	6.93	13.3	23.6
F ₄	18.5	6.28	15.6	27.3
F ₅	24.7	9.78	21.5	37.8
F ₆	20.9	9.32	18.4	29.5
F ₇	12.1	3.85	9.09	16.9
F ₈	15.8	6.07	12.8	23.1
F ₉	17.5	6.84	14.0	25.5
F ₁₀	17.7	7.09	13.7	23.8
F ₁₁	18.4	9.19	15.8	24.5
F ₁₂	14.1	4.09	10.0	19.2
F ₁₃	12.9	0	7.92	20.9
F ₁₄	22.2	12.7	22.3	30.5
F ₁₅	15.8	3.25	9.53	23.4
F ₁₆	17.9	7.29	13.9	25.2
F ₁₇	16.4	5.07	13.2	24.4
F ₁₈	11.6	0	7.26	16.9
F ₁₉	14.3	4.49	10.5	19.9
F ₂₀	25.1	4.86	18.2	43.8
F ₂₁	12.8	0	7.00	19.2
F ₂₂	11.8	3.0	8.33	16.8
F ₂₃	19.2	8.14	16.0	26.2
F ₂₄	15.0	0	9.35	21.9
F ₂₅	18.1	0	7.70	28.6
F ₂₆	19.2	0	13.4	30.3
F ₂₇	12.4	0	6.03	18.1
F ₂₈	19.6	9.19	17.25	26.4
F ₂₉	15.2	2.70	11.3	21.9
F ₃₀	15.6	0	8.04	21.6

TABLE S2 Details of chromosomal inbreeding effects (standard errors in brackets) on the seven traits studied and 30 chromosomes (Omy). We considered chromosomal inbreeding effects over generations with total (F_{omy}), middle ($F_{omy,G6}$) and recent ($F_{omy,G3}$) inbreeding coefficients in the full population.

Traits	PW			FL			SD			CF			SW			EN			EW		
	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)	F_{omy} effect (%)	$F_{omy,G6}$ effect (%)	$F_{omy,G3}$ effect (%)
1	0.03 (0.36)	0.03 (0.37)	-0.08 (0.40)	0.01 (0.10)	-0.02 (0.10)	-0.08 (0.12)	-1.05 (1.19)	-1.38 (1.14)	-0.61 (1.31)	-3.49 (2.28)	2.12 (2.25)	2.54 (2.56)	-0.39 (0.63)	-0.51 (0.63)	-0.78 (0.67)	-0.12 (0.62)	-0.45 (0.61)	0.65 (0.65)	-0.42 (0.34)	-0.21 (0.34)	-0.34 (0.37)
2	0.20 (0.33)	0.33 (0.32)	0.42 (0.32)	0.03 (0.09)	0.09 (0.09)	0.14 (0.09)	0.82 (1.08)	0.23 (0.98)	0.26 (1.05)	-2.88 (2.02)	-2.25 (1.91)	-1.01 (2.06)	0.27 (0.58)	0.25 (0.54)	0.31 (0.54)	0.43 (0.56)	0.56 (0.53)	0.52 (0.53)	-0.28 (0.31)	-0.36 (0.29)	-0.33 (0.30)
3	0.15 (0.35)	0.22 (0.34)	0.56 (0.36)	0.05 (0.10)	0.10 (0.10)	0.17 (0.10)	0.37 (1.16)	0.11 (1.07)	0.11 (1.18)	-1.46 (2.19)	-0.63 (2.10)	0.59 (2.27)	0.70 (0.61)	0.75 (0.58)	0.76 (0.59)	0.55 (0.59)	0.58 (0.56)	0.51 (0.59)	0.10 (0.33)	0.12 (0.31)	0.33 (0.33)
4	-0.10 (0.36)	-0.28 (-0.36)	-0.001 (0.38)	-0.09 (0.10)	-0.12 (0.10)	-0.06 (0.11)	1.07 (1.18)	0.94 (1.11)	1.09 (1.26)	-1.54 (2.20)	-2.65 (2.19)	-3.57 (2.45)	-0.22 (0.63)	-0.44 (0.62)	-1.03 (0.64)	-0.18 (0.61)	-0.41 (0.60)	-0.62 (0.63)	-0.29 (0.33)	-0.29 (0.33)	-0.56 (0.35)
5	-0.24 (0.38)	-0.09 (0.37)	-0.19 (0.37)	-0.03 (0.11)	0.01 (0.11)	0.01 (0.11)	-0.28 (1.26)	-0.67 (1.09)	-0.58 (1.20)	-1.79 (2.14)	-0.03 (2.13)	-0.40 (2.34)	0.30 (0.73)	-0.12 (0.63)	-0.43 (0.61)	0.001 (0.69)	-0.40 (0.61)	-0.48 (0.60)	0.15 (0.37)	0.12 (0.34)	-0.02 (0.34)
6	-0.46 (0.34)	-0.29 (0.34)	-0.46 (0.36)	-0.10 (0.10)	-0.05 (0.10)	-0.11 (0.11)	-0.62 (1.13)	-0.77 (1.05)	-0.48 (1.20)	-1.72 (2.14)	-2.91 (2.07)	-3.73 (2.35)	0.17 (0.60)	-0.07 (0.57)	0.18 (0.61)	-0.19 (0.58)	-0.20 (0.56)	-0.10 (0.60)	0.13 (0.32)	-0.07 (0.31)	0.10 (0.33)
7	-0.28 (0.41)	-0.20 (0.43)	0.24 (0.49)	-0.01 (0.12)	0.04 (0.12)	0.15 (0.14)	0.88 (1.26)	0.43 (1.36)	1.79 (1.62)	-0.99 (2.64)	-1.31 (2.70)	1.87 (3.21)	0.01 (0.70)	0.18 (0.71)	0.42 (0.81)	-0.13 (0.69)	0.08 (0.70)	0.18 (0.80)	0.2 (0.38)	0.01 (0.39)	0.22 (0.45)
8	-0.01 (0.40)	-0.03 (0.39)	0.06 (0.40)	-0.04 (0.11)	0.03 (0.11)	0.02 (0.12)	0.31 (1.30)	0.49 (1.23)	0.09 (1.33)	-0.87 (2.48)	-0.40 (2.43)	0.33 (2.60)	-1.16 (0.68)	-1.21 (0.66)	-0.85 (0.67)	-0.46 (0.67)	-0.74 (0.64)	-0.55 (0.66)	-0.64 (0.37)	-0.53 (0.36)	-0.26 (0.37)
9	0.48 (0.35)	0.57 (0.35)	0.54 (0.37)	0.08 (0.10)	0.13 (0.10)	0.12 (0.11)	0.41 (1.16)	-0.46 (1.09)	-0.15 (1.21)	3.52 (2.22)	3.42 (2.16)	4.61 (2.38)	-0.39 (0.61)	-0.05 (0.59)	-0.12 (0.61)	-0.09 (0.60)	0.06 (0.57)	-0.07 (0.66)	-0.19 (0.33)	-0.05 (0.32)	-0.08 (0.34)
10	-0.52 (0.35)	-0.54 (0.35)	-0.49 (0.39)	-0.20 (0.10)	-0.23 (0.10)	-0.18 (0.11)	0.81 (1.15)	0.74 (1.11)	0.38 (1.28)	0.67 (2.15)	-0.36 (2.16)	-0.82 (2.47)	-1.73 (0.61)	-1.68 (0.60)	-1.39 (0.65)	-1.10 (0.60)	-0.93 (0.59)	-0.82 (0.64)	-0.67 (0.33)	-0.72 (0.33)	-0.69 (0.36)
11	-0.36 (0.39)	-0.23 (0.38)	-0.17 (0.40)	-0.06 (0.11)	0.02 (0.11)	0.04 (0.12)	0.76 (1.28)	0.60 (1.19)	0.02 (1.31)	3.94 (2.46)	3.85 (2.35)	3.93 (2.57)	0.47 (0.67)	0.28 (0.64)	-0.25 (0.66)	0.38 (0.65)	0.20 (0.62)	-0.20 (0.65)	0.01 (0.36)	0.08 (0.35)	0.16 (0.37)
12	0.04 (0.37)	-0.14 (0.37)	-0.16 (0.39)	0.02 (0.11)	0.01 (0.11)	0.01 (0.11)	1.98 (1.22)	1.75 (1.14)	1.74 (1.27)	-0.61 (2.25)	-0.31 (2.18)	0.25 (2.39)	0.50 (0.65)	0.39 (0.64)	0.79 (0.65)	0.94 (0.63)	0.69 (0.62)	1.05 (0.63)	-0.34 (0.35)	-0.19 (0.34)	-0.17 (0.35)
13	0.49 (0.28)	0.56 (0.29)	-0.17 (0.39)	0.16 (0.08)	0.17 (0.08)	0.01 (0.11)	-0.60 (0.94)	-1.16 (0.93)	-1.63 (1.29)	-0.17 (1.83)	0.21 (1.87)	0.07 (2.65)	0.25 (0.48)	0.65 (0.49)	0.02 (0.64)	0.08 (0.48)	0.30 (0.48)	-0.10 (0.64)	0.13 (0.26)	0.26 (0.27)	0.13 (0.36)
14	-0.57 (0.37)	-0.29 (0.36)	-0.14 (0.37)	-0.17 (0.11)	-0.17 (0.10)	-0.12 (0.11)	1.53 (1.21)	1.14 (1.13)	0.89 (1.21)	-1.48 (2.22)	-1.08 (2.18)	-3.11 (2.36)	-0.86 (0.62)	-0.60 (0.59)	-0.53 (0.60)	-1.02 (0.61)	-0.70 (0.58)	-0.64 (0.60)	-0.03 (0.34)	-0.12 (0.32)	-0.18 (0.33)
15	0.02 (0.33)	-0.15 (0.33)	-0.33 (0.39)	-0.07 (0.09)	-0.12 (0.10)	-0.15 (0.11)	0.90 (1.07)	-0.39 (1.02)	-0.86 (1.28)	2.11 (1.94)	1.61 (1.91)	1.54 (2.41)	-1.21 (0.56)	-1.29 (0.55)	-1.36 (0.64)	-0.83 (0.54)	-0.83 (0.54)	-1.14 (0.63)	-0.34 (0.30)	-0.36 (0.30)	-0.18 (0.35)
16	-0.22 (0.35)	-0.28 (0.36)	-0.06 (0.38)	-0.15 (0.10)	-0.16 (0.10)	-0.13 (0.11)	1.37 (1.16)	1.70 (1.12)	2.16 (1.26)	-2.16 (2.20)	-1.96 (2.20)	-1.24 (2.45)	-0.52 (0.61)	-0.46 (0.60)	-0.30 (0.64)	-0.56 (0.59)	-0.56 (0.59)	-0.61 (0.63)	0.15 (0.33)	-0.05 (0.33)	0.20 (0.35)
17	-0.61 (0.37)	-0.61 (0.38)	-0.49 (0.42)	-0.25 (0.11)	-0.24 (0.11)	-0.19 (0.12)	-0.81 (1.21)	-0.95 (1.21)	-1.45 (1.38)	-1.08 (2.31)	-2.54 (2.47)	-1.65 (2.85)	0.43 (0.63)	0.74 (0.64)	0.45 (0.69)	0.23 (0.61)	0.33 (0.63)	0.07 (0.68)	0.09 (0.34)	0.26 (0.35)	0.34 (0.38)
18	-0.17 (0.35)	-0.06 (0.36)	0.05 (0.38)	-0.01 (0.10)	0.04 (0.10)	0.06 (0.11)	1.47 (1.17)	1.63 (1.15)	1.27 (1.25)	3.49 (2.25)	2.09 (2.25)	2.82 (2.43)	-0.50 (0.60)	-0.25 (0.60)	-0.17 (0.62)	-0.05 (0.59)	0.38 (0.59)	0.36 (0.62)	-0.42 (0.33)	-0.64 (0.33)	-0.57 (0.35)
19	0.03 (0.36)	0.04 (0.36)	0.14 (0.40)	-0.10 (0.10)	-0.07 (0.10)	-0.05 (0.12)	0.12 (1.17)	-0.41 (1.14)	-0.27 (1.31)	2.05 (2.24)	2.54 (2.22)	2.55 (2.55)	-0.28 (0.61)	0.12 (0.61)	0.22 (0.62)	-0.34 (0.60)	-0.27 (0.60)	0.27 (0.65)	0.02 (0.33)	0.35 (0.33)	0.12 (0.36)
20	0.20 (0.29)	0.04 (0.28)	-0.31 (0.29)	0.07 (0.08)	0.05 (0.08)	-0.04 (0.08)	1.23 (0.94)	1.50 (0.87)	0.51 (0.95)	0.85 (1.58)	1.74 (1.65)	2.72 (1.85)	1.08 (0.51)	0.84 (0.47)	-0.05 (0.48)	0.66 (0.49)	0.27 (0.46)	-0.22 (0.47)	0.22 (0.27)	0.38 (0.26)	0.18 (0.26)
21	0.53 (0.30)	0.21 (0.32)	0.02 (0.39)	0.09 (0.09)	0.03 (0.09)	0.02 (0.11)	0.77 (0.99)	0.32 (1.01)	-0.06 (1.28)	0.81 (1.97)	-0.62 (2.08)	-1.96 (2.63)	0.11 (0.52)	-0.16 (0.54)	-0.36 (0.64)	0.35 (0.51)	0.30 (0.53)	-0.05 (0.63)	-0.24 (0.28)	-0.38 (0.29)	-0.34 (0.35)
22	0.23 (0.37)	0.39 (0.41)	1.01 (0.51)	0.08 (0.11)	0.15 (0.12)	0.30 (0.15)	1.55 (1.24)	1.30 (1.32)	2.93 (1.69)	0.90 (2.42)	1.35 (2.62)	-0.37 (3.31)	-0.28 (0.63)	0.13 (0.68)	0.74 (0.83)	-0.07 (0.62)	0.37 (0.67)	1.33 (0.83)	-0.15 (0.35)	-0.06 (0.38)	-0.45 (0.47)
23	0.34 (0.31)	0.37 (0.30)	0.21 (0.33)	0.05 (0.09)	0.04 (0.09)	-0.01 (0.09)	-1.45 (1.02)	-1.36 (0.96)	-0.26 (1.08)	2.28 (1.95)	2.19 (1.88)	3.61 (2.10)	-0.26 (0.52)	-0.12 (0.50)	0.10 (0.54)	0.27 (0.52)	0.42 (0.50)	0.66 (0.53)	-0.43 (0.28)	-0.35 (0.28)	-0.22 (0.30)
24	-0.30 (0.28)	-0.19 (0.28)	-0.16 (0.29)	-0.06 (0.08)	-0.03 (0.08)	-0.01 (0.08)	0.71 (0.93)	0.57 (0.88)	0.48 (0.94)	-0.64 (1.76)	-0.43 (1.74)	-0.40 (1.85)	0.25 (0.48)	0.43 (0.47)	0.03 (0.47)	-0.09 (0.47)	0.003 (0.46)	-0.20 (0.47)	0.28 (0.26)	0.25 (0.26)	0.17 (0.26)
25	0.31 (0.30)	0.37 (0.29)	0.29 (0.27)	0.07 (0.08)	0.10 (0.08)	0.13 (0.08)	0.25 (0.97)	0.23 (0.90)	-0.20 (0.89)	1.59 (1.80)	2.43 (1.74)	-0.27 (1.70)	-0.63 (0.51)	-0.69 (0.48)	-1.04 (0.45)	-0.34 (0.50)	-0.44 (0.47)	-0.63 (0.45)	-0.16 (0.27)	-0.11 (0.26)	-0.29 (0.25)
26	-0.12 (0.25)	-0.26 (0.25)	-0.13 (0.28)	-0.04 (0.07)	-0.08 (0.07)	0.003 (0.08)	1.23 (1.54)	1.05 (1.52)	0.23 (1.84)	1.23 (1.54)	1.05 (1.52)	0.23 (1.84)	0.44 (0.43)	0.31 (0.42)	0.63 (0.47)	0.50 (0.42)	0.30 (0.41)	0.55 (0.46)	-0.21 (0.23)	-0.14 (0.23)	-0.09 (0.26)
27	0.01 (0.29)	-0.07 (0.30)	-0.14 (0.33)	-0.01 (0.09)	-0.04 (0.09)	-0.06 (0.09)	0.45 (0.96)	0.83 (0.95)	1.20 (1.08)	-0.73 (1.84)	0.29 (1.86)	0.53 (2.10)	0.07 (0.51)	0.07 (0.52)	0.34 (0.55)	0.15 (0.50)	0.08 (0.50)	0.32 (0.54)	0.05 (0.27)	0.11 (0.28)	0.23 (0.30)
28	0.87 (0.34)	0.70 (0.32)	0.34 (0.35)	0.30 (0.10)	0.28 (0.09)	0.16 (0.10)	0.50 (1.11)	1.05 (1.01)	-0.01 (1.17)	-1.32 (2.08)	-3.00 (1.95)	-3.93 (2.27)	-0.97 (0.56)	0 (0)	0.08 (0.58)	-0.64 (0.56)	-0.11 (0.52)	0.16 (0.58)	-0.36 (0.31)	-0.15 (0.29)	-0.28 (0.32)
29	-0.42 (0.30)	-0.30 (0.31)	-0.25 (0.36)	-0.08 (0.09)	-0.05 (0.09)	0.004 (0.10)	-0.98 (1.00)	-0.93 (1.01)	-1.37 (1.19)	-2.10 (1.95)	-2.07 (2.01)	-1.65 (2.37)	-0.34 (0.51)	0.03 (0.52)	0.07 (0.59)	-0.51 (0.50)	-0.19 (0.51)	-0.11 (0.59)	0.26 (0.28)	0.31 (0.29)	0.16 (0.33)
30	0.07 (0.33)	-0.01 (0.33)	0.11 (0.32)	0.02 (0.10)	-0.02 (0.10)	-0.02 (0.09)	0.44 (1.10)	0.18 (1.05)	-0.12 (1.06)	-1.03 (2.10)	-2.78 (2.06)	-1.97 (2.01)	0.77 (0.57)	0.15 (0.56)	0.22 (0.53)	0.80 (0.56)	0.36 (0.55)	0.09 (0.53)	-0.22 (0.31)	-0.34 (0.31)	-0.10 (0.29)

TABLE S3 Summary statistics of local inbreeding $F_{\text{omy_region}}$ values, for regions of 20Mb on specific omy1, omy10 and omy28, in the 1346 rainbow trout genotyped.

F_{omy_region}	Mean (%)	Q1 (%)	Median (%)	Q3 (%)
F _{1_r1}	11.9	0	0	15.0
F _{1_r2}	16.5	0	7.1	27.5
F _{1_r3}	20.3	0	5.8	26.4
F _{1_r4}	17.4	0	9.8	24.5
F _{10_r1}	13.4	0	9.6	16.2
F _{10_r2}	21.1	0	11.4	27.9
F _{10_r3}	20.2	0	16.8	29.0
F _{10_r4}	6.9	0	0	0
F _{28_r1}	13.8	0	6.1	22.4
F _{28_r2}	25.2	11.1	23.8	37.8

Supplementary figures :

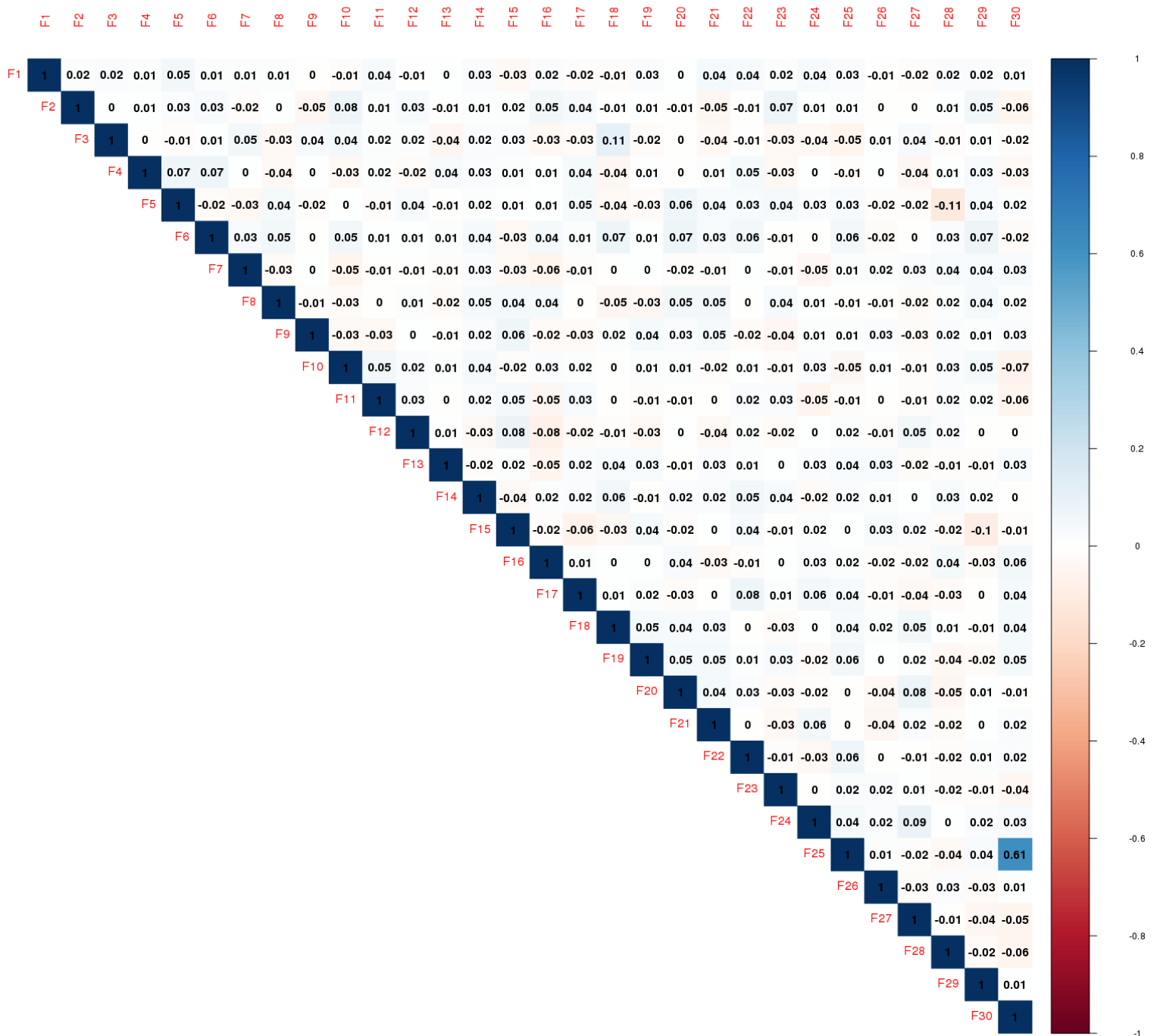


FIGURE S1 Correlations between chromosomal inbreeding coefficients (F_{omy}). Except for the correlation between F_{25} and F_{30} (p -value $> 2.26 \times 10^{-16}$), all correlations were not significant (p -value > 0.05).

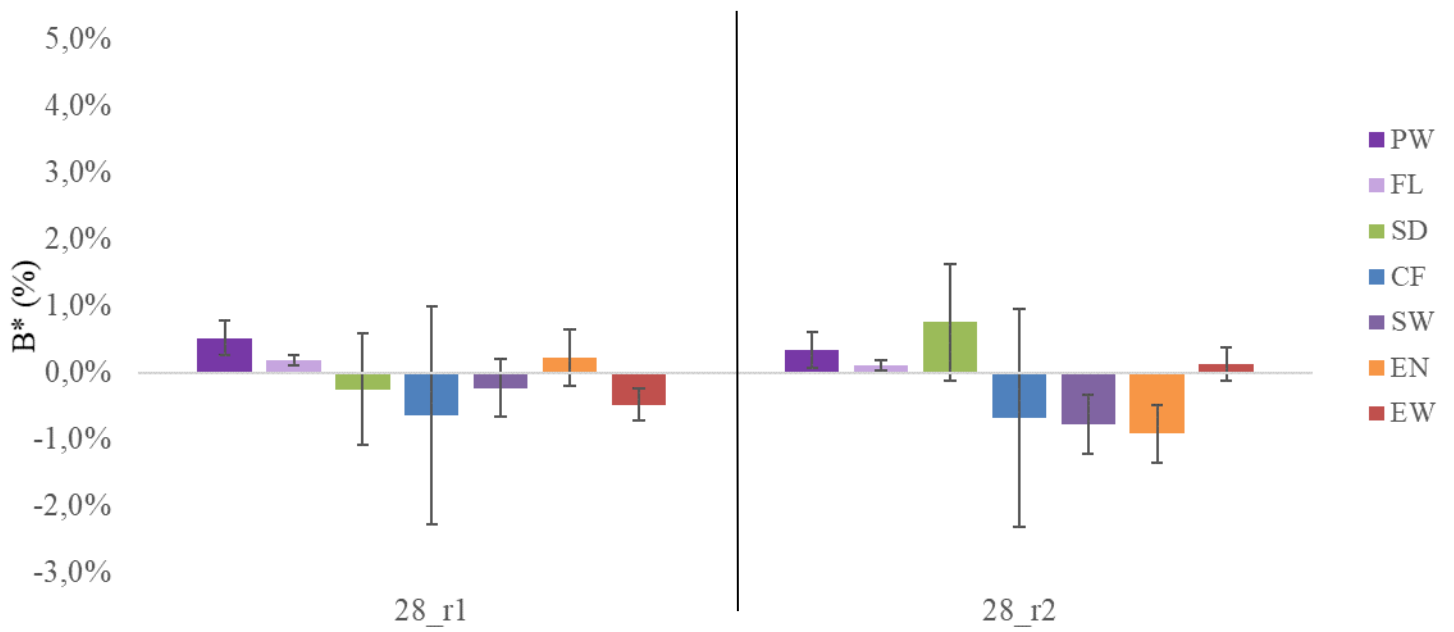


FIGURE S2 Variations along chromosome 28 in local inbreeding effects on female size and reproduction traits.

Chapitre 4

Détection de signatures de sélection positive ou équilibrante communes à quatre populations domestiquées de truite arc-en-ciel

Au fil du temps les populations évoluent, s'adaptent à leur environnement, laissant des traces spécifiques dans le génome. Alors que de nombreuses espèces de poissons font encore l'objet d'un processus de domestication, la domestication de la truite arc-en-ciel est plus ancienne (cf Chapitre 1, partie 1.1.2) et des programmes de sélection existent depuis déjà une trentaine d'années en France. L'objectif de cette étude est d'identifier des régions génomiques sous signatures de sélection dans différentes populations de truites, soit simplement domestiquées, soit soumises à sélection artificielle depuis au moins 8 générations.

Dans ce dernier chapitre présentant les travaux de thèse, nous avons cherché par la détection de traces de sélection communes à différentes populations domestiquées à répondre aux questions suivantes : **Existe-t-il des régions du génome des truites arc-en-ciel, très homozygotes, sous sélection positive conservées entre plusieurs populations domestiquées ? Ou bien des régions hétérozygotes, sous sélection équilibrante également partagées entre plusieurs populations domestiquées ? Ces régions sont-elles simplement en lien avec des fonctions essentielles dans la survie et la reproduction des truites ? Ou sont-elles liées fortement au processus de domestication des populations ?**

Nous avons utilisé deux méthodes, basées sur les îlots de consanguinité (ROH) et l'Homozygotie d'Haplotype Étendue (EHH), pour détecter des signatures de sélection récente (positive ou équilibrante) communes à quatre populations de truite arc-en-ciel domestiquées (originaires d'Amérique du Nord et de France). Nous avons ensuite identifié les fonctions des

gènes contenus dans les régions fortement homozygotes ou hétérozygotes, communes à ces quatre populations et avons regardé si ces gènes avaient déjà été identifiées comme des gènes clés dans des processus de domestication ou sélection d'animaux d'élevage.

Conclusion du Chapitre 4

Treize régions ont été identifiées sous sélection positive ou équilibrante dans les quatre populations de truites arc-en-ciel étudiées (originaire d'Amérique du Nord et de France). Alors que dans les régions fortement homozygotes, de nombreux gènes dont les fonctions sont variables (comportement, croissance, immunité, reproduction, développement embryonnaire et organisation cellulaire) et touchent toutes les fonctions liées à la survie et la domestication de l'espèce ont été identifiées sous sélection positive, les gènes identifiés sous sélection équilibrante (présents dans les régions hétérozygotes) semblent avant tout liés à la seule à la *fitness* des individus (organisation cellulaire, développement embryonnaire et immunité).

Cette étude nous permet de confirmer l'importance d'un large ensemble de gènes déjà détectés comme étant sous sélection positive dans diverses espèces animales et d'identifier de nouveaux gènes candidats importants pour la survie et la domestication chez la truite arc-en-ciel. Ces résultats doivent permettre d'établir une première carte génomique des régions où la diversité génétique est soit à éviter, soit à privilégier dans les populations de truite arc-en-ciel domestiquées en particulier quand elles sont soumises à sélection artificielle.

Les différents résultats obtenus dans ce chapitre ont été présentés lors d'une communication orale au congrès *Génétique en Aquaculture* (GIA) en 2022 et ils ont également fait l'objet d'un projet d'article « *Genome-wide detection of positive and balancing selection signatures shared by four domesticated *Oncorhynchus mykiss* populations* » présenté ici et soumis à *Molecular ecology*.

Genome-wide detection of positive and balancing selection signatures shared by four domesticated rainbow trout populations (*Oncorhynchus mykiss*)

Paul, K.¹, Restoux, G.¹, Phocas, F.¹

1. Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

Corresponding author: Florence Phocas - florence.phocas@inrae.fr

Acknowledgements

We thank the two breeding companies “Viviers de Sarrance” and “Milin Nevez” that allow us to use their HD genotypes to perform the study. This study was partly supported by the European Maritime and Fisheries Fund and FranceAgrimer (Hypotemp project, n° P FEA470019FA1000016).

Abstract

Evolutionary processes leave footprints across the genome over time. Highly homozygous regions may correspond to positive selection of favourable alleles, while maintenance of heterozygous regions may be due to balancing selection phenomena. We analyzed 176 genomes coming from 20 sequenced US fish and 156 fish from three different French lines that were genotyped using a HD Axiom Trout Genotyping 665K SNP Array. Using methods based on either Run of Homozygosity or Extended Haplotype Homozygosity, we detected selection signals in four domesticated rainbow trout populations. Nine genomic regions composed of 253 genes, mainly located on chromosome 2 but also on chromosomes 12, 15, 16, and 20, were identified under positive selection in all four populations. In addition, four heterozygous regions containing 29 genes putatively under balancing selection were also shared by the four populations and located on chromosomes 10, 13, and 19. Whatever the homozygous or heterozygous nature of the region, we always found some genes highly conserved among vertebrates due to their critical roles in cellular and nuclear organisation, embryonic development or immunity. We identify new promising candidate genes involved in rainbow trout fitness, as well as genes already detected under positive selection in other fishes (*auts2*, *atp1b3*, *zp4*, *znf135*, *igf-1 α* , *brd2*, *col9a2*, *mrap2*, *pbx1*, *emilin-3*). These findings represent a genome-wide map of signatures of selection common over rainbow trout populations, which is the foundation to understand the processes in action and to identify what kind of diversity should be preserved, or conversely avoided in breeding programs, in order to maintain or improve essential biological functions in domesticated rainbow trout populations.

Keywords: Runs of Homozygosity, Extended Haplotype Homozygosity, domestication, fitness, selection, fish.

1| Introduction

Any population, whether animal or plant, wild or domesticated, evolved through continuous and cumulative changes over time (Wright, 1931). It relies on various evolutionary forces, mutation, migration, selection, and genetic drift, whose relative effects may vary depending on population history and structure. For example, genetic drift is more substantial when the effective population size is small and randomly induces fixation of alleles, which may lead to degeneration and extinction due to the fixation of deleterious alleles in small populations (Smith & Haigh, 1974). When modifications of environmental conditions occur, allele frequencies will change to a new relevant equilibrium, as a result of natural selection. Indeed, favorable alleles in a particular environment due to either new mutations or standing variation, will be positively selected. In wild populations, favourable alleles are generally affecting fitness through individual survival, mating, or fertility (East, 1918; Fisher, 1958). Natural selection can also act by negative (or purifying) selection that hinders the spread of deleterious alleles (Charlesworth et al., 1995). These two processes tend to reduce the genetic diversity at the target genes but had different effect on the genome, positive selection leading to stronger selection signatures (selective sweep) than negative one. Conversely, the population's polymorphism can be actively maintained in some rare genomic regions through balancing selection that keeps an equilibrium in the frequencies of alleles. The two main biological causes of balancing selection are heterozygote advantage at a single locus, known as overdominance effect, and frequency-dependent selection with a rare-allele advantage that tends to restore a frequency equilibrium between alleles at the population level (Charlesworth, 2006, Fijarczyk & Babik, 2015).

Domestication is the evolutionary process of genetic adaptation over generations of a wild population to handling by humans and breeding in captive environments (Darwin, 1859, 1868; Price, 1984). During domestication, humans exert artificial selection pressure on the initial population by choosing and organizing the reproduction of the most adapted individuals to cohabitation or more globally to those whose aptitudes correspond the best to their expectations (Price, 1999; Russell, 2002), such as a less fearfulness of humans (Price, 2002; Harri et al., 2007). Domestication induces severe genetic bottlenecks due to the selection and reproduction of only a few adapted animals from the wild population. Thus, many genetic evolutionary processes, such as selection, genetic drift, and inbreeding, have a significant role in the evolution of farmed animal populations (Helmer, 1992; Mignon-Grasteau et al., 2005). The domestication process affects life history traits due to changes in morphological, physiological, reproductive, behavioural, and immune functions (Mignon-Grasteau et al., 2005; Pulcini et al.,

2013; Milla et al., 2021 for review in fishes) compared to their wild relatives (Darwin, 1859, 1868). Wilkins et al. (2014) suggest that these specific modifications, called domestication syndrome, may be due to mild deficit of neural-crest cells during embryonic development in domesticated animals. In addition, both natural and artificial selection in domesticated species leaves footprints across the genome, known as selection signatures, which can point to regions harboring essential genes for domestication or natural fitness (Dobney & Larson, 2006; Qanbari & Simianer, 2014; Wright, 2015).

Compared to domestication in terrestrial animals (Mignon-Grasteau et al., 2005), fish domestication is recent and was first documented with carp about 2,000 years ago. The precise date and location (Neolithic China or at the Roman period in Central and East Europe) of the carp domestication are still debated (Balon, 1995; Balon, 2004). However, most farmed fish species have only been domesticated since the last century. The rainbow trout is native to the Pacific drainages of North America and to Kamchatka in Russia and its domestication started in the 1870s in California (Hershberger, 1992; Gall & Crandel, 1992). It was then introduced in Western Europe at the beginning of the 20th century (Fabrice, 2018).

Numerous studies have been carried out over the last ten years to detect signatures of selection in farmed fish species (Channel Catfish: Sun et al., 2014; Atlantic salmon: Mäkinen et al., 2015; Gutierrez et al., 2016; Liu et al., 2017a; Pritchard et al., 2018; López et al., 2019; Carp: Su et al., 2018; Nile Tilapia: Hong et al., 2015; Cádiz et al. 2020; Yu et al., 2022; Rainbow trout: Cádiz et al., 2021; Coho salmon : López et al., 2021; Australasian snapper: Baesjou & Wellenreuther, 2021; Brown trout: Magris et al., 2022) in order to identify genomic regions involved in recent adaptation or domestication processes (Smith & Haigh, 1974; Pennings & Hermisson, 2006). In this study, we were interested in farmed rainbow trout populations as it is one of the oldest farmed fish and the analysis of genes under either positive or balancing subsequent selection in. Indeed, only a few studies on selection signatures were performed in rainbow trout. Three of them only focused on wild populations and showed signatures of selection linked to life-history variation, egg development, spawning time (Martínez et al., 2011), immune response (Limborg et al., 2012), and smoltification (Weinstein et al., 2019). The first study in domesticated rainbow trout was performed on a single Chilean population (Cádiz et al., 2021) genotyped with a 57K SNP array; identified signatures of selection were associated with early development, growth, reproduction and immune system. Recently, a high-density array (665K SNPs) was developed for rainbow trout (Bernard et al., 2022), allowing us to potentially more accurately detect signatures of selection and to compare

them across various domesticated rainbow trout populations. The existence of signatures of selection shared by farmed populations from different geographical areas is essential to understand the importance of genetic diversity in several genomic regions in rainbow trout and then to identify genes having key roles in either the domestication process or fitness because conserved by all populations (Bruford et al., 2003; Yáñez et al., 2022).

Various approaches have been developed to reveal selection signatures within population based on site frequency spectrum, linkage disequilibrium (LD), or reduction of local variability (Vitti et al., 2013; Saravanan et al., 2020). Among these approaches, we will use two strategies, one based on the reduction of local variability using Run of Homozygosity (ROH) metrics and the second one relying on allele frequencies and the extent of LD based on Extended Haplotype Homozygosity (EHH). ROH is a large homozygous stretch in the genome of an individual inherited from a common ancestor to his parents (McQuillan et al., 2008; Purfield et al., 2012), while EHH measures the extent of shared haplotypes through the association between a single core haplotype and multiple loci at various distances from the core region (Sabeti et al., 2002). In our study, we considered four populations: one INRAE experimental line (with no intentional selection), two French selected lines from two different breeding companies, and a pooled American population gathering samples from one wild river and four hatchery populations, all coming from the North-West of the USA and closely genetically linked (Gao et al., 2018). Our work aimed to discover the main genomic regions sharing strong homozygosity (positive selection) or heterozygosity (balancing selection) across the four rainbow trout populations and to get further insights into the nature of genes spanning these regions.

2 | Material and methods

2.1 | Populations

Three French populations were considered: 14 breeding females from the INRAE synthetic line SY and, 90 and 72 females from two selected lines LB and LC from the breeding companies “Bretagne Truite” (Plouigneau, France) and “Viviers de Sarrance” (Sarrance, France) respectively. The SY was developed by intercrossing several domesticated lines of rainbow trout, in order to create a population with a large diversity (D’Ambrosio et al., 2019).

In addition, we considered an American pooled population, hereafter named HA, using the

whole genome sequence data of 20 fishes obtained by Gao et al. (2018). The sampling strategy consisted in collecting DNA from 4 individuals in each of five locations from the North-West of the USA: wild fish from Elwha River, and farmed fish from Dworshak, L. Quinault, Quinault, and Shamina hatcheries. We pooled together the 20 individuals, as these five populations were genetically close to each other (Supplementary Figure 1; Gao et al., 2018) and greatly distant from the three French populations (Figure 1).

2.2 / Genotyping and quality control

High-density genotypes were obtained at the INRAE genotyping Platform Gentyane (Clermont-Ferrand, France) for all the 176 French samples using the Affymetrix 665K SNP array recently developed for rainbow trout (Bernard et al., 2022). We only considered the genotypes for the 576,118 SNPs of the Rainbow Trout Axiom® 665K SNP array that were positioned on the Arlee genome (GCA_013265735.3, Gao et al., 2021; Bernard et al., 2022). From the whole-genome sequence information of the 20 American samples (Gao et al., 2018), we extracted the genotypes for the same 576,118 SNPs of the HD chip.

Among the 177 French genotyped fish, 19 individuals with more than 30% identity-by-state (IBS) with other individuals were removed from the dataset. We thus kept for the analysis 76, 67, 20, and 14 fish sampled from LB, LC, HA, and SY populations, respectively.

Then, SNP quality control was performed using PLINK v1.9 software (Chang et al., 2015). Note that, to avoid limitations due to the low number of individuals in SY, quality filters were made considering LC and SY together, as both populations were genotyped on the same SNP plate and are close genetically (D'Ambrosio et al., 2019). About 4,000 SNPs randomly distributed over all the genome were removed for all populations due to extreme deviation from Hardy-Weinberg equilibrium (p -value $< 10^{-7}$). It allowed us to discard SNPs with high risk of wrong genotyping, in addition to the edit for SNP call rate lower than 97%. We retained 571,319 SNPs, 569,030 SNPs, and 573,793 SNPs on LB, 'LC- SY', and HA populations, respectively. Finally, crossing the three SNP lists, we kept the 546,903 common SNPs for the analysis.

2.3 / Genetic structure of the populations

Genetic differentiation between populations was measured with a pairwise F_{st} estimate using the VCFtools v0.1.13 software (Danecek et al., 2011). In addition, a principal component analysis (PCA) was performed with the R package *Adegenet* (function *glPca*) (Jombart & Ahmed, 2011) to visualize the genetic structure of the populations.

2.4 | Runs of homozygosity

Runs of homozygosity (ROH) were identified for each fish using the PLINK v1.9 *homozyg* function (Chang et al., 2015) with the following options ‘*--homozyg-kb 500 --homozyg-window-snp 40 --homozyg-snp 40 --homozyg-gap 500 --homozyg-density 40 --homozyg-het 1*’. ROH was defined by a sliding window with a minimum length of 500 kb containing at least 40 homozygous SNPs. This minimum number of homozygous SNPs was chosen using the formula described by Purfield et al. (2012) in order to limit the number of ROHs that might only occur by chance.

2.4.1 | Estimation of inbreeding coefficients

The individual inbreeding coefficients (F_{ROH}) were calculated according McQuillan et al (2008) as $F_{i,ROH} = \frac{\sum length(ROH_i)}{L_{Genome}}$

With $\sum length(ROH_i)$ the sum of ROH length in an individual i and L_{Genome} the total length of the autosomal genome covered by SNPs.

2.4.2 | Identification of ROH islands

For each SNP, the number of individuals with this SNP included in a ROH was calculated in order to identify the regions of the genome that were frequently homozygous in each population, i.e. constituting ROH islands (Nothnagel et al., 2010). These ROH hotspots may then be considered as signatures of positive selection (Saravanan et al., 2021).

To allow the comparison of ROH islands across populations, we implemented population-specific thresholds based on the ROH occurrence to define ROH islands, as proposed in many studies (Purfield et al., 2017; Mastrangelo et al., 2017; Zhang et al., 2018; Peripolli et al., 2018; Grilz-Seger et al., 2018; Gorssen et al., 2021; Illa et al., 2022). The number of individuals corresponding to the top 5% of SNPs most often found in a ROH within each population was chosen as a threshold to define a ROH island.

These top 5% values were equivalent to 35, 27, 5, and 10 individuals for LB, LC, SY, and HA, respectively. Values chosen within each population corresponded to 48.6%, 40.3%, 35.7%, and 50% of individuals with a ROH in LB, LC, SY, and HA, respectively. In addition, two close SNPs in the top 5% were considered in the same ROH island if there were separated by a

distance lower than 500 kb with less than 40 SNPs in the gap stretch. The ROH island was delimited by a number of individuals, with ROH falling below the top 10% of the SNPs, which correspond to 30, 22, 3, and 7 individuals for LB, LC, SY, and HA populations, respectively.

2.4.3 / Detection of balancing selection signals based on regions without ROH

We used the ROH occurrence information per SNP to detect extreme heterozygous regions, i.e. without ROH. In these regions, we have an enrichment of heterozygous SNP relative to the genome-wide prevalence that may be due to balancing selection phenomena (Szpiech et al., 2013).

Applying the same criteria as for defining ROH, the minimal size and number of SNPs to define a heterozygous region were fixed to 500 kb and 40 SNPs, respectively. Moreover, two successive SNPs were considered in the same region if they were separated by a distance lower than 50 kb. A region was detected in extreme heterozygosity if less than 5% of individuals (per population) have SNPs in ROH in the region, corresponding to a maximum of respectively 4 and 3 individuals with a ROH in LB and LC populations, and to 0 individual with a ROH in SY and HA.

2.5 / Detection of signatures of selection based on Extended Haplotype Homozygosity (EHH)

For a given core allele, the EHH is defined as ‘the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for the entire interval from the core region to the point x’ (Sabeti et al., 2002). EHH measures the association between a single allele from the study locus (the core region) with multiple loci at various distances x (Sabeti et al., 2002). The iHS (Integrated Haplotype Homozygosity Score) proposed by Voight et al. (2006) aims to compare the integrated EHH profiles obtained for a SNP in the ancestral versus derived states. An extreme value of iHS corresponds to a positive selection because a core haplotype with unusually high EHH and high frequency in the population indicates the presence of a mutation that has spread through the population faster than the haplotype broke down.

EHH methodology requires haplotype information. Thus, genotype data must be phased before their calculation. We used FImpute3 (Sargolzaei et al., 2014) to phase the genotypes of the study females, considering all parents (including our study females) and offspring genotyped

in LB, LC, and SY populations for different purposes (see respectively Prchal et al., 2022, Lagarde et al., 2022 and Paul et al., 2022). All parents (except 8 SY sires) were genotyped with the HD chip (Bernard et al., 2022), while offsprings (and 8 SY sires) were genotyped with a 57K chip only (Palti et al., 2015). Information used for phasing is given in Table 1. Due to the lack of genotyped offsprings, only the HD genotypes information was used to phase the genotypes of the HA population.

TABLE 1. Data information used to phase the HD genotypes of the study females that belong to the parental cohorts. Number of individuals and SNPs available after quality control

Line	Status of individuals	Number of individuals	Number of SNP used
LB	parents	288	571,319
	offsprings	1,297	29,091
LC	parents	173	569,03
	offsprings	1,350	30,379
SY	parents (dams + 1 sire)	16	569,03
	offsprings (+ 8 sires)	866	32,725

Once phasing was performed, the *rehh* R package (Gautier & Vitalis, 2012; Gautier et al., 2017) was used to conduct EHH-based analyses. EHH detection was stopped when the EHH value declined under 0.1 or when the gap between two consecutive SNPs was higher than 20 kb (*scan_hh* function with the following options: *limehh* = 0.1; *maxgap*=20 kb).

2.5.1 | Cross population Extended Haplotype Homozygosity (XP-EHH)

From EHH information, we used the XP-EHH statistics (*ies2xpehh* function) to compare the integrated EHH profiles (iES), two by two, between a French (popA) and the HA (popB) populations at the same focal SNP (Sabeti et al., 2007) as:

$$XP_EHH = \frac{\ln\left(\frac{iES_{popA}}{iES_{popB}}\right) - Med\left[\ln\left(\frac{iES_{popA}}{iES_{popB}}\right)\right]}{SD\left[\ln\left(\frac{iES_{popA}}{iES_{popB}}\right)\right]}$$

The median (*Med*) and standard deviation (*SD*) of $\ln(iES_A/iES_B)$ were computed over all the analysed SNPs.

2.5.2 / Integrated Haplotype Homozygosity Score (iHS)

In the same way, we used the iHS test (Voight et al., 2006) to evaluate the evidence of positive selection based on haplotype frequencies in a single population, using the *ihh2ihs* function of the package *rehh*. This statistic was based on the log-ratio of the integrated EHH (iHH) for haplotypes with the ancestral (A) versus the derived (D) alleles and was computed for each

$$\text{autosomal SNP as } iHS = \frac{\ln\left(\frac{iHH_A}{iHH_D}\right) - Mean_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD_p\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}$$

The average ($Mean_p$) and standard deviation (SD_p) of $\ln(iHH_A/iHH_D)$ were computed over all the SNPs with a derived allele frequency p similar to that of the core SNP. In our study, the ancestral allele state is unknown. Therefore, we assumed that the most frequent allele represents the ancestral state as proposed by Bahbahani et al. (2015).

2.5.3 / Detection of candidate regions

To detect candidate regions for signatures of selection based on the iHS test, we used the *calc_candidate_region* function of the R package *rehh* (Gautier & Vitalis, 2012). We considered windows of 500 kb across the genome containing at least 30 SNPs, with 10 kb of overlapping. A region was considered as under positive selection if at least one SNP had a $-\log(p\text{-value}) > 4$ and extreme iHS value *i.e.* $|iHS| \geq 2.5$.

2.6 / Identification of common regions under positive selection

ROH islands and regions identified by iHS were pooled within each population. Then, the intersection set of the regions identified by one or another method across the four studied populations was established. We eliminated an intersection from the study if one population does not have at least one SNP with an $|iHS| \geq 2.5$ or enough individuals with an ROH in the intersected region. So, only intersections containing either ROH island or extreme iHS ($|iHS| \geq 2.5$) for the four populations were thus further analyzed.

2.7 | Gene analysis

The genes annotated in the regions under positive or balancing selection were identified from the NCBI *Oncorhynchus mykiss* genome assembly (GCA_013265735.3). Gene symbols were checked, and, if necessary, familiar names were added using the information available from GeneCards (<https://www.genecards.org/>).

Gene ontology (GO) terms study was performed with 'g:profiler' (Raudvere et al., 2019; <https://biit.cs.ut.ee/gprofiler/gost>) for the list of genes identified in the regions of interest.

Percent identity of rainbow trout proteins with nine other vertebrate species (human, mouse, cow, goat, pig, chicken, zebrafish, medaka, and Atlantic salmon) was established using the blastp tool (Basic Local Alignment Search Tool on proteins).

3| Results

3.1 | Genetic diversity within and across populations

The ROH statistics and inbreeding coefficients are presented in **Table 2** for all the populations. The average number of ROH per individual varied between 141 (SY) and 168 (LB). French selected lines had larger average sizes of ROH than populations SY and HA. The average inbreeding coefficients of HA individuals were between three (compared to SY) and five (compared to LB) times lower than those of the French lines.

TABLE 2. ROH statistics and inbreeding coefficients of the four studied populations (Standard deviations are indicated in brackets).

Population	Average number of ROH	Average size of ROH (in kb)	Average F_{ROH}
LB	168 (14.6)	2,770 (270.8)	0.20 (0.02)
LC	157 (15.9)	2,485 (326.8)	0.17 (0.03)
SY	141 (33.5)	1,860 (291.2)	0.12 (0.05)
HA	167 (65.6)	1,433 (145.6)	0.04 (0.03)

Based on genome-wide F_{st} values, large differentiation of around 0.28 was observed between HA and any of the French populations (**Table 3**). In the PCA figure (**Figure 1**), the three French lines differed strongly from the American pooled populations, and the first two PCA axes explained 29% of the total genetic variation. In addition, F_{st} values indicated that all the French lines were moderately differentiated (0.104 – 0.122).

TABLE 3. Genome-wide F_{st} statistics derived two-by-two between the four populations.

	LC	LB	HA
SY	0.104	0.122	0.275
LC		0.122	0.275
LB			0.289

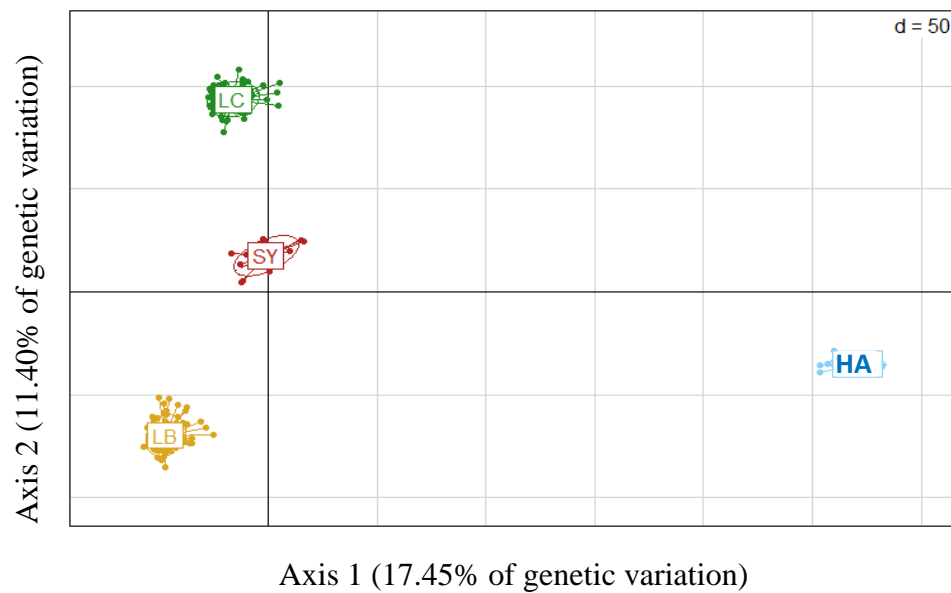


FIGURE 1. Principal component analysis (PCA) of the genetic diversity of the four rainbow trout populations (LB, LC, SY, and HA) based on 546,903 SNPs.

Using the XP-EHH statistics, we identified 93, 105 and, 135 regions that strongly discriminated HA from LB, LC and SY, respectively. Among these regions, 34 regions were shared, spanning about 32 Mb in total over 21 chromosomes, and differentiated any of the French lines from the American HA pooled population (Supplementary information S1).

The distribution of the proportion of individuals having a ROH at each SNP position is presented in Figure 2. In average, ROH were more shared between individuals for selected lines (LB and LC, on average, 23.39% and 19.82% of individuals respectively) than for other populations (SY and HA, on average, 13.67% and 8.91% of individuals respectively). Probably linked to the composite nature of the HA population (5 sub-groups of 4 individuals), HA contained the lowest number of shared ROH among the individuals but also showed the most shared ROH among individuals.

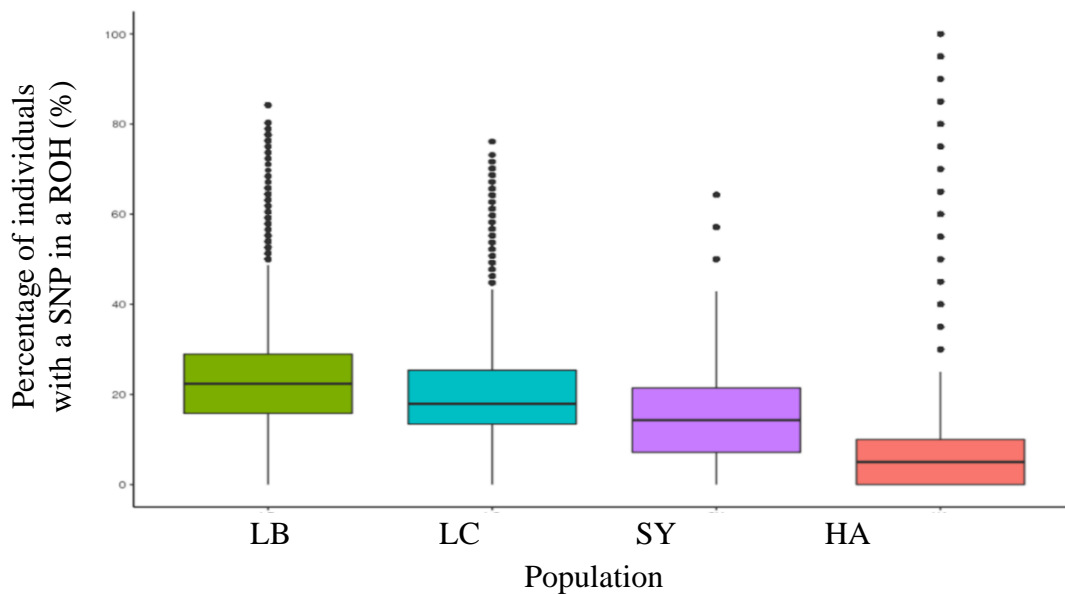


FIGURE 2. Box-plots of the occurrence of ROH (number of individuals having this ROH) per SNP for each rainbow trout population LB, LC, SY, and HA.

3.2 | Signatures of positive selection

3.2.1 | ROH islands

The sharing of ROH among individuals, regardless of the population considered, was presented Figure 3. Eight ROH islands were shared by at least 2 populations, and a minimum of 50% of individuals concerned. However, only three of these regions were defined as ROH island in each of the 4 populations.

We listed all ROH islands within each population which resulted in the identification of 270 ROH islands distributed among the four populations (Supplementary Informations S2 to S5,

for LB, LC, SY, and HA, respectively). The ROH islands were not evenly distributed across populations and chromosomes. The average ROH island size was 2,737 kb, varying from 1,593 kb to 4,465 kb, depending on the population. The longest ROH island was observed in SY (21.4 Mb), while the shortest one was observed in LC (16.1 kb).

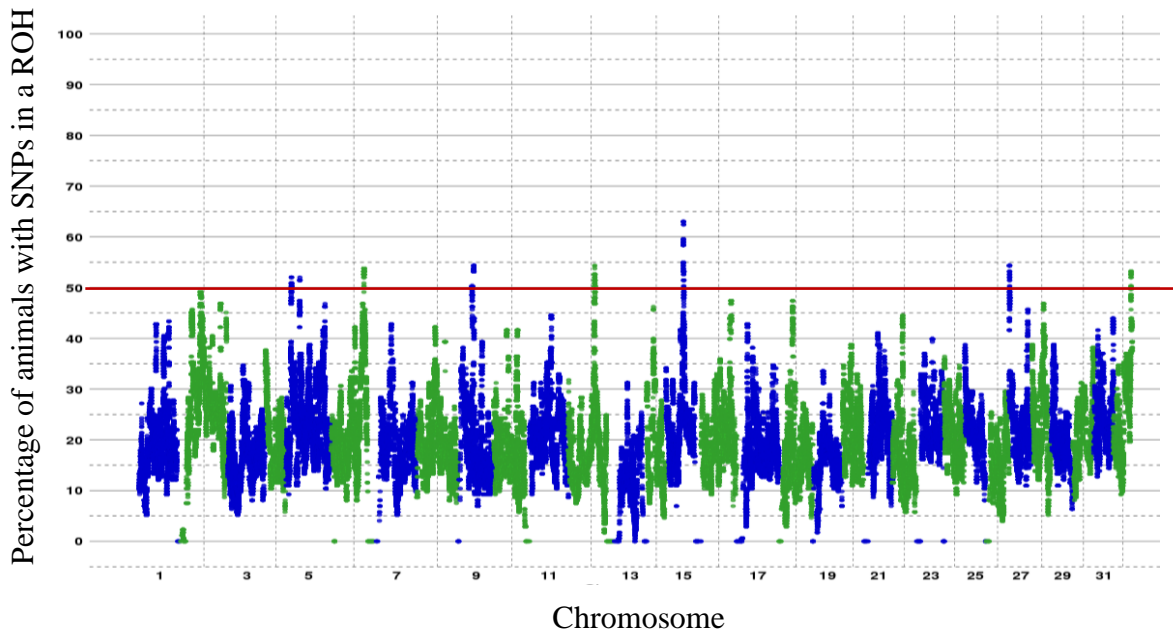


FIGURE 3. Manhattan plot of the occurrence of ROH per SNP across chromosomes (gathering all rainbow trout populations). The red line highlights the ROH islands.

3.2.2 | Identifying selection signatures using *iHS*

The $-\log(p\text{-values})$ of *iHS* calculated along the genome are presented in **Figure 4** for each population (all regions identified with *calc-candidate_region* are described in **Supplementary Informations S6-S9**). While numerous regions have been detected as under positive selection overall, fewer candidate regions were detected for the French lines (LB, LC and, SY) than for the American pooled population (HA). The genome-wide highest estimated values of $|iHS|$ were 8.97, 7.24, 5.67, and 9.09 for LB, LC, SY, and HA, respectively (with $-\log(p\text{-values}) > 7.8$).

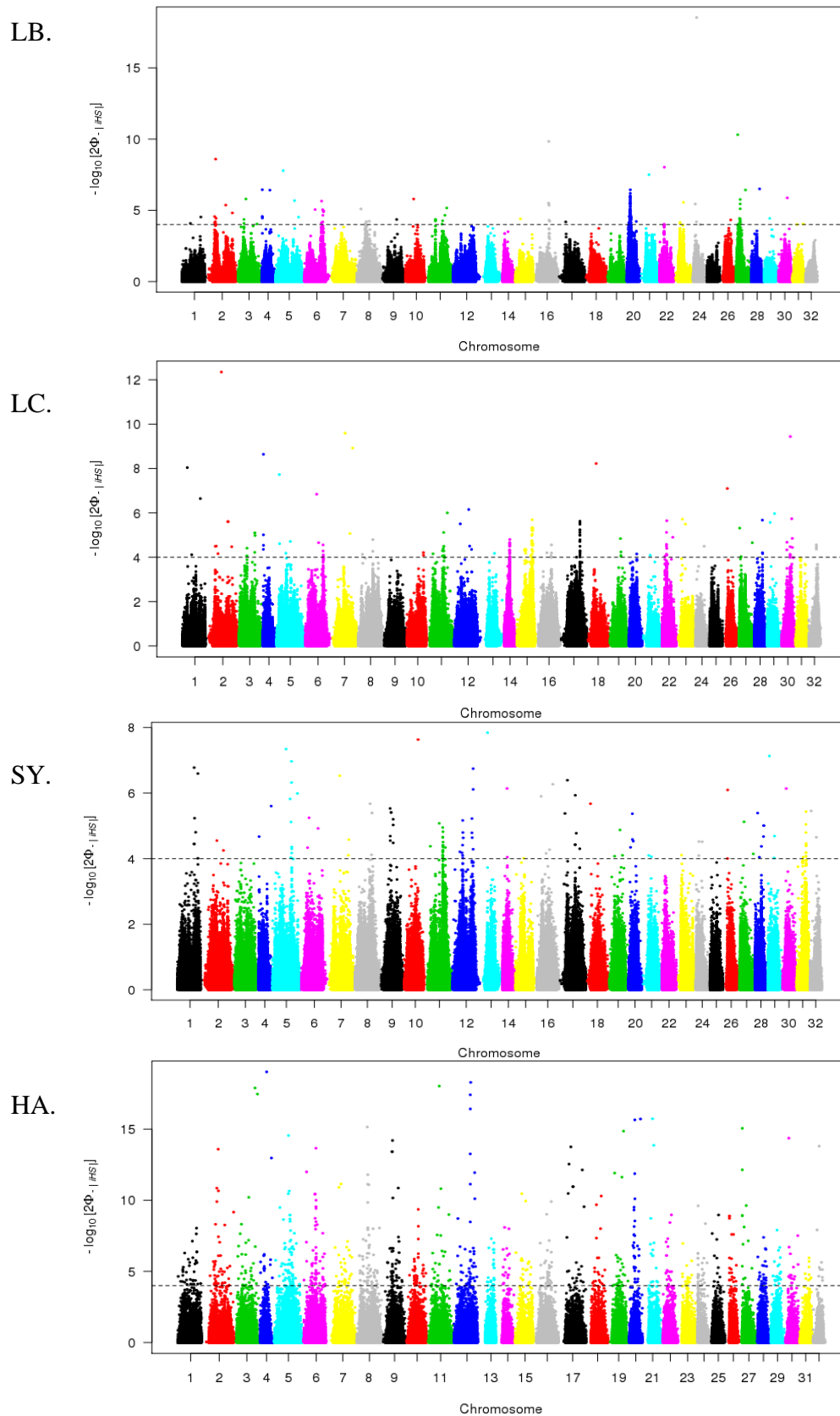


FIGURE 4. Genome-wide distribution of $-\log(\text{p-value})$ for standardized $i\text{HS}$ for each population (LB, LC, SY, HA). The dashed line indicates the $-\log(\text{p-value})$ significance threshold set to 4 to identify regions under positive selection

In total, 72, 68, 76, and 54 ROH islands were identified in LB, LC, SY, and HA populations respectively (**Figure 5**). Using iHS statistics, 55, 69, 73, and 362 signatures of selection were detected for LB, LC, SY, and HA populations, respectively. Only 10.4%, 8.7%, 8.0%, and 5.6% of the regions were detected by both methods (ROH + iHS) for LB, LC, SY, and HA populations, respectively.

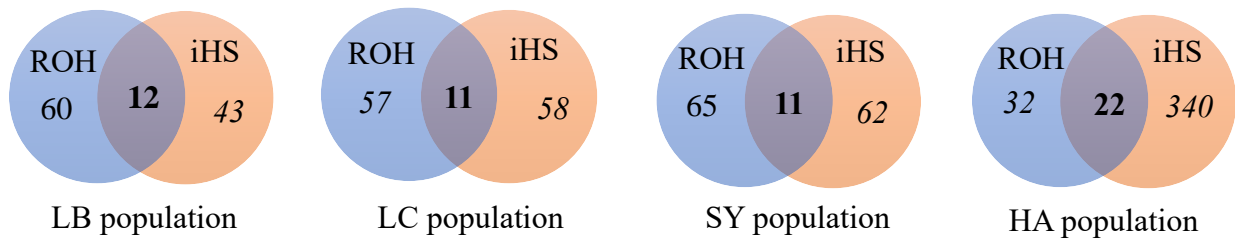


FIGURE 5. Venn Diagram of the number of regions identified as ROH island or iHS signature of selection for each rainbow trout population.

3.2.3 / Regions under positive selection shared by the four studied populations

Among the numerous regions identified for each population by either ROH or iHS methods, only nine regions were shared by the four studied populations (Table 4). The average size of these shared regions was 1,135 kb. Five of them were located on chromosome 2, and the four other regions were on chromosomes 12, 15, 16, and 20, respectively.

Depending on the population, six regions were identified by either ROH or iHS metrics. Two regions, chr2_c and chr15_a, were only detected by ROH in all four populations, while a single region, chr16_a, was only identified through significant iHS statistics in all the four populations (Supplementary Information S10). The list of genes annotated in the nine shared genomic regions is given in Supplementary Informations S11.

TABLE 4. Homozygous regions under positive selection in the four studied populations.

Region	CHR	Start (Mb)	End (Mb)	Size (kb)
chr2_a	2	25.40	26.30	900
chr2_b	2	31.60	34.20	2600
chr2_c	2	46.00	46.66	664
chr2_d	2	69.70	71.20	1500
chr2_e	2	88.46	89.34	878
chr12_a	12	57.97	59.10	1138
chr15_a	15	38.96	39.57	610
chr16_a	16	45.80	47.00	1200
chr20_a	20	19.10	19.83	726

3.3 / Signatures of balancing selection

3.3.1. Regions under balancing selection detected within population

In total, 14, 24, 158, and 265 hot spots of polymorphism (i.e. without ROH) were identified in LB, LC, SY, and HA populations, respectively. The numbers of heterozygous regions detected for SY and HA populations were drastically larger than those observed for the LB and LC selected lines. The average size of the detected heterozygous regions was 1,400 kb, varying from 1,086 kb to 1,828 kb, depending on the population.

Tables listing all heterozygous regions within each population are presented in Supplementary Informations S12 to S15, for LB, LC, SY, and HA, respectively.

3.3.2. Regions under balancing selection shared by the four studied populations

A substantial lack of ROH was observed in four regions of all studied rainbow trout populations (Table 5). Two of them, chr10_a and chr19_a, were particularly small (53 kb and 70 kb, respectively), but still contained at least 20 SNPs. The region chr10_a only encoded one of the introns of the *ctnna2* (= catenin alpha 2) gene while chr19_a was composed of two genes, *smarca5* (=SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5) and *frem2* (= FRAS1-related extracellular matrix protein 2). A second heterozygous region on chromosome 19 was larger (163 kb) but contained a single annotated gene, *pou4f2* (= POU domain, class 4, transcription factor 2-like). A last region chr13_a spanned over 1,100 kb) on chromosome 13 and was composed of 25 genes. The list of genes annotated in the four shared genomic regions is given in Supplementary Informations S16.

TABLE 5. Highly heterozygous regions shared by the four studied populations.

Region	CHR	Start (Mb)	End (Mb)	Size (kb)	SNP number	SNP density per Mb
chr10_a	10	56.314	56.366	53	20	379
chr13_a	13	46.959	48.071	1,112	446	401
chr19_a	19	10.753	10.823	70	24	342
chr19_b	19	11.354	11.517	163	52	319

3.4 | Identification and roles of genes underlying the regions under selection across all populations

3.4.1. Common homozygous regions under positive selection

The nine common homozygous regions contained a total of 253 genes (listed in Supplementary Information S11). A gene ontology (GO) study was performed and showed a significant over-representation (p -value < 0.01) among the 253 genes of functions related to the following GO terms: membrane (GO:0016020, CC: Cellular Component, p -value = $1.3e10^{-5}$), intrinsic and integral component of membrane (GO:0031224; GO:0016021, CC, p -value = $0.001/0.005$), ion binding (GO:0043167, MF: Molecular Function, p -value = 0.002), and nuclear speck (GO:0016607, CC, p -value = 0.008).

Among the nine studied regions, the three regions chr2_a, chr2_c, and chr15_a, that contain less than ten genes annotated in each, were analyzed in further detail to accurately define the roles of underlying genes. The 17 genes located in these three regions are listed in **Table 6** with their associated biological functions. These genes play key roles in protein transduction/maturation, genome stability, embryonic development, growth, energetic function, reproduction, or immune function. In addition to this list of genes, a subset of 15 genes in the six other homozygous regions already identified as signatures of selection in the literature were further studied in terms of their biological functions. Detailed information for these genes is also given in Table 6.

TABLE 6. List and functions of the 17 genes annotated in three homozygous regions (chr2_a, chr2_c and chr15_a) shared by the four rainbow trout populations, and the 15 genes in the six other regions already identified as signatures of selection in the literature. *SS : *Identify by signature of selection in that study*

Region	Gene name	Protein name		General functions	References	
chr2_a	<i>mrap2</i>	melanocortin-2 accessory protein 2A	receptor	Cellular organization and growth	May regulate both receptor trafficking and activation in response to ligands. Link to energy homeostasis control and body weight regulation. Linked to severe obesity in many species	Liu et al., 2013 (human, zebrafish, rodent); Zhu et al., 2019 (sea lamprey); Wen et al., 2021 (snakehead); SS : Cadiz et al., 2021 (rainbow trout), Cumer et al., 2021 (goat)
	<i>cep162</i>	centrosomal protein of 162 kDa		Cellular and nuclear organization	Involved in cilium assembly (promote transition at the cilia base). Acts by specifically recognizing and binding the axonemal microtubule.	Wang et al., 2013
	<i>uncharacterized LOC110539089</i>					
	<i>adgrb1</i>	adhesion G protein-coupled receptor B1		Neuronal and embryonic development	Essential for growth and metastasis of solid tumors (zebrafish). Plays a role during brain/neuron development, associated with autism in mice and human (BA11 synonymous of <i>adgrb1</i>).	Purcell, 2017 (human); Cazorla-Vázquez & Engel, 2018 (from zebrafish to human); Shiu et al., 2022 (mice)
	<i>tsnare1</i>	t-SNARE protein 1	domain-containing	Cellular organization and neuronal development	Predicted to be involved in intracellular protein transport; vesicle docking; vesicle fusion; and integral component of membrane. Neurodevelopment function.	Fromer et al., 2016 (zebrafish and human)
	<i>pttg1ip</i>	pituitary tumor-transforming gene 1 protein-interacting protein		Cellular organization and growth	Participates in metaphase-anaphase transition of the cell cycle and facilitates translocation of <i>ptg1</i> into the nucleus + allow to predict breast cancer survival + induced transcriptional activation of transcriptional basic fibroblast growth factor (when coexpressing with <i>ptg1</i>).	Repo et al., 2017 (human)
	<i>Cdk14</i>	cyclin-dependent kinase 14		Neuronal and embryonic development	Regulator of cell cycle progression and proliferation + role in meiosis, neuron differentiation/craniofacial development (Wnt signaling pathway)	Margarit et al., 2014 (zebrafish); Yin et al., 2021 (human)
chr2_b	<i>pbx1</i>	pre-B-cell leukemia transcription factor 1		Neuronal and embryonic development	Related to early development in zebrafish. Mutations in this gene generally cause major malformations, which seem to play an essential role in survival in various species.	Teoh et al., 2010 (zebrafish); Selleri et al., 2001 (mouse); Le Tanno et al., 2017 (human); SS : Cadiz et al., 2021 (rainbow trout)
	<i>col9a2</i>	collagen alpha-2(IX) chain		Neuronal and embryonic development	Component of cartilage, implicated in human intervertebral disc degeneration (IVDD) and seems also related to growth. Mutations in this gene may cause diverse syndromes, such as multiple epiphyseal dysplasias and ocular, skeletal, orofacial, and auditory abnormalities in humans.	Muragaki et al., 1996; Baker et al., 2011; Xu et al., 2022 (human); SS : Lopez et al., 2018 (atlantic salmon)
	<i>brd2</i>	bromodomain containing 2		Nuclear and cellular organization, neuronal and embryonic development	Associated with transcription complexes and acetylated chromatin during mitosis. Potential role in oogenesis, egg-to-embryo transition, and proper development of the digestive and central nervous systems (zebrafish). And involved in spermatogenesis or folliculogenesis, as demonstrated in situ on mice cells.	DiBenedetto et al., 2008 (zebrafish); Rhee et al., 1998 (mouse); SS : Lopez et al., 2018 (atlantic salmon)
	<i>scap</i>	sterol regulatory element-binding protein cleavage-activating protein		Cellular organization	Binds to sterol regulatory element binding proteins (SREBPs) and transports them from the ER to the Golgi.	Howarth et al., 2013 (zebrafish); SS : Wang et al., 2022 (Pacific White Shrimp)

Table 6 (continued)

Region	Gene name	Protein name		General functions	References
	<i>tnf-α-ip8l2b</i>	tumor necrosis factor, alpha-induced protein 8-like protein 2 B	Immunity	Predicted to be involved in the negative regulation of T-cell activation, inflammatory response, innate and adaptive immunity by maintaining immune homeostasis.	Umasuthan et al., 2014 (Oplegnathus fasciatus); Sullivan et al., 2017 (vertebrates)
	<i>atg5</i>	autophagy protein 5	Immunity	Involved in several cellular processes linked to the immune response, such as autophagic vesicle formation, innate antiviral immune response, lymphocyte development and proliferation in mice.	Miller et al., 2008; Ye et al., 2018 (mouse)
chr2_c	<i>brsk2a</i>	serine/threonine-protein kinase BRSK2	cellular organization, neuronal and embryonic development	Enable in several functions: ATP/ATPase binding activity, protein kinase activity. Key role in polarization of neurons and axonogenesis, cell cycle progress (apoptotic signaling pathway) and insulin secretion (metabolic process). This gene is related to autism spectrum disorder (social deficit) and locomotor defects (larval phase and adulthood) in zebrafish	Hiatt et al., 2019 (human); Deng et al., 2022 (human); Liu et al., 2022 (zebrafish)
	<i>abtb2b</i>	Ankyrin Repeat And BTB Domain Containing 2b	Cellular organization	Predicted to be involved in SMAD protein signal transduction, heterodimerization activity. Act upstream of or within cellular response to toxic substance.	
	<i>b4galnt4a</i>	N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-acetylgalactosaminyltransferase 1	Cellular organization	Enables acetylgalactosaminyltransferase activity. Predicted to be located in Golgi cisterna membrane. Predicted to be integral component of membrane.	
chr2_d	<i>igf-1a</i>	insulin-like growth factor 1a receptor	Growth	Plays a critical role in transformation events. Cleavage of the precursor generates alpha and beta subunits. It is highly overexpressed in most malignant tissues where it functions as an anti-apoptotic agent by enhancing cell survival.	SS : Wayne & vonHoldt, 2012 (dog); Lopez et al., 2019 (atlantic salmon)
chr2_e	<i>znf135</i>	gastrula zinc finger protein XICGF26.1	Neuronal development and cellular organization	Involved in cytoskeleton organization, regulation of cell morphogenesis, and RNA-binding. A mutation of znf135 is related to neurological symptoms in humans.	Raghuram et al., 2017 (human); SS : Gutierrez et al., 2016 (atlantic salmon)
chr12_a	<i>grxcr1</i>	glutaredoxin domain-containing cysteine-rich protein 1-like	Neuronal development and cellular organization	Involved in actin organization in hair cells and is associated with a non-syndromic hearing impairment and the regulation of hair bundle morphogenesis in mouse. A mutant for this gene was identified in mice and linked to hyperactivity (modifies behaviour).	Liu et al., 2021, Lorente-Cánovas et al., 2022 (mouse); SS: Saravanan et al., 2021 (cattle)
chr15_a	<i>chn1</i>	N-chimaerin	Embryonic development	Encodes GTPase-activating protein. Plays an important role in neuronal signal-transduction mechanisms. Implication during embryonic development: cell polarity and lack of yolk extension. In zebrafish, a morpholino knockdown of chn1 reveals its crucial role in early development, revealing severe abnormalities (development of round somites, lack of yolk extension, and kinked posterior notochord).	Leskow et al., 2006 (zebrafish); Miyake et al., 2008; Ip et al., 2012 (human)
	<i>atp5mc1</i>	ATP synthase lipid-binding protein, mitochondrial	Energetic function	Loss of ATP synthase -> aberrant mitochondria cristae morphology + energy metabolism correlated to immune system	Palmer et al., 2011; Miller et al., 2019 (human); Wang et al., 2020 (chicken)
	<i>zc3h15</i>	zinc finger CCCH domain-containing protein 15	Embryonic development and cellular organization	Embryonic development (positive regulation of GTPase activity) / Elongation processivity (high tumor progression in melanoma). In addition, in vitro (mice cells) that zc3h15 knockdown had an inhibitory effect on HIV-1 replication and then on HIV infection.	Capalbo et al., 2010 (mice); Li et al 2021 (human)

Table 6 (continued)

Region	Gene name	Protein name	General functions	References	
	<i>zp4</i>	zona pellucida sperm-binding protein 4-like	Reproduction	Extracellular matrix that surrounds the oocytes and early embryo. Plays vital roles during oogenesis, gamete development, fertilization and preimplantation development. Mutation in this gene induces infertility in both males and females in mammals.	Wassarman & Litscher, 2018 (fish); Li et al., 2021 (zebrafish); SS : Lopez et al., 2019 (atlantic salmon)
	<i>uncharacterized protein LOC110490841</i>				
	<i>nid2</i>	nidogen-2-like	Cellular and nuclear organization	Cell-adhesion protein that binds collagens I and IV and laminin and may be involved in maintaining the structure of the basement membrane. Linked to ovarian cancer	Torky et al., 2018 (human); Zhang et al., 2022 (zebrafish, mouse)
	<i>brca2</i>	breast cancer type 2 susceptibility protein	Genome stability and cellular organization	Essential for efficient homology-directed ADN repair. Impaired homology-directed repair caused by <i>brca2</i> deficiency leads to chromosomal instability and tumorigenesis through lack of repair or misrepair of DNA damage. plays an essential role in ovarian development and tumorigenesis of reproductive tissues	Shive et al., 2010 (zebrafish); Rodriguez-Mari et al., 2011 (zebrafish); Moynahan et al., 2001; Chen et al., 2018 (human)
	<i>atp1b3</i>	sodium/potassium-transporting ATPase subunit beta-1-interacting protein 3	Cellular and nuclear organization	ATPase responsible for establishing and maintaining the electrochemical gradient of Na ⁺ and K ⁺ ions across the plasma membrane, essential for osmoregulation.	Zhang et al., 2019 (human); SS : Naval-Sanchez et al., 2020 (atlantic salmon)
	<i>Dnajc5</i>	dnaJ homolog subfamily C member 5-like	Cellular and nuclear organization	Regulated the ATPase activity of 70kDa heat shock proteins and plays a role in membrane trafficking and protein folding. This protein has been shown to have also anti-neurodegenerative properties in human with a gene expression study.	Nosková et al., 2011 (human); SS : Signer-Hasler et al., 2022 (goat)
	<i>samD10</i>	sterile alpha motif domain-containing protein 10-like	Cellular and nuclear organization	Linked to binding activity and transmembranaire pathway	SS : Signer-Hasler et al., 2022 (goat)
chr16_a	<i>nol4</i>	nucleolar protein 4-like	Cellular and nuclear organization	Predicted to enable RNA binding activity	SS : Signer-Hasler et al., 2022 (goat)
	<i>tpd54</i> (= <i>TPD52L2</i>)	tumor protein D54	Cellular and nuclear organization	Related to cellular organization, are characterized by an N-terminal coiled-coil motif that forms homo and heteromeric complexes and affects cell proliferation, adhesion, and invasion.	Mukudai et al., 2013; Zhuang et al., 2019 (human); SS : Signer-Hasler et al., 2022 (goat)
	<i>magi2</i>	membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2	Neuronal and embryonic development	Plays a role in regulating activin-mediated signaling in neuronal cells. In zebrafish, the protein of this gene plays a vital role in embryogenesis.	Borah et al., 2016 (zebrafish); SS : Cumer et al., 2021 (sheep); Hou et al., 2012 (cattle)
	<i>emilin3</i> (= <i>emilin-2</i>)	EMILIN-3	Embryonic development and growth	Played a role in extracellular matrix organization and elastic fiber formation. Its gene expression was related to embryonic development and involved in muscle fiber development in zebrafish.	Milanetto et al., 2008 (zebrafish); SS : Baesjou & Wellenreuther, 2021 (australasian snapper)
chr20_a	<i>auts2</i>	autism susceptibility gene 2 protein homolog	Neuronal development	Related to central nervous system development and is associated with autism in humans.	Oksenberg et al., 2013; Hori et al., 2021 (human); SS : Lopez et al., 2018 (atlantic salmon); Consortium, bovine hapmap, 2009 (cattle)

We studied the degree of protein identity among 10 vertebrate species for all the 17 genes of regions chr2_a, chr2_c, and chr15_a (Table 7), considering a protein as highly conserved if its identity between rainbow trout and other species was higher than 85%. Except for the proteins linked to *cep162* (centrosomal protein of 162 kDa) and *zp4* (zona pellucida sperm-binding

protein 4-like) genes, all other proteins were highly conserved at least between the two studied salmonids. In each of the three regions, one or two genes were highly conserved across the all ten study species: in chr2_a, rainbow trout *cdk14* (cyclin-dependent kinase 14) protein had a percent identity between 86 and 99.6% with the other species; in chr2_c, rainbow trout *brsk2a* (serine/threonine-protein kinase *brsk2*) protein had between 92 and 96.3 % of percent identity with the other species; in chr15_a, two genes, *chn1*(n-chimaerin) and *atp5mc1* (ATP synthase lipid-binding protein, mitochondrial), also had protein percent identity ranging from 85% to 98% depending of the species.

TABLE 7. Percentage of protein identity between rainbow trout and nine other vertebrate species for all genes annotated in homozygous regions chr2_a, chr2_c and chr15_a.

Region	gene_ID	Human	Mouse	Goat	Cattle	Pig	Chicken	Zebrafish	Medaka	Atlantic salmon
chr2_a	<i>mrp2a</i>	45.71	43.52	43.87	44.98	45.45	42.20	56.22	50.45	87.55
	<i>cep162</i>	36.61	37.50	40.95	40.95	50.00	53.73	38.61	62.50	80.66
	<i>adgrb1</i>	62.37	63.44	62.12	61.95	61.74	67.72	84.02	80.94	98.16
	<i>tsnare1</i>	54.78	31.50	56.99	55.79	56.02	60.74	85.62	78.55	98.29
	<i>pttg1IP</i>	60.00	57.89	57.04	57.04	58.82	59.74	70.92	66.03	93.89
	<i>cdk14</i>	87.05	87.05	86.44	86.02	85.99	87.24	88.96	91.08	99.58
chr2_c	<i>brsk2a</i>	92.12	92.50	92.66	92.19	92.66	93.82	96.14	92.05	96.26
	<i>abtb2b</i>	71.54	70.76	71.93	71.74	71.74	72.46	81.05	61.68	97.35
	<i>b4galnt4a</i>	63.35	65.38	57.14	64.63	64.95	66.37	66.06	79.96	96.55
chr15_a	<i>chn1</i>	88.80	86.59	88.04	87.32	88.04	88.10	85.29	85.01	98.04
	<i>atp5mc1</i>	97.37	87.10	91.76	94.44	93.33	86.17	91.30	97.87	94.12
	<i>zc3h15</i>	69.35	68.57	67.55	67.55	67.55	66.90	74.33	71.57	97.30
	<i>zp4</i>	29.67	29.61	31.87	37.47	30.21	31.46	45.60	49.74	74.74
	<i>nid2</i>	52.00	51.16	51.11	51.11	51.22	55.00	58.14	52.78	97.87
	<i>brca2</i>	46.24	43.72	37.12	32.42	45.61	45.85	38.68	54.08	92.05

Some other rainbow trout proteins (*tsnare1*, t-SNARE domain-containing protein 1; *pttg1IP*, pituitary tumor-transforming gene 1 protein-interacting protein) were conserved to a lesser extent (minimum 65% of percent identity) with the three other fish species, some being also conserved at least with chicken (*adgrb1*, adhesion G protein-coupled receptor B1; *b4galnt4a*, N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-acetylgalactosaminyltransferase 1) or even with all the nine study species (*zc3h15*, zinc finger CCCH domain-containing protein 15).

3.4.2. Common heterozygous regions under balancing selection

The four common heterozygous regions (Table 5) contained 29 genes (listed in Supplementary Information S16). A gene ontology (GO) terms study showed no significant over-representation of specific GO terms.

The degree of protein percent identity among 10 various vertebrate species for these 29 genes are presented in supplementary information S17.

Regions chr10_a, chr19_a, and chr19_b contained only a few genes and were then analyzed in further detail to accurately determine the role of underlying genes (Table 8). These genes play key roles in cellular and nuclear organisation and in embryonic development.

TABLE 8. List and functions of the 4 genes annotated in three heterozygous regions (chr10_a, chr19_a, and chr19_b) shared by the four rainbow trout populations.

Region	Gene name	Protein name	General function	References
chr10_a	<i>ctnna2</i>	catenin alpha 2	Enables actin filament binding activity, and involved in negative regulation of Arp2/3 complex-mediated actin nucleation. Regulation of neuron migration and of neuron projection development. Implicated in brain malformations. Seems implicated in vertebral development(/deformities) in Yunlong grouper.	Uvarov et al., 2014 (vertebrates); Li et al., 2022 (Yunlong grouper)
chr19_a	<i>smarca5</i>	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	The protein encoded by this gene is a member of the SWI/SNF family of proteins. Members of this family have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering the chromatin structure around those genes. The protein encoded by this gene is a component of the chromatin remodeling and spacing factor RSF, a facilitator of the transcription of class II genes by RNA polymerase II	Ding et al., 2021 (zebrafish); Limi et al., 2018 (mice); Armas et al., 2013 (zebrafish)
	<i>frem2</i>	FRAS1-related extracellular matrix protein 2	Plays a role in epidermal-dermal interactions -> important for the integrity of skin and renal epithelia.	Gautier et al., 2008 (zebrafish)
chr19_b	<i>pou4f2</i>	POU domain, class 4, transcription factor 2-like	May be involved in maintaining visual system neurons in the retina, and in the lateral line. The level of the encoded protein is also elevated in a majority of breast cancers, resulting in accelerated tumor growth. Seems link to cardiac development in zebrafish	DeCarvalho et al., 2004 (zebrafish); Maskell et al. 2017 (zebrafish)

4| Discussion

The objective of our study was to detect signatures of selection into domestic rainbow trout. To reach that goal we studied four genetically distinct populations coming from different locations either in France or in the North-West of the USA. We used two different approaches, ROH and EHH, to detect the genomic regions shared by all populations using a HD SNP. We were able to detect 9 very conserved regions and 4 hotspots of polymorphism, corresponding to 253 and 29 annotated genes, respectively.

4.1 | Genetic structure

First, we described the genetic structure of the populations under scrutiny. The three French lines were moderately differentiated with F_{st} ranging from 0.10 to 0.12. These estimations were congruent with those computed by D'Ambrosio et al.'s study (2019) with the same populations that ranged between 0.09 and 0.14, but were estimated using a 38K SNPs array. These moderate differences between the 3 French populations were consistent with the PCA we performed and the history of these populations with a partly common INRAE origin (D'Ambrosio et al., 2019). This trend is shared between European populations with for instance an average F_{st} of 0.13 between 12 European rainbow trout strains (Gross et al., 2007) or 0.12 among 9 Norwegian populations (Glover, 2009). Similarly, US farmed populations are also weakly to moderately differentiated with an average F_{st} of about 0.09 (Silverstein et al., 2004) or 0.13 (Liu et al., 2017b) and pairwise values ranging from 0.06 and 0.16. We observed a similar pattern in the present study with the HA population that consisted in samples from 5 locations, which all clustered together in the PCA. Reversely, we observed a large differentiation between our French and US populations revealed by large F_{st} values (0.27-0.29). This is likely the result of numerous factors, including selection, genetic drift and absence of gene flow between these very geographically distant populations. In addition, the European farmed populations originated from Californian domesticated strains, that have been shown to differ from strains of North-Western USA (Stanković et al., 2015). We found 34 haplotypes distributed over 21 chromosomes that differed between the American pooled population (HA) and all French populations (Supplementary information S1).

Due to the moderate to large differentiation between the 4 populations, the conserved regions across all populations are likely to be the result of ancient natural selection traces.

4.2 / Comparison of methods to detect common signatures of positive selection

We used a double check of positive selection traces in the genome by using both ROH and EHH approaches. However, for each population, only a few regions were identified by both methods. These regions detected by more than one method represent stronger evidence of selection signatures since outlier markers detected by various genome scan methods help to uncover true selection signatures by reducing the number of false positives.

Even if both methods evaluate the homozygous large stretches in the genome, iHS also considers information based on haplotypic version and linkage disequilibrium from a core SNP. ROH approach detects homozygous regions regardless of their haplotypic versions, contrary to iHS. Thus, it may detect a signature of positive selection even if various haplotypes were present at the homozygous state in the population. In addition, while the ROH approach only detects the homozygous large stretches (at least 500 kb in the present study), iHS can detect small regions under positive selection as the only limitation in EHH region size is based on a threshold value for a minimum LD (0.10). Consequently, the sizes of the detected homozygous region varied between 1,065 kb and 2,857 kb based on ROH metrics and between 1,000 kb and 1,600 kb with iHS statistics.

The high number of regions (55, 69, 73, and 362) detected by iHS in our study was consistent with numbers detected in either Atlantic salmon (López et al., 2019) or cattle (Saravanan et al., 2021). However, these two previous studies used a lower threshold than ours ($-\log(\text{p-value}) = 3$ and 2 , respectively vs 4 in the present study). Lower numbers of regions were previously detected by iHS in rainbow trout by Cádiz et al. (2021) and in Coho salmon by López et al. (2021). We speculate that these differences in the numbers of detected signals may be linked to the lower density of SNPs they used in both studies (57K or 200K chip versus 665K chip for our study) and the subsequent lower ability to detect LD and haplotypes at fine scale. Indeed, in the Chilean rainbow trout study (Cádiz et al., 2021), only one signal of positive selection was detected by iHS located at 6.398-14.936 Mb on chromosome 20 of the Swanson reference genome, which corresponds to the region 7.488-16.111 Mb on chromosome 20 of the Arlee reference genome. Nevertheless, we also detected by iHS signals of selection in each of our four studied populations, located at 10.5-16.5 Mb for LB, 11.2 – 13.3 Mb for LC, 13.0 – 14.2 Mb for SY and 12.3-13.2 Mb for HA (Supplementary Informations S6 to S9 for LB, LC, SY, and HA, respectively). Thus, all these signals were consistent with the larger region identified by Cádiz et al. (2021).

A common putative selection signature located at 13.0-13.2Mb could also be shared by all studied populations. In this 200kb-region, we observed at least one *iHS* value over $|2.5|$ for LB, LC and SY lines, but not for HA population. In this region of 200 kb on chromosome 20, six genes were identified (*lgi1*, *noc3l*, *plce1*, *slc35g1*, *fra10ac1*, *tbc1d12*). Among these genes, Cádiz et al. (2021) identified two candidate genes associated with domestication, *noc3l* (nucleolar complex protein 3 homolog) and *plce1* (1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase epsilon-1). Both are related to early development traits in zebrafish (*noc3l*: Walter et al., 2009; *plce1*: Zhou et al., 2009).

4.3 / Biological functions of genes under positive or balancing selection

Among the 282 genes in the 13 regions detected under either positive or balancing selection, most genes seem to play essential roles in fitness as expected with such a dataset comprising both European and US populations. They are related to all main biological functions (genome stability, cell organization, neuronal and embryonic development, energy metabolism, growth, reproduction, and immunity). All identified biological functions were already described in other studies of signatures of selection in farmed rainbow trout (Cádiz et al., 2021) and other domesticated species (López et al., 2018, 2019, 2021; Naval-Sanchez et al., 2020; Baesjou & Wellenreuther, 2021; Signer-Hasler et al., 2022).

4.3.1 Hotspots of heterozygosity and balancing selection for fitness traits

In livestock species, many variants under balancing selection are known to improve performance in heterozygote state but cause defect in homozygous state (Hedrick, 2015; Georges et al., 2019). However, in such cases of balancing selection, there is generally only one homozygous state, which is deleterious at a locus level, while the alternative homozygous state is observed in the population. In our study, we highlight four regions potentially involved in balancing selection for which we observed a lack of any kind of long stretches of homozygosity. Even if these regions are extremely heterozygous, the proteins associated with these genes are highly conserved among vertebrates (Supplementary information S17). Many processes may explain these surprising observations at first glance. First of all, these regions may concentrate polymorphism in non-coding parts of the genome. Polymorphism in intronic regions may promote various proteins by allowing alternative splicing. We may also observe an excess of synonymous polymorphism in exons without effects on proteins. Further analyses must be conducted to better understand the mechanisms underlying the maintenance of extreme

polymorphism, whether to validate the hypothesis of balancing selection or the existence of high mutation and recombination rates in these specific regions.

In the heterozygous region chr10_a, the gene *ctnna2* (Table 8) enables actin filament binding activity and is involved in the regulation of neuron migration and neuron projection development. Thus, *ctnna2* plays an essential role in brain development among vertebrates (Uvarov et al., 2014). In yonlong grouper, *ctnna2* seems implicated in vertebral development, because significantly differentially expressed between normal and fish with lordosis (Li et al., 2022). In mice, a homozygous for a mutation of *ctnna2* reduced body weight, male fertility, and induced brain abnormalities (hypoplastic cerebellum, abnormal foliation pattern, ectopic Purkinje cells, and abnormal pyramidal cells in the hippocampus). While the protein associated to this gene is highly conserved among vertebrates (Uvarov et al., 2014; Supplementary information S17), the gene exhibits a strong polymorphism in all the four studied rainbow trout populations. However, a large part of its polymorphism is located in one intronic region (intron 6-7) of *ctnna2*. In the zfin database, five transcripts of this gene were identified (three mRNA and two non-coding RNA). We hypothesize that the polymorphism in the intronic region of *ctnna2* is essential for alternative splicing.

In the heterozygous region chr13_a (Supplementary information S16), *mmd* and *map2k4* are identified as highly conserved across vertebrates (Supplementary information S17). The gene *mmd* plays an important role in maturing macrophages, which is essential for immune response as observed in mice (Lin et al., 2021). The gene *map2k4* is implicated in a variety of cellular processes (proliferation, differentiation, transcription regulation, development), seems to play a role in liver organogenesis and embryonic development during gastrulation, as demonstrated by morpholino-mediated knockdown in zebrafish (Seo et al., 2010), and implicated in immune response in yellow catfish (Zheng et al., 2022). The inflammatory process in immune response seems linked to the polymorphism of the *map2k4* gene, which is consistent with our hypothesis of balancing selection, and more precisely potential ancestral trans-species polymorphism in this genomic region (Gu et al., 2016; Fijarczyk & Babik, 2015). Trans-species polymorphism is a crucial evolutionary mechanism for sharing adaptative genetic variation across taxa (Klein et al., 1998). The study of this mechanism has primarily concentrated on major histocompatibility complex genes, but a few studies described this process for other immune genes (Ferrer-Admetlla, et al., 2008; Leffler et al., 2013; Těšický & Vinkler, 2015). Maintaining genetic diversity in regions related to the immune system may be

essential to resilience against various pathogens. In addition, this region of chromosome 13 has been recently detected as a significant QTL playing a role on resistance to temperature (Lagarde et al., 2022).

In the heterozygous region chr19_a (Table 8), the protein encoded by *smarca5* is a component of chromatin remodeling and spacing factor RSF, a facilitator of the transcription of class II genes by RNA polymerase II (zebrafish: Armas et al., 2013; Ding et al., 2021; mice: Limi et al., 2018). The protein is highly conserved among vertebrates (Supplementary information S17), which is consistent with its essential role thought to regulate the transcription of many genes by altering the chromatin structure around those genes. In the same region chr19_a, *frem2* codes for an extracellular matrix protein required for maintenance of the integrity of skin and renal epithelia in zebrafish (Gautier et al., 2008). This protein is moderately conserved across vertebrates (Supplementary information S17). In a study searching for genomic regions with ancestral trans-species polymorphism shared between humans and chimpanzees (Leffler et al., 2013), *frem3*, an important paralog of *frem2*, was identified under balancing selection. However, further studies should test the hypothesis of trans-species conservation of *map2k4* and *frem2* genes that may help to understand the various cellular processes in which the gene is implicated.

In the heterozygous region chr19_b (Table 8), *pou4f2* protein is highly conserved among vertebrates (Supplementary information S17) and is a tissue-specific DNA-binding transcription factor involved in the development and differentiation of specific cells. It maintains the visual system neurons in the retina and the lateral line (DeCarvalho et al., 2004) and seems also related to cardiac development in zebrafish (Maskell et al., 2017).

4.3.2 Hotspots of homozygosity and positive selection for essential biological functions

4.3.2.1 | Regions and genes involved in cellular and nuclear organization, and embryonic development

In homozygous region chr2_a, three genes plays important roles in embryonic development and then fitness (*cep162*, *tsnare1*, *mrp2*, Table 6). In the homozygous region chr2_b (Table 6), the gene *pbx1* (pre-B-cell leukemia transcription factor 1) is related to early development in zebrafish (Teoh et al., 2010). Mutations in this gene generally cause major malformations, which seem to play an essential role in survival in various species (zebrafish: Teoh et al., 2010; mouse: Selleri et al., 2001; human: Le Tanno et al., 2017). It was detected under positive selection in a Chilean farmed rainbow trout population (Cádiz et al., 2021). However only

moderate percent identity (> 65%) is observed between *pbx1* proteins across vertebrates.

In the homozygous region chr15_a, many genes (*chn1*, *atp5mc1*, *zc3h15*, *nid2* and *brca2*) were playing essential roles in cell functioning and early development (Table 6). However only two of them were highly conserved among vertebrates (> 85%; *chn1* and *atp5mc1*). The gene *atp5mc1* is a crucial gene for mitochondrial cristae morphology, and plays important roles in metabolic processes associated to growth (Table 6; Palmer et al., 2011; Miller et al., 2019; Wang et al., 2020). In zebrafish, a morpholino knockdown of *chn1* reveals its crucial role in early development, revealing severe abnormalities (development of round somites, lack of yolk extension, and kinked posterior notochord) (Leskow et al., 2006).

Three genes located in close vicinity in region chr16_a (between 46.42 and 46.53 Mb; Supplementary information S11), *samd10* (sterile alpha motif domain-containing protein 10-like), *dnajc5* (dnaJ homolog subfamily C member 5-like), and *tpd54* (tumor protein D54) were also detected in close chromosomal vicinity and under positive selection in ten modern goat breeds and one wild Bezoar goat (Signer-Hasler et al., 2022). This cluster of genes has a significant role in survival and cellular processes (Table 6). In addition, in this region chr16_a, the protein of the gene *magi2* (membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2, Table 6) plays a vital role in embryogenesis in zebrafish (Borah et al., 2016). The gene *magi2* was also identified under positive selection in a domesticated sheep breed compared to the wild Asiatic mouflon (Cumer et al., 2021).

4.3.2.2 | Regions and genes involved in neural and brain development, and behaviour

In total, we identified 7 genes as primarily associated to neural and brain development in both regions detected under positive selection (*tsnare1*, *cdk14*, *brsk2a*, *auts2*, *brd2*, *znf135*, and *grxcr1*). Some genes (*brsk2a*, *znf135*, *grxcr1*, *auts2*; Table 6), related to brain development may induce behavior modifications in farmed animals, that may be related to domestication processes (Pasquet, 2018; Milla et al., 2021; Deng et al., 2022; Liu et al., 2022). This is in line with Żarski et al. (2020) study showing that domestication modulates gene expression involved in neurogenesis.

In particular, the gene *auts2* gene was previously identified under positive selection both in cattle (Consortium, bovine Hapmap, 2009) and in domesticated Atlantic salmon populations from Canada and Scotland compared to their wild Atlantic salmon counterpart (López et al., 2018). The gene *znf135* was also detected under positive selection in a farmed population of Atlantic salmon compared to a wild-type population (Gutierrez et al., 2016). The gene *grxcr1*

was detected under positive selection in the Tharparkar cattle (Saravanan et al., 2021). It strongly suggests that all these genes play a key role in domestication processes and may act on essential behaviours in both terrestrial and aquatic farmed animals.

4.3.2.3 / Regions and genes involved in growth metabolism

Genes related to growth metabolism were only identified in four regions under positive selection, none of them were detected in high heterozygosity regions.

In the homozygous region chr2_a (Table 6), the protein *mrap2* (*melanocortin-2 receptor accessory protein 2A*) is associated to growth. A lack of this gene exhibit severe obesity in many species (human, zebrafish, rodent: Liu et al., 2013; sea lamprey: Zhu et al., 2019; snakehead: Wen et al., 2021). Yoshida et al. (2017) detected a growth-QTL in Atlantic salmon and considered *mrap2* as a candidate gene for growth up to 25 months. In addition, *mrap2* was identified in the Chilean farmed rainbow trout population as under positive selection (Cádiz et al., 2021). A QTL related to sea lice resistance in rainbow trout (Cáceres et al., 2021) was also detected in the region chr2_a (located from 10.43 Mb to 11.81 Mb of the Swanson reference genome, which corresponds to 26.69 Mb – 28.07 Mb of the Arlee reference genome). Cáceres et al. (2021) explained that having a high potential for growth seem essential for sea lice resistance. Indeed, proteomic investigations allow to establish a link between growth and immune function in salmonids (Causey, 2018).

In homozygous region chr2_b (Table 6), the *col9a2* (collagen alpha-2(IX) chain) gene is a component of cartilage and seems also related to growth (Xu et al., 2022).-This gene is detected under positive selection in a Scottish farmed population of Atlantic salmon (López et al., 2018). In addition, the gene *scap* (sterol regulatory element-binding protein cleavage-activating protein) was already identified under positive selection in six farmed Pacific white Shrimp populations (Wang et al., 2022).

In the homozygous region chr2_d (Table 6), the gene *igf-1a* (insuline like growth factor receptor 1a) plays an important role in growth and transformation events. In salmonids, expressions of *igf-1a* and growth hormone were demonstrated to be modified between domesticated and wild populations of rainbow trout and coho salmon (Tymchuk et al., 2009). The same observation was made with a higher expression of *igf-1a* between larvae from domesticated spawners than larvae from wild spawners of the Eurasian perch (Palińska-Żarska et al., 2021). In addition, *igf-1* was also observed as a marker of domestication in dogs (Wayne & vonHoldt, 2012).

In the homozygous region chr16_a, the *emilin-3a* gene (elastin microfibril interfacier 3a, Table 6) plays a role in extracellular matrix organization and elastic fiber formation. Its gene expression is related to embryonic development and involved in muscle fiber development in zebrafish (Milanetto et al., 2008). *Emilin-3a* had already been identified as under positive selection in one population of F2 Australian snapper farmed population compared to the first generation (F1) of domestication of a wild population (Baesjou & Wellenreuther, 2021). Thus, this signature of selection can be considered as a result of the domestication process.

All identified growth-related genes seem associated with domestication. This assertion is confirmed for five genes (*mrp2*, *col9a2*, *scap*, *igf-1a*, *emilin-3*) that were also identified under positive selection in various farmed populations with favorable alleles linked to better growth phenotypes (Table 6).

4.3.2.4 / Regions and genes involved in reproduction

Very few genes directly related to reproduction traits were only identified in highly homozygous regions.

In the homozygous region chr2_b, the *brd2* (bromodomain-containing protein 2, Table 6) gene is implicated in several biological process (see section 4.4.1.). It seems related to oogenesis and egg-to-embryo transition in zebrafish (DiBenedetto et al., 2008), which is consistent with a QTL detected for egg size in this region in rainbow trout (D'Ambrosio et al., 2020). Moreover, it seems that *brd2* is involved in spermatogenesis or folliculogenesis, as demonstrated in situ on mice cells (Rhee et al., 1998). Khendek et al. (2017) compared the reproductive performances (egg size, gonadal histology, hormonal levels) between domesticated and F1 with wild broodstock of Eurasian perch populations. They showed that domestication may have increased the oocyte diameter and the level of 17 β -Estradiol, but decreased the embryo survival of domesticated fish. This gene was also identified under positive selection in a selected Canadian population of Atlantic salmon (López et al., 2018).

In the homozygous region chr15_a (Table 6), the gene *zp4* has already been identified under positive selection in a farmed Scottish population of Atlantic salmon compared to a wild population (López et al., 2018), and may be related to domestication process.

4.3.2.5 / Regions and genes involved in immunity

Magris et al. (2022) observed that regions under positive selection revealed an enrichment of KEGG terms related to viral infection in farmed brown trout. However, it should be noticed that in our study, few genes related to immune function were identified and no enrichment in immune terms was observed in GO analysis.

Genes related to immune function were mainly identified in three different regions detected as putatively under positive selection for two of them and under balancing selection for the last one.

In the homozygous region chr2_b (Table 6), genes *mfaip8l2b* (tumor necrosis factor, alpha-induced protein 8-like protein 2 B) and *atg5* (autophagy protein 5) are related to immune functions. Note that *atg5* is well conserved across vertebrates (> 80%).

In the region chr15_a, the gene *zc3h15* (Table 6) seems to have an inhibitory effect on HIV-1 replication and then on HIV infection in vitro (mice cells) (Capalbo et al., 2010).

A last gene in the region chr16_a, the gene *atp1b3* (sodium/potassium-transporting ATPase subunit beta-1-interacting protein 3) was also identified under positive selection in farmed Atlantic Salmon (Naval-Sanchez et al., 2020). In the Senegalese sole, *atp1b3a* and *atp1b3b* paralogs have been hypothesized to be involved in response to low salinity (Armesto et al., 2015). In addition, this gene is involved in some immune responses. It was shown in cell culture study, that *atp1b3* inhibits hepatitis B virus replication via inducing NF-kappa B activation (human: Zhang et al., 2021) and is involved in numerous viral propagation such as HIV and EV71 (Zheng et al., 2020) in cell culture experiments.

4.4. Conclusion

To sum up, we identified 13 regions under selection with numerous genes strongly involved in essential biological functions. By identifying signatures of selection shared by our four studied populations, we have focused our detection on regions related to ancient evolutionary processes that are essentially important for species survival. We only identified nine homozygous regions presumably under positive selection and four heterozygous regions putatively under balancing selection in four different rainbow populations. While common homozygous regions may be associated with important biological functions underlying both fish fitness and domestication, the heterozygous regions seem mainly linked to fitness functions (cell organization, embryonic

development, and immunity) which are involved at different developmental stages or to cope with various pathogens or abiotic stressors. Maintaining genetic diversity in these regions could be essential for the species survival.

This study allows us to confirm the importance of a large set of 17 genes already detected as under positive selection in previous studies, among which 10 genes were identified in fishes (*auts2*, *atp1b3*, *zfp4*, *znf135*, *igf-1a*, *brd2*, *col9a2*, *mrp2*, *pbx1* and *emilin-3*). We also identify new promising candidate genes as important for rainbow trout fitness. In our opinion, this study substantially increases knowledge of evolutionary processes and helps to determine the genomic location and the nature of the genetic variation that must be maintained in rainbow trout populations for domestication and selection purposes.

References

- Armas, P., Margarit, E., Mouguelar, V. S., Allende, M. L., & Calcaterra, N. B. (2013). Beyond the binding site: in vivo identification of *tbx2*, *smarca5* and *wnt5b* as molecular targets of CNBP during embryonic development. *PloS one*, 8(5), e63234. <https://doi.org/10.1371/journal.pone.0063234>
- Armesto, P., Infante, C., Cousin, X., Ponce, M., & Manchado, M. (2015). Molecular and functional characterization of seven Na⁺/K⁺-ATPase β subunit paralogs in Senegalese sole (*Solea senegalensis* Kaup, 1858). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 182, 14-26. <https://doi.org/10.1016/j.cbpb.2014.11.011>
- Baesjou, J. P., & Wellenreuther, M. (2021). Genomic signatures of domestication selection in the Australasian snapper (*Chrysophrys auratus*). *Genes*, 12(11), 1737. <https://doi.org/10.3390/genes12111737>
- Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M. N., Van Tassell, C., Sonstegard, T., ... & Hanotte, O. (2015). Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Scientific reports*, 5(1), 1-13. <https://doi.org/10.1038/srep11729>
- Baker, S., Booth, C., Fillman, C., Shapiro, M., Blair, M. P., Hyland, J. C., & Ala-Kokko, L. (2011). A loss of function mutation in the COL9A2 gene causes autosomal recessive Stickler syndrome. *American Journal of Medical Genetics Part A*, 155(7), 1668-1672. <https://doi.org/10.1002/ajmg.a.34071>
- Balon, E. K. (1995). Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture*, 129(1-4), 3-48. [https://doi.org/10.1016/0044-8486\(94\)00227-F](https://doi.org/10.1016/0044-8486(94)00227-F)

Balon, E. K. (2004). About the oldest domesticates among fishes. *Journal of fish Biology*, 65, 1-27. <https://doi.org/10.1111/j.0022-1112.2004.00563.x>

Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M., ... & Phocas, F. (2022). Development of a high-density 665K SNP array for rainbow trout genome-wide genotyping. *Frontiers in Genetics*, 13, 1664-8021 <https://doi.org/10.3389/fgene.2022.941340>

Borah, S., Barrodia, P., & Swain, R. K. (2016). Nucleolar protein 4-like has a complex expression pattern in zebrafish embryos. *International Journal of Developmental Biology*, 60(1-2-3), 53-56. <https://doi.org/10.1387/ijdb.150307rs>

Bovine HapMap Consortium, Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., ... & Dodds, K. G. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324(5926), 528-532.

Browning, S. R., & Browning, B. L. (2011). Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics*, 12(10), 703-714. <https://doi.org/10.1038/nrg3054>

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4(11), 900-910. <https://doi.org/10.1038/nrg1203>

Cáceres, P., Barría, A., Christensen, K. A., Bassini, L. N., Correa, K., Garcia, B., ... & Yáñez, J. M. (2021). Genome-scale comparative analysis for host resistance against sea lice between Atlantic salmon and rainbow trout. *Scientific reports*, 11(1), 1-11. <https://doi.org/10.1038/s41598-021-92425-3>

Cadieux-Dion, M., Andermann, E., Lachance-Touchette, P., Ansorge, O., Meloche, C., Barnabe, A., ... & Cossette, P. (2013). Recurrent mutations in DNAJC5 cause autosomal dominant Kufs disease. *Clinical genetics*, 83(6), 571-575.

Cádiz, M. I., López, M. E., Díaz-Domínguez, D., Cáceres, G., Yoshida, G. M., Gomez-Uchida, D., & Yáñez, J. M. (2020). Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (*Oreochromis niloticus*). *Scientific reports*, 10(1), 1-14. <https://doi.org/10.1038/s41598-020-68064-5>

Cádiz, M. I., López, M. E., Díaz-Domínguez, D., Cáceres, G., Marin-Nahuelpi, R., Gomez-Uchida, D., ... & Yáñez, J. M. (2021). Detection of selection signatures in the genome of a farmed population of anadromous rainbow trout (*Oncorhynchus mykiss*). *Genomics*, 113(5), 3395-3404. <https://doi.org/10.1016/j.ygeno.2021.07.027>

Capalbo, G., Müller-Kuller, T., Dietrich, U., Hoelzer, D., Ottmann, O. G., & Scheuring, U. J. (2010). Inhibition of X4-tropic HIV type 1 replication by knockdown of the cellular protein LEREPO4. *AIDS research and human retroviruses*, 26(10), 1155-1161. <https://doi.org/10.1089/aid.2010.0041>

Causey, D. R. (2018). Proteomic and molecular investigations of links between growth and immune function in salmonids (Doctoral dissertation, University of Aberdeen).

Cazorla-Vázquez, S., & Engel, F. B. (2018). Adhesion GPCRs in kidney development and disease. *Frontiers in Cell and Developmental Biology*, 6, 9. <https://doi.org/10.3389/fcell.2018.00009>

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742-015. <https://doi.org/10.1186/s13742-015-0047-8>

Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, 141(4), 1619-1632. <https://doi.org/10.1093/genetics/141.4.1619>

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS genetics*, 2(4), e64. <https://doi.org/10.1371/journal.pgen.0020064>

Chen, C. C., Feng, W., Lim, P. X., Kass, E. M., & Jasin, M. (2018). Homology-directed repair and the role of BRCA1, BRCA2, and related proteins in genome integrity and cancer. *Annual review of cancer biology*, 2, 313-336. <https://doi.org/10.1146/annurev-cancerbio-030617-050502>

Christie, M. R., Marine, M. L., Fox, S. E., French, R. A., & Blouin, M. S. (2016). A single generation of domestication heritably alters the expression of hundreds of genes. *Nature communications*, 7(1), 1-6. <https://doi.org/10.1038/ncomms10676>

Cumer, T., Boyer, F., & Pompanon, F. (2021). Genome-wide detection of structural variations reveals new regions associated with domestication in small ruminants. *Genome biology and evolution*, 13(8), evab165. <https://doi.org/10.1093/gbe/evab165>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*, Murray, London

Darwin, C. (1868). The variation of animals and plants under domestication. Volume 2 (London: J. Murray).

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., ... & Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51(1), 1-15. <https://doi.org/10.1186/s12711-019-0468-4>

D'Ambrosio, J., Morvezen, R., Brard-Fudulea, S., Bestin, A., Acin Perez, A., Guéméné, D., ... & Phocas, F. (2020). Genetic architecture and genomic selection of female reproduction traits in rainbow trout. *BMC genomics*, 21(1), 1-14. <https://doi.org/10.1186/s12864-020-06955-7>

DeCarvalho, A. C., Cappendijk, S. L., & Fadool, J. M. (2004). Developmental expression of the POU domain transcription factor Brn-3b (Pou4f2) in the lateral line and visual system of zebrafish. *Developmental dynamics: an official publication of the American Association of Anatomists*, 229(4), 869-876. <https://doi.org/10.1002/dvdy.10475>

Deng, J., Wang, Y., Hu, M., Lin, J., Li, Q., Liu, C., & Xu, X. (2022). Deleterious Variation in BR Serine/Threonine Kinase 2 Classified a Subtype of Autism. *Frontiers in Molecular Neuroscience*, 15. <https://doi.org/10.3389/fnmol.2022.904935>

DiBenedetto, A. J., Guinto, J. B., Ebert, T. D., Bee, K. J., Schmidt, M. M., & Jackman, T. R. (2008). Zebrafish brd2a and brd2bare paralogous members of the bromodomain-ET (BET) family of transcriptional coregulators that show structural and expression divergence. *BMC developmental biology*, 8(1), 1-19. <https://doi.org/10.1186/1471-213X-8-39>

Ding, Y., Wang, W., Ma, D., Liang, G., Kang, Z., Xue, Y., ... & Liu, F. (2021). Smarca5-mediated epigenetic programming facilitates fetal HSPC development in vertebrates. *Blood*, 137(2), 190-202. <https://doi.org/10.1182/blood.2020005219>

Dobney, K., & Larson, G. (2006). Genetics and animal domestication: new windows on an elusive process. *Journal of zoology*, 269(2), 261-271. <https://doi.org/10.1111/j.1469-7998.2006.00042.x>

Durland, E., De Wit, P., & Langdon, C. (2021). Temporally balanced selection during development of larval Pacific oysters (*Crassostrea gigas*) inherently preserves genetic diversity within offspring. *Proceedings of the Royal Society B*, 288(1958), 20203223. <https://doi.org/10.1098/rspb.2020.3223>

East, E. M. (1918). The role of reproduction in evolution. *The American Naturalist*, 52(618/619), 273-289. <https://www.journals.uchicago.edu/doi/pdf/10.1086/279670>

Fabrice, T. (2018). Fish domestication: an overview. *Animal domestication*. <https://doi.org/10.5772/intechopen.79628>

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marquès-Bonet, T., Ramírez-Soriano, A., Muntasell, A., ... & Casals, F. (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *The Journal of Immunology*, 181(2), 1315-1322. <https://doi.org/10.4049/jimmunol.181.2.1315>

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular ecology*, 24(14), 3529-3545. <https://doi.org/10.1111/mec.13226>

Fisher, RA. (1958). The genetical theory of natural selection.

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., ... & Sklar, P. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience*, 19(11), 1442-1453. <https://doi.org/10.1038/nn.4399>

Gall, G.A.E., Crandell, P.A. (1992). The rainbow trout. *Aquaculture*, 100(1-3), 1-10. [https://doi.org/10.1016/0044-8486\(92\)90333-G](https://doi.org/10.1016/0044-8486(92)90333-G)

Gao, G., Nome, T., Pearse, D. E., Moen, T., Naish, K. A., Thorgaard, G. H., ... & Palti, Y. (2018). A new single nucleotide polymorphism database for rainbow trout generated through whole genome resequencing. *Frontiers in genetics*, 9, 147. <https://doi.org/10.3389/fgene.2018.00147>

Gao, G., Magadan, S., Waldbieser, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., ... & Palti, Y. (2021). A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. *G3*, 11(4), jkab052. <https://doi.org/10.1093/g3journal/jkab052>

Gautier, M., & Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28(8), 1176-1177. <https://doi.org/10.1093/bioinformatics/bts115>

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplement of the R package rehh to detect positive selection from haplotype structure. *Molecular ecology resources*, 17(1), 78-90. <https://doi.org/10.1111/1755-0998.12634>

Gautier, P., Naranjo-Golborne, C., Taylor, M. S., Jackson, I. J., & Smyth, I. (2008). Expression of the *fras1/frem* gene family during zebrafish development and fin morphogenesis. *Developmental dynamics: an official publication of the American Association of Anatomists*, 237(11), 3295-3304. <https://doi.org/10.1002/dvdy.21729>

Georges, M., Charlier, C., & Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nature Reviews Genetics*, 20(3), 135-156.

Glover, K. A., Otterå, H., Olsen, R. E., Slinde, E., Taranger, G. L., & Skaala, Ø. (2009). A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. *Aquaculture*, 286(3-4), 203-210. <https://doi.org/10.1016/j.aquaculture.2008.09.023>

Gorssen, W., Meyermans, R., Janssens, S., & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, 53(1), 1-10. <https://doi.org/10.1186/s12711-020-00599-7>

Grilz-Seger, G., Mesarič, M., Cotman, M., Neuditschko, M., Druml, T., & Brem, G. (2018). Runs of homozygosity and population history of three horse breeds with small population size. *Journal of Equine Veterinary Science*, 71, 27-34. <https://doi.org/10.1016/j.jevs.2018.09.004>

Gross, M. R. (1998). One species with two biologies: Atlantic salmon (*Salmo salar*) in the wild and in aquaculture. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(S1), 131-144.

Gross, R., Lulla, P., & Paaver, T. (2007). Genetic variability and differentiation of rainbow trout (*Oncorhynchus mykiss*) strains in northern and Eastern Europe. *Aquaculture*, 272, S139-S146.

<https://doi.org/10.1016/j.aquaculture.2007.08.004>

Gu, L., Wu, Y., Hu, S., Chen, Q., Tan, J., Yan, Y., ... & Tang, N. (2016). Analysis of association between MAP2K4 gene polymorphism rs3826392 and IL-1b serum level in southern Chinese Han Ischemic Stroke Patients. *Journal of Stroke and Cerebrovascular Diseases*, 25(5), 1096-1101. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.12.035>

Gutierrez, A. P., Yáñez, J. M., & Davidson, W. S. (2016). Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Marine genomics*, 26, 41-50. <https://doi.org/10.1016/j.margen.2015.12.007>

Hansen, M. M., & Mensberg, K. L. D. (1998). Genetic differentiation and relationship between genetic and geographical distance in Danish sea trout (*Salmo trutta* L.) populations. *Heredity*, 81(5), 493-504. <https://doi.org/10.1046/j.1365-2540.1998.00408.x>

Harris, D. R. (2007). Agriculture, cultivation and domestication: exploring the conceptual framework of early food production. *Rethinking agriculture: archaeological and ethnoarchaeological perspectives*, 16, 35.

Hedrick, P. W. (2015). Heterozygote advantage: the effect of artificial selection in livestock and pets. *Journal of Heredity*, 106(2), 141-154.

Helmer, D. (1992). *La domestication des animaux par les hommes préhistoriques*. FeniXX.

Hershberger, W. K. (1992). Genetic variability in rainbow trout populations. *Aquaculture*, 100(1-3), 51-71. [https://doi.org/10.1016/0044-8486\(92\)90339-M](https://doi.org/10.1016/0044-8486(92)90339-M)

Hiatt, S. M., Thompson, M. L., Prokop, J. W., Lawlor, J. M., Gray, D. E., Bebin, E. M., ... & Cooper, G. M. (2019). Deleterious variation in *brsk2* associates with a neurodevelopmental disorder. *The American Journal of Human Genetics*, 104(4), 701-708. <https://doi.org/10.1016/j.ajhg.2019.02.002>

Hong Xia, J., Bai, Z., Meng, Z., Zhang, Y., Wang, L., Liu, F., ... & Hua Yue, G. (2015). Signatures of selection in tilapia revealed by whole genome resequencing. *Scientific reports*, 5(1), 1-10. <https://doi.org/10.1038/srep14168>

Hori, K., Shimaoka, K., & Hoshino, M. (2021). *AUTS2* gene: keys to understanding the pathogenesis of neurodevelopmental disorders. *Cells*, 11(1), 11. <https://doi.org/10.3390/cells11010011>

Howarth, D. L., Yin, C., Yeh, K., & Sadler, K. C. (2013). Defining hepatic dysfunction parameters in two models of fatty liver disease in zebrafish larvae. *Zebrafish*, 10(2), 199-210. <https://doi.org/10.1089/zeb.2012.0821>

Illa, S. K., Mumtaz, S., Mukherjee, S., Nath, S., & Mukherjee, A. (2022). Genome-wide evaluation of runs of homozygosity in Indian Sahiwal cattle. *The 12th World Congress on Genetic Applied to Livestock*.

Ioerger, T. R., Clark, A. G., & Kao, T. H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences*, 87(24), 9732-9735. <https://doi.org/10.1073/pnas.87.24.9732>

Ip, J. P., Shi, L., Chen, Y., Itoh, Y., Fu, W. Y., Betz, A., ... & Ip, N. Y. (2012). α 2-chimaerin controls neuronal migration and functioning of the cerebral cortex through CRMP-2. *Nature neuroscience*, 15(1), 39-47. <https://doi.org/10.1038/nn.2972>

Ishikawa, K., Azuma, S., Ikawa, S., Semba, K., & Inoue, J. I. (2005). Identification of DRG family regulatory proteins (DFRPs): specific regulation of DRG1 and DRG2. *Genes to Cells*, 10(2), 139-150.

Jombart, T., & Ahmed, I. (2011). Adegnet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071. <https://doi.org/10.1093/bioinformatics/btr521>

Khendek, A., Alix, M., Viot, S., Ledoré, Y., Rousseau, C., Mandiki, R., ... & Milla, S. (2017). How does a domestication process modulate oogenesis and reproduction performance in Eurasian perch?. *Aquaculture*, 473, 206-214. <https://doi.org/10.1016/j.aquaculture.2017.02.003>

Kishi, M., Pan, Y. A., Crump, J. G., and Sanes, J. R. (2005). Mammalian SAD kinases are required for neuronal polarization. *Science* 307, 929–932. <https://doi.org/10.1126/science.1107403>

Klein, J., Sato, A., Nagl, S., & O'hUigín, C. (1998). Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, 1-C1. <https://www.jstor.org/stable/221700>

Leffler, E. M., Gao, Z., Pfeifer, S., Ségurel, L., Auton, A., Venn, O., ... & Przeworski, M. (2013). Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, 339(6127), 1578-1582.

Lagarde, H., Lallias, D., Patrice, P., Prchal, M., Francois, Y., D'Ambrosio, J., ... & Phocas, F. (2022). Genetic architecture of acute hyperthermia resistance in juvenile rainbow trout (*Oncorhynchus mykiss*) and genetic correlations with production traits. *bioRxiv*. <https://doi.org/10.1101/2022.11.14.516387>

Le Tanno, P., Breton, J., Bidart, M., Satre, V., Harbuz, R., Ray, P. F., ... & Coutton, C. (2017). PBX1 haploinsufficiency leads to syndromic congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *Journal of Medical Genetics*, 54(7), 502-510. <http://dx.doi.org/10.1136/jmedgenet-2016-104435>

Leskow, F. C., Holloway, B. A., Wang, H., Mullins, M. C., & Kazanietz, M. G. (2006). The zebrafish homologue of mammalian chimerin Rac-GAPs is implicated in epiboly progression during development. *Proceedings of the National Academy of Sciences*, 103(14), 5373-5378. <https://doi.org/10.1073/pnas.0508585103>

Li, Q., Hou, J., Guo, C., Zhang, Y., Liu, L., Hu, H., ... & Cui, H. (2021a). ZC3H15 Correlates with a Poor Prognosis and Tumor Progression in Melanoma. *BioMed Research International*, 2021. <https://doi.org/10.1155/2021/8305299>

Li, H., Zhu, Q., Chen, R., Liu, M., & Xu, D. (2021b). Identification and Characterization of Dimorphic Expression of Sex-Related Genes in Rock Bream, a Fish With Multiple Sex Chromosomes. *Frontiers in genetics*, 12. <https://doi.org/10.3389/fgene.2021.791179>

Li, Z., Tian, Y., Wang, L., Li, Z., Chen, S., Li, L., ... & Zhai, J. (2022). Comparative transcriptomics analyses and revealing candidate networks and genes involved in lordosis of the Yunlong grouper (*Epinephelus moara*♀× *Epinephelus lanceolatus*♂). *Aquaculture*, 550, 737866. <https://doi.org/10.1016/j.aquaculture.2021.737866>

Limi, S., Senecal, A., Coleman, R., Lopez-Jones, M., Guo, P., Polumbo, C., ... & Cvekl, A. (2018). Transcriptional burst fraction and size dynamics during lens fiber cell differentiation and detailed insights into the denucleation process. *Journal of Biological Chemistry*, 293(34), 13176-13190. <https://doi.org/10.1074/jbc.RA118.001927>

Lin, Y., Huang, M., Wang, S., You, X., Zhang, L., & Chen, Y. (2021). PAQR11 modulates monocyte-to-macrophage differentiation and pathogenesis of rheumatoid arthritis. *Immunology*, 163(1), 60-73. <https://doi.org/10.1111/imm.13303>

Liu, T., Elmquist J.K. & Williams K.W. (2013) Mrap2: an accessory protein linked to obesity. *Cell Metabolism* 18, 309–11. <https://doi.org/10.1016/j.cmet.2013.08.016>

Liu, L., Ang, K. P., Elliott, J. A., Kent, M. P., Lien, S., MacDonald, D., & Boulding, E. G. (2017a). A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: Testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evolutionary applications*, 10(3), 276-296. <https://doi.org/10.1111/eva.12450>

Liu, S., Palti, Y., Martin, K. E., Parsons, J. E., & Rexroad III, C. E. (2017b). Assessment of genetic differentiation and genetic assignment of commercial rainbow trout strains using a SNP panel. *Aquaculture*, 468, 120-125. <https://doi.org/10.1016/j.aquaculture.2016.10.004>

Liu, S., Huang, J., Zhang, Y., Liu, Y., Zuo, S., & Li, R. (2019). MAP2K4 interacts with Vimentin to activate the PI3K/AKT pathway and promotes breast cancer pathogenesis. *Aging (Albany NY)*, 11(22), 10697. <https://doi.org/10.18632/aging.102485>

Liu, C., & Zhao, B. (2021). Murine GRXCR1 has a different function than GRXCR2 in the morphogenesis of stereocilia. *Frontiers in Cellular Neuroscience*, 276. <https://doi.org/10.3389/fncel.2021.714070>

Liu, Y., Wang, Y., Li, N., and Jiang, S. (2022). Avobenzone and nanoplastics affect the development of zebrafish nervous system and retinal system and inhibit their locomotor behavior. *Sci. Total Environ.* 806:150681. <https://doi.org/10.1016/j.scitotenv.2021.150681>

López, M. E., Neira, R., & Yáñez, J. M. (2015). Applications in the search for genomic selection signatures in fish. *Frontiers in genetics*, 5, 458. <https://doi.org/10.3389/fgene.2014.00458>

López, M. E., Benestan, L., Moore, J. S., Perrier, C., Gilbey, J., Di Genova, A., ... & Yáñez, J. M. (2018). Comparing genomic signatures of domestication in two Atlantic salmon (*Salmo salar* L.) populations with different geographical origins. *Evolutionary Applications*, *12*(1), 137-156. <https://doi.org/10.1111/eva.12689>

López, M. E., Linderoth, T., Norris, A., Lhorente, J. P., Neira, R., & Yáñez, J. M. (2019). Multiple selection signatures in farmed atlantic salmon adapted to different environments across hemispheres. *Frontiers in Genetics*, *10*, 901. <https://doi.org/10.3389/fgene.2019.00901>

López, M. E., Cádiz, M. I., Rondeau, E. B., Koop, B. F., & Yáñez, J. M. (2021). Detection of selection signatures in farmed coho salmon (*Oncorhynchus kisutch*) using dense genome-wide information. *Scientific reports*, *11*(1), 1-13. <https://doi.org/10.1038/s41598-021-86154-w>

Limborg, M. T., Blankenship, S. M., Young, S. F., Utter, F. M., Seeb, L. W., Hansen, M. H., & Seeb, J. E. (2012). Signatures of natural selection among lineages and habitats in *Oncorhynchus mykiss*. *Ecology and Evolution*, *2*(1), 1-18. <https://doi.org/10.1002/ece3.59>

Lorente-Cánovas, B., Eckrich, S., Lewis, M. A., Johnson, S. L., Marcotti, W., & Steel, K. P. (2022). *Grxcr1* regulates hair bundle morphogenesis and is required for normal mechano-electrical transduction in mouse cochlear hair cells. *Plos one*, *17*(3), e0261530. <https://doi.org/10.1371/journal.pone.0261530>

Magris, G., Marroni, F., D'Agaro, E., Vischi, M., Chiabà, C., Scaglione, D., ... & Morgante, M. (2022). ddRAD-seq reveals the genetic structure and detects signals of selection in Italian brown trout. *Genetics Selection Evolution*, *54*(1), 1-14. <https://doi.org/10.1186/s12711-022-00698-7>

Mäkinen, H., Vasemägi, A., McGinnity, P., Cross, T. F., & Primmer, C. R. (2015). Population genomic analyses of early-phase Atlantic salmon (*Salmo salar*) domestication/captive breeding. *Evolutionary applications*, *8*(1), 93-107. <https://doi.org/10.1111/eva.12230>

Margarit, E., Armas, P., Siburu, N. G., & Calcaterra, N. B. (2014). CNBP modulates the transcription of Wnt signaling pathway components. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1839*(11), 1151-1160. <https://doi.org/10.1016/j.bbagr.2014.08.009>

Martínez, A., Garza, J. C., & Pearse, D. E. (2011). A microsatellite genome screen identifies chromosomal regions under differential selection in steelhead and rainbow trout. *Transactions of the American Fisheries Society*, *140*(3), 829-842. <https://doi.org/10.1080/00028487.2011.588094>

Maskell, L. J., Qamar, K., Babakr, A. A., Hawkins, T. A., Heads, R. J., & Budhram-Mahadeo, V. S. (2017). Essential but partially redundant roles for POU4F1/Brn-3a and POU4F2/Brn-3b transcription factors in the developing heart. *Cell Death & Disease*, *8*(6), e2861-e2861. <https://doi.org/10.1038/cddis.2017.185>

Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutura, A. M., Di Gerlando, R., & Portolano, B. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution*, *49*(1), 1-10. <https://doi.org/10.1186/s12711->

017-0360-z

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., ... & Wilson, J. F. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83(3), 359-372. <https://doi.org/10.1016/j.ajhg.2008.08.007>

Mignon-Grasteau, S., Boissy, A., Bouix, J., Faure, J. M., Fisher, A. D., Hinch, G. N., ... & Beaumont, C. (2005). Genetics of adaptation and domestication in livestock. *Livestock Production Science*, 93(1), 3-14. <https://doi.org/10.1016/j.livprodsci.2004.11.001>

Milla, S., Pasquet, A., El Mohajer, L., & Fontaine, P. (2021). How domestication alters fish phenotypes. *Reviews in Aquaculture*, 13(1), 388-405. <https://doi.org/10.1111/raq.12480>

Milanetto, M., Tiso, N., Braghetta, P., Volpin, D., Argenton, F., & Bonaldo, P. (2008). Emilin genes are duplicated and dynamically expressed during zebrafish embryonic development. *Developmental dynamics: an official publication of the American Association of Anatomists*, 237(1), 222-232. <https://doi.org/10.1002/dvdy.21402>

Miller, B. C., Zhao, Z., Stephenson, L. M., Cadwell, K., Pua, H. H., Lee, H. K., ... & Virgin, IV, H. W. (2008). The autophagy gene ATG5 plays an essential role in B lymphocyte development. *Autophagy*, 4(3), 309-314. <https://doi.org/10.4161/auto.5474>

Miller, T. E., Henkels, K. M., Huddleston, M., Salisbury, R., Hussain, S. M., Sasaki, A. T., & Cho, K. J. (2019). Depletion of phosphatidylinositol 4-phosphate at the Golgi translocates K-Ras to mitochondria. *Journal of cell science*, 132(16), jcs231886. <https://doi.org/10.1242/jcs.231886>

Miyake, N., Chilton, J., Psatha, M., Cheng, L., Andrews, C., Chan, W. M., ... & Engle, E. C. (2008). Human CHN1 mutations hyperactivate $\alpha 2$ -chimaerin and cause Duane's retraction syndrome. *Science*, 321(5890), 839-843.

Moynahan, M. E., Pierce, A. J., & Jasin, M. (2001). BRCA2 is required for homology-directed repair of chromosomal breaks. *Molecular cell*, 7(2), 263-272. [https://doi.org/10.1016/S1097-2765\(01\)00174-5](https://doi.org/10.1016/S1097-2765(01)00174-5)

Mukudai, Y., Kondo, S., Fujita, A., Yoshihama, Y., Shirota, T., & Shintani, S. (2013). Tumor protein D54 is a negative regulator of extracellular matrix-dependent migration and attachment in oral squamous cell carcinoma-derived cell lines. *Cellular Oncology*, 36(3), 233-245. <https://doi.org/10.1007/s13402-013-0131-y>

Muragaki, Y., Mariman, E., van Beersum, S. E., Perälä, M., van Mourik, J., Warman, M. L., ... & Hamel, B. C. (1996). A mutation in the gene encoding the $\alpha 2$ chain of the fibril-associated collagen IX, COL9A2, causes multiple epiphyseal dysplasia (EDM2). *Nature genetics*, 12(1), 103-105. <https://doi.org/10.1038/ng0196-103>

Nakanishi, K., Niida, H., Tabata, H., Ito, T., Hori, Y., Hattori, M., et al. (2019). Isozyme-Specific Role of SAD-A in Neuronal Migration During Development of Cerebral Cortex. *Cereb. Cortex* 29, 3738–3751.

<https://doi.org/10.1093/cercor/bhy253>

Naval-Sanchez, M., McWilliam, S., Evans, B., Yáñez, J. M., Houston, R. D., & Kijas, J. W. (2020). Changed patterns of genomic variation following recent domestication: selection sweeps in farmed Atlantic salmon. *Frontiers in genetics, 11*, 264. <https://doi.org/10.3389/fgene.2020.00264>

Nosková, L., Stránecký, V., Hartmannová, H., Přistoupilová, A., Barešová, V., Ivánek, R., ... & Kmoch, S. (2011). Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *The American Journal of Human Genetics, 89*(2), 241-252. <https://doi.org/10.1016/j.ajhg.2011.07.003>

Nothnagel, M., Lu, T. T., Kayser, M., & Krawczak, M. (2010). Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human molecular genetics, 19*(15), 2927-2935. <https://doi.org/10.1093/hmg/ddq198>

Oksenberg, N., Stevison, L., Wall, J. D., & Ahituv, N. (2013). Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS genetics, 9*(1), e1003221. <https://doi.org/10.1371/journal.pgen.1003221>

Palińska-Żarska, K., Król, J., Woźny, M., Kamaszewski, M., Szudrowicz, H., Wiechetek, W., ... & Żarski, D. (2021). Domestication affected stress and immune response markers in *Perca fluviatilis* in the early larval stage. *Fish & Shellfish Immunology, 114*, 184-198. <https://doi.org/10.1016/j.fsi.2021.04.028>

Palmer, C. S., Osellame, L. D., Stojanovski, D., & Ryan, M. T. (2011). The regulation of mitochondrial morphology: intricate mechanisms and dynamic machinery. *Cellular signalling, 23*(10), 1534-1545. <https://doi.org/10.1016/j.cellsig.2011.05.021>

Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., ... & Rexroad III, C. E. (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular ecology resources, 14*(3), 588-596. <https://doi.org/10.1111/1755-0998.12204>

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., ... & Moen, T. (2015). The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. *Molecular ecology resources, 15*(3), 662-672. <https://doi.org/10.1111/1755-0998.12337>

Pasquet, A. (2018). Effects of domestication on fish behaviour. In *Animal domestication*. IntechOpen. <https://doi.org/10.5772/intechopen.78752>

Paul, K., Pelissier P., Goardon L., Dechamp N., Danon J., Jaffrelo L., Poncet C., Dupont-Nivet M., Phocas, F. (*In review*) Maternal and genetic effects on embryonic survival from fertilization to swim up stage and reproductive success in a farmed rainbow trout line. *Aquaculture Reports*.

Pennings, P. S., & Hermisson, J. (2006). Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. *Molecular biology and evolution*, 23(5), 1076-1084. <https://doi.org/10.1093/molbev/msj117>

Peripolli, E., Stafuzza, N. B., Munari, D. P., Lima, A. L. F., Irgang, R., Machado, M. A., ... & da Silva, M. V. G. B. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC genomics*, 19(1), 1-13. <https://doi.org/10.1186/s12864-017-4365-3>

Pulcini, D., Wheeler, P. A., Cataudella, S., Russo, T., & Thorgaard, G. H. (2013). Domestication shapes morphology in rainbow trout *Oncorhynchus mykiss*. *Journal of Fish Biology*, 82(2), 390-407. <https://doi.org/10.1111/jfb.12002>

Prchal, M., D'Ambrosio, J., Lagarde, H., Lallias, D., Patrice, P., François, Y., ... & Phocas, F. (2022). Genome-wide association study and genomic prediction of tolerance to acute hypoxia in rainbow trout. *Aquaculture*, 739068. <https://doi.org/10.1016/j.aquaculture.2022.739068>

Price, E. O. (1984). Behavioral aspects of animal domestication. *The quarterly review of biology*, 59(1), 1-32.

Price, E. O. (1999). Behavioral development in animals undergoing domestication. *Applied Animal Behaviour Science*, 65(3), 245-271. [https://doi.org/10.1016/S0168-1591\(99\)00087-8](https://doi.org/10.1016/S0168-1591(99)00087-8)

Price, E. O. (2002). *Animal domestication and behavior*. Cabi. <http://sherekashmir.informaticspublishing.com/479/1/9780851995977.pdf>

Pritchard, V. L., Mäkinen, H., Vähä, J. P., Erkinaro, J., Orell, P., & Primmer, C. R. (2018). Genomic signatures of fine-scale local selection in Atlantic salmon suggest involvement of sexual maturation, energy homeostasis and immune defence-related genes. *Molecular Ecology*, 27(11), 2560-2575. <https://doi.org/10.1111/mec.14705>

Purcell, R. (2017). Activation and regulation of the brain-expressed adhesion G protein-coupled receptors *ADGRB1/BA11* and *ADGRB2/BA12*: Implications for human disease (Doctoral dissertation, Emory University). <https://etd.library.emory.edu/concern/etds/mc87pq78v?locale=fr>

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC genetics*, 13(1), 1-11. <https://doi.org/10.1186/1471-2156-13-70>

Purfield, D. C., McParland, S., Wall, E., & Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One*, 12(5), e0176780. <https://doi.org/10.1371/journal.pone.0176780>

Qanbari, S. & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livest. Sci.* 166, 133–143. <https://doi.org/10.1016/j.livsci.2014.05.003>

Raghuram, V., Weber, S., Raber, J., Chen, D. H., Bird, T. D., Maylie, J., & Adelman, J. P. (2017). Assessment of mutations in KCNN2 and ZNF135 to patient neurological symptoms. *Neuroreport*, *28*(7), 375. <https://doi.org/10.1097/WNR.0000000000000754>

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research*, *47*(W1), W191-W198. <https://doi.org/10.1093/nar/gkz369>

Repo, H., Gurvits, N., Löyttyniemi, E., Nykänen, M., Lintunen, M., Karra, H., ... & Kronqvist, P. (2017). PTTG1-interacting protein (PTTG1IP/PBF) predicts breast cancer survival. *BMC cancer*, *17*(1), 1-8. <https://doi.org/10.1186/s12885-017-3694-6>

Rhee, K., Brunori, M., Besset, V., Trousdale, R., & Wolgemuth, D. J. (1998). Expression and potential role of Fsr_g1, a murine bromodomain-containing homologue of the Drosophila gene female sterile homeotic. *Journal of Cell Science*, *111*(23), 3541-3550. <https://doi.org/10.1242/jcs.111.23.3541>

Rodríguez-Marí, A., Wilson, C., Titus, T. A., Canestro, C., BreMiller, R. A., Yan, Y. L., ... & Postlethwait, J. H. (2011). Roles of brca2 (fancd1) in oocyte nuclear architecture, gametogenesis, gonad tumors, and genome stability in zebrafish. *PLoS genetics*, *7*(3), e1001357. <https://doi.org/10.1371/journal.pgen.1001357>

Russell, N. (2002). The wild side of animal domestication. *Society & Animals*, *10*(3), 285-302. <https://doi.org/10.1163/156853002320770083>

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., ... & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832-837. <https://doi.org/10.1038/nature01140>

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913-918. <https://doi.org/10.1038/nature06250>

Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science*, *241*, 104257. <https://doi.org/10.1016/j.livsci.2020.104257>

Saravanan, K. A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G. K., ... & Singh, R. K. (2021). Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, *113*(3), 955-963. <https://doi.org/10.1016/j.ygeno.2021.02.009>

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC genomics*, *15*(1), 1-12. <https://doi.org/10.1186/1471-2164-15-478>

Selleri L, Depew MJ, Jacobs Y, Chanda SK, Tsang KY, Cheah KSE, Rubenstein JLR, O'Gorman S, Cleary ML. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* 128: 3543– 3557. <https://doi.org/10.1242/dev.128.18.3543>

Seo, J., Asaoka, Y., Nagai, Y., Hirayama, J., Yamasaki, T., Namae, M., ... & Nishina, H. (2010). Negative regulation of wnt11 expression by Jnk signaling during zebrafish gastrulation. *Journal of cellular biochemistry*, 110(4), 1022-1037. <https://doi.org/10.1002/jcb.22616>

Shiu, F. H., Wong, J. C., Yamamoto, T., Lala, T., Purcell, R. H., Owino, S., ... & Escayg, A. (2022). Mice lacking full length Adgrb1 (Bai1) exhibit social deficits, increased seizure susceptibility, and altered brain development. *Experimental Neurology*, 351, 113994. <https://doi.org/10.1016/j.expneurol.2022.113994>

Shive, H. R., West, R. R., Embree, L. J., Azuma, M., Sood, R., Liu, P., & Hickstein, D. D. (2010). brca2 in zebrafish ovarian development, spermatogenesis, and tumorigenesis. *Proceedings of the National Academy of Sciences*, 107(45), 19350-19355. <https://doi.org/10.1073/pnas.1011630107>

Signer-Hasler, H., Henkel, J., Bangerter, E., Bulut, Z., Drögemüller, C., Leeb, T., & Flury, C. (2022). Runs of homozygosity in Swiss goats reveal genetic changes associated with domestication and modern selection. *Genetics Selection Evolution*, 54(1), 1-11. <https://doi.org/10.1186/s12711-022-00695-w>

Silverstein, J. T., Rexroad III, C. E., & King, T. L. (2004). Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). *Aquaculture Research*, 35(1), 40-48. <https://doi.org/10.1111/j.1365-2109.2004.00979.x>

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35. <https://doi.org/10.1017/S0016672300014634>

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *The American Journal of Human Genetics*, 93(1), 90-102. <https://doi.org/10.1016/j.ajhg.2013.05.003>

Stanković, D., Crivelli, A. J., & Snoj, A. (2015). Rainbow trout in Europe: introduction, naturalization, and impacts. *Reviews in Fisheries Science & Aquaculture*, 23(1), 39-71. <https://doi.org/10.1080/23308249.2015.1024825>

Su, S., Li, H., Du, F., Zhang, C., Li, X., Jing, X., ... & Bouzoualegh, R. (2018). Combined QTL and genome scan analyses with the help of 2b-RAD identify growth-associated genetic markers in a new fast-growing carp strain. *Frontiers in genetics*, 9, 592. <https://doi.org/10.3389/fgene.2018.00592>

Sullivan, C., Lage, C. R., Yoder, J. A., Postlethwait, J. H., & Kim, C. H. (2017). Evolutionary divergence of the vertebrate TNFAIP8 gene family: Applying the spotted gar orthology bridge to understand ohnolog loss in teleosts. *Plos one*, 12(6), e0179517. <https://doi.org/10.1371/journal.pone.0179517>

Sun, L., Liu, S., Wang, R., Jiang, Y., Zhang, Y., Zhang, J., ... & Liu, Z. (2014). Identification and analysis

of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). *PLoS one*, 9(10), e109666. <https://doi.org/10.1371/journal.pone.0109666>

Teoh, P. H., Shu-Chien, A. C., & Chan, W. K. (2010). Pbx1 is essential for growth of zebrafish swim bladder. *Developmental dynamics: an official publication of the American Association of Anatomists*, 239(3), 865-874. <https://doi.org/10.1002/dvdy.22221>

Těšický, M., & Vinkler, M. (2015). Trans-species polymorphism in immune genes: general pattern or MHC-restricted phenomenon? *Journal of immunology research*, 2015. <https://doi.org/10.1155/2015/838035>

Torky, H. A., Sherif, A., Abo-Louz, A., Ali, M., Ahmed, A., & Ali, A. (2018). Evaluation of serum nidogen-2 as a screening and diagnostic tool for ovarian cancer. *Gynecologic and Obstetric Investigation*, 83(5), 461-465. <https://doi.org/10.1159/000481798>

Tymchuk, W. E., Beckman, B., & Devlin, R. H. (2009). Altered expression of growth hormone/insulin-like growth factor I axis hormones in domesticated fish. *Endocrinology*, 150(4), 1809-1816. <https://doi.org/10.1210/en.2008-0797>

Umasuthan, N., Revathy, K. S., Whang, I., Kim, E., Oh, M. J., Jung, S. J., ... & Lee, J. (2014). Genomic identification and molecular characterization of a non-mammalian TNFAIP8L2 gene from *Oplegnathus fasciatus*. *Gene*, 542(1), 52-63. <https://doi.org/10.1016/j.gene.2014.02.047>

Uvarov, P., Kajander, T., & Airaksinen, M. S. (2014). Origin and loss of nested LRR/α-catenin genes during vertebrate evolution. *PLoS one*, 9(2), e89910. <https://doi.org/10.1371/journal.pone.0089910>

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual review of genetics*. <https://doi.org/10.1146/annurev-genet-111212-133526>

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS biology*, 4(3), e72. <https://doi.org/10.1371/journal.pbio.0040072>

Walters, K. B., Dodd, M. E., Mathias, J. R., Gallagher, A. J., Bennin, D. A., Rhodes, J., ... & Huttenlocher, A. (2009). Muscle degeneration and leukocyte infiltration caused by mutation of zebrafish Fad24. *Developmental dynamics: an official publication of the American Association of Anatomists*, 238(1), 86-99. <https://doi.org/10.1002/dvdy.21821>

Wang, W. J., Tay, H. G., Soni, R., Perumal, G. S., Goll, M. G., Macaluso, F. P., ... & Bryan Tsou, M. F. (2013). CEP162 is an axoneme-recognition protein promoting ciliary transition zone assembly at the cilia base. *Nature cell biology*, 15(6), 591-601. <https://doi.org/10.1038/ncb2739>

Wang, G. D., Xie, H. B., Peng, M. S., Irwin, D., & Zhang, Y. P. (2014). Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.*, 2(1), 65-84. <https://doi.org/10.1146/annurev-animal-022513-114129>

Wang, Y., Miao, X., Li, H., Su, P., Lin, L., Liu, L., & Li, X. (2020). The correlated expression of immune and energy metabolism related genes in the response to *Salmonella enterica* serovar Enteritidis inoculation in chicken. *BMC veterinary research*, *16*(1), 1-9. <https://doi.org/10.1186/s12917-020-02474-5>

Wang, H., Teng, M., Liu, P., Zhao, M., Wang, S., Hu, J., ... & Zeng, Q. (2022). Selection Signatures of Pacific White Shrimp *Litopenaeus vannamei* Revealed by Whole-Genome Resequencing Analysis. *Advances in the Biology, Aquaculture, and Conservation of Threatened Marine Species and their Application in Human Health and Nutrition*. <https://doi.org/10.3389/fmars.2022.844597>

Wassarman, P. M., & Litscher, E. S. (2018). The mouse egg's zona pellucida. *Current topics in developmental biology*, *130*, 331-356. <https://doi.org/10.1016/bs.ctdb.2018.01.003>

Wassarman, P. M., & Litscher, E. S. (2021). Zona pellucida genes and proteins: Essential players in mammalian oogenesis and fertility. *Genes*, *12*(8), 1266. <https://doi.org/10.3390/genes12081266>

Wayne, R. K., & Vonholdt, B. M. (2012). Evolutionary genomics of dog domestication. *Mammalian Genome*, *23*(1), 3-18. <https://doi.org/10.1007/s00335-011-9386-7>

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Weinstein, S. Y., Thrower, F. P., Nichols, K. M., & Hale, M. C. (2019). A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. *PloS one*, *14*(9), e0223018. <https://doi.org/10.1371/journal.pone.0223018>

Wellenreuther, M., Le Luyer, J., Cook, D., Ritchie, P. A., & Bernatchez, L. (2019). Domestication and temperature modulate gene expression signatures and growth in the Australasian snapper *Chrysophrys auratus*. *G3: Genes, Genomes, Genetics*, *9*(1), 105-116. <https://doi.org/10.1534/g3.118.200647>

Wen, Z. Y., Liu, T., Qin, C. J., Zou, Y. C., Wang, J., Li, R., & Tao, Y. X. (2021). MRAP2 interaction with melanocortin-4 receptor in snakehead (*Channa argus*). *Biomolecules*, *11*(3), 481. <https://doi.org/10.3390/biom11030481>

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, *197*(3), 795-808. <https://doi.org/10.1534/genetics.114.171975>

Williams, J. L., Hall, S. J., Del Corvo, M., Ballingall, K. T., Colli, L. I. C. I. A., Ajmone Marsan, P. A. O. L. O., & Biscarini, F. (2016). Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Animal genetics*, *47*(1), 19-27. <https://doi.org/10.1111/age.12376>

Wright, D., Nakamichi, R., Krause, J., & Butlin, R. K. (2006). QTL analysis of behavioral and morphological differentiation between wild and laboratory zebrafish (*Danio rerio*). *Behavior genetics*, *36*(2), 271-

284.

Wright, D. (2015). Article commentary: the genetic architecture of domestication in animals. *Bioinformatics and biology insights*, 9, BBI-S28902. <https://doi.org/10.4137/BBI.S28902>

Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, 16(2), 97.

Xu, H., Dong, R., Zeng, Q., Fang, L., Ge, Q., Xia, C., ... & Tong, P. (2022). Col9a2 gene deletion accelerates the degeneration of intervertebral discs. *Experimental and Therapeutic Medicine*, 23(3), 1-13. <https://doi.org/10.3892/etm.2022.11130>

Yan, Q., Huang, C., Jiang, Y., Shan, H., Jiang, R., Wang, J., ... & Sun, H. (2018). Calpain7 impairs embryo implantation by downregulating β 3-integrin expression via degradation of HOXA10. *Cell death & disease*, 9(3), 1-12.

Yáñez, J. M., Xu, P., Carvalheiro, R., & Hayes, B. (2022). Genomics applied to livestock and aquaculture breeding. *Evolutionary Applications*, 15(4), 517. <https://doi.org/10.1111/eva.13378>

Yao, X., Liu, X., Zhang, Y., Li, Y., Zhao, C., Yao, S., & Wei, Y. (2017). Gene therapy of adult neuronal ceroid lipofuscinoses with CRISPR/Cas9 in zebrafish. *Human gene therapy*, 28(7), 588-597.

Ye, X., Zhou, X. J., & Zhang, H. (2018). Exploring the role of autophagy-related gene 5 (ATG5) yields important insights into autophagy in autoimmune/autoinflammatory diseases. *Frontiers in immunology*, 9, 2334. <https://doi.org/10.3389/fimmu.2018.02334>

Yin, X., Huo, Z., Yan, S., Wang, Z., Yang, T., Wu, H., & Zhang, Z. (2021). MiR-205 inhibits sporadic vestibular schwannoma cell proliferation by targeting cyclin-dependent kinase 14. *World Neurosurgery*, 147, e25-e31. <https://doi.org/10.1016/j.wneu.2020.11.043>

Yoshida, G. M., Lhorente, J. P., Carvalheiro, R., & Yáñez, J. M. (2017). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar* L.). *Animal Genetics*, 48(6), 698-703. <https://doi.org/10.1111/age.12621>

Yu, X., Setyawan, P., Bastiaansen, J. W., Liu, L., Imron, I., Groenen, M. A., ... & Megens, H. J. (2022). Genomic analysis of a Nile tilapia strain selected for salinity tolerance shows signatures of selection and hybridization with blue tilapia (*Oreochromis aureus*). *Aquaculture*, 560, 738527. <https://doi.org/10.1016/j.aquaculture.2022.738527>

Żarski, D., Le Cam, A., Nynca, J., Klopp, C., Ciesielski, S., Sarosiek, B., ... & Bobe, J. (2020). Domestication modulates the expression of genes involved in neurogenesis in high-quality eggs of Sander lucioperca. *Molecular Reproduction and Development*, 87(9), 934-951. <https://doi.org/10.1002/mrd.23414>

Zhang, Z., Zhang, Q., Xiao, Q., Sun, H., Gao, H., Yang, Y., ... & Pan, Y. (2018). Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data. *Animal*

genetics, 49(6), 579-591. <https://doi.org/10.1111/age.12730>

Zhang, S. H., Liu, D. X., Wang, L., Li, Y. H., Wang, Y. H., Zhang, H., ... & Chen, Y. H. (2019). A CASPR1–ATP1B3 protein interaction modulates plasma membrane localization of Na⁺/K⁺-ATPase in brain microvascular endothelial cells. *Journal of Biological Chemistry*, 294(16), 6375-6386. <https://doi.org/10.1074/jbc.RA118.006263>

Zhang, J., Zheng, T., Zhou, X., Wang, H., Li, Z., Huan, C., ... & Zhang, W. (2021). ATP1B3 Restricts Hepatitis B Virus Replication Via Reducing the Expression of the Envelope Proteins. *Virologica Sinica*, 36(4), 678-691. <https://doi.org/10.1007/s12250-021-00346-2>

Zhang, J. L., Richetti, S., Ramezani, T., Welcker, D., Lütke, S., Pogoda, H. M., ... & Hammerschmidt, M. (2022). Vertebrate extracellular matrix protein hemicentin-1 interacts physically and genetically with basement membrane protein nidogen-2. *Matrix Biology*. <https://doi.org/10.1016/j.matbio.2022.08.009>

Zheng, B., Zhang, J., Zheng, T., Wang, H., Li, Z., Huan, C., ... & Zhang, W. (2020). ATP1B3 cooperates with BST-2 to promote hepatitis B virus restriction. *Journal of medical virology*, 92(2), 201-209. <https://doi.org/10.1002/jmv.25599>

Zheng, T., Song, Z., Tao, Y., Qiang, J., Ma, J., Lu, S., & Xu, P. (2022). Transport stress induces innate immunity responses through TLR and NLR signaling pathways and increases mucus cell number in gills of hybrid yellow catfish (*Tachysurus fulvidraco*♀ × *Pseudobagrus vachellii*♂). *Fish & Shellfish Immunology*, 127, 166-175. <https://doi.org/10.1016/j.fsi.2022.06.015>

Zhou, W., & Hildebrandt, F. (2009). Molecular cloning and expression of phospholipase C epsilon 1 in zebrafish. *Gene Expression Patterns*, 9(5), 282-288. <https://doi.org/10.1016/j.gep.2009.03.003>

Zhu, M., Xu, B., Wang, M., Liu, S., Zhang, Y., & Zhang, C. (2019). Pharmacological modulation of MRAP2 protein on melanocortin receptors in the sea lamprey. *Endocrine connections*, 8(4), 378-388. <https://doi.org/10.1530/EC-19-0019>

Zhuang, Y., Ly, R. C., Frazier, C. V., Yu, J., Qin, S., Fan, X. Y., ... & Wang, L. (2019). The novel function of tumor protein D54 in regulating pyruvate dehydrogenase and metformin cytotoxicity in breast cancer. *Cancer & metabolism*, 7(1), 1-15. <https://doi.org/10.1186/s40170-018-0193-4>

Data Accessibility and Benefit-Sharing

Restrictions applied to the availability of the data that support the findings of this study, which were used under license and so are not publicly available. The data can be made available for reproduction of the results from Florence Phocas (florence.phocas@inrae.fr) on request via a material transfer agreement and with permission of the two breeding companies “Viviers de Sarrance” (Sarrance, France) and “Milin Nevez” (Plouigneau, France).

Author contributions

Katy Paul: Investigations, Methodology, Formal analysis, Writing - Original Draft;

Gwendal Restoux: Conceptualization, Methodology, Draft Reviewing;

Florence Phocas: Supervision, Conceptualization, Methodology, Investigation, Formal analysis, Resources, Writing - Original Draft.

List of appendices

- **Supplementary figure 1.** Principal component analysis (PCA) of the genetic diversity of SY, and HA sub-populations based on 546,903 SNPs.
- **Supplementary information S1.** Commons signatures of selection between French lines (LB, LC and SY) and American pooled population (HA) detected by XP-EHH statistic.
- **Supplementary Informations S2.** Listing of ROH islands in LB population.
- **Supplementary Informations S3.** Listing of ROH islands in LC population.
- **Supplementary Informations S4.** Listing of ROH islands in SY population.
- **Supplementary Informations S5.** Listing of ROH islands in HA population.
- **Supplementary Informations S6.** Listing of significant signals of selection identified by iHS in LB population.
- **Supplementary Informations S7.** Listing of significant signals of selection identified by iHS in LC population.
- **Supplementary Informations S8.** Listing of significant signals of selection identified by iHS in SY population.
- **Supplementary Informations S9.** Listing of significant signals of selection identified by iHS in HA population.

- **Supplementary Information S10.** iHS and ROH values per population within the 9 overlapping regions detecting in the four populations.
- **Supplementary Information S11.** Listing of the 253 genes which composed the nine shared genomic regions. In bold, genes identified linked to domestication in the literature.
- **Supplementary Informations S12.** Listing of heterozygous regions (i.e., large region without ROH) in LB population.
- **Supplementary Informations S13.** Listing of heterozygous regions (i.e., large region without ROH) in LC population.
- **Supplementary Informations S14.** Listing of heterozygous regions (i.e., large region without ROH) in SY population.
- **Supplementary Informations S15.** Listing of heterozygous regions (i.e., large region without ROH) in HA population.
- **Supplementary Information S16.** Listing of the 29 genes which composed the four heterozygous genomic regions.
- **Supplementary information S17.** Percentage of protein homology between rainbow trout and nine other vertebrate species for all genes annotated in heterozygous regions chr10_a, chr13_a, chr19_a and chr19_b.

Chapitre 5

Discussion Générale

Dans un contexte de changement climatique et de diversification des objectifs de sélection, il est essentiel de gérer et de préserver la diversité génétique des lignées sélectionnées de truites arc-en-ciel, dont la baisse observée ces dernières décennies devient inquiétante (D'Ambrosio et al., 2019). Dans ce chapitre, à la lumière des connaissances sur la diversité génétique acquises au cours de cette thèse, nous proposerons des pistes d'amélioration de sa gestion dans les programmes de sélection truiticole. La discussion s'organisera donc selon les étapes d'un programme d'amélioration génétique, en abordant les enjeux de sa gestion, ainsi que les méthodes et les outils disponibles pour gérer ces ressources au sein des populations sélectionnées de truite arc-en-ciel.

Tout programme de sélection se construit en quatre étapes clefs (Phocas, 2011). Premièrement, il faut déterminer des objectifs et des critères de sélection, puis réaliser l'évaluation génétique des individus, permettant un choix objectivé des reproducteurs. Ensuite, il est nécessaire d'organiser et d'optimiser les différentes actions de création du progrès génétique, c'est-à-dire sélectionner les reproducteurs et réaliser les plans d'accouplement. Pour finir, la dernière étape d'un programme de sélection est la diffusion du progrès génétique acquis dans les élevages de production.

Les schémas de sélection aquacole actuels sont divisés en deux catégories : des schémas de sélection massale basés sur les seuls phénotypes des candidats à la sélection, et des schémas de sélection généalogique. La sélection généalogique en aquaculture a été initiée dans les

années 70 en Norvège chez le saumon (Gjedrem, 2010), où l'élevage est fait en familles séparées. Avant l'essor de la génomique, cette stratégie avait de nombreux avantages : améliorer la précision de l'évaluation des caractères grâce à la possibilité d'utiliser l'information familiale ; sélectionner sur des caractères létaux mesurés sur des collatéraux apparentés ; et gérer la diversité génétique (Gjedrem, 2010). Cependant, cette stratégie présente deux inconvénients majeurs. D'une part, elle nécessite d'investir dans des structures d'élevage induisant un surcoût important, limitant le nombre de familles produites (Vandeputte & Haffray, 2014). D'autre part, dans ce système, il n'est pas possible de dissocier les effets génétiques additifs des effets environnementaux et maternels, ce qui induit des pertes de précisions et une surestimation de la composante génétique dans l'évaluation des reproducteurs.

Aujourd'hui, la sélection dans les élevages truiticoles français se fait à l'aide d'une combinaison de sélection massale sur la croissance et la morphologie selon la méthode PROSPER (Chevassus et al., 2004), et d'une sélection sur une indexation généalogique BLUP. Depuis 2004, l'assignation de parenté assistée par marqueurs génétiques est utilisée en routine pour permettre l'indexation BLUP – modèle Animal (Vandeputte et Haffray 2014). Depuis 2018, l'essor et la baisse des coûts du génotypage ont également permis aux sélectionneurs de commencer la mise en place de la sélection génomique GBLUP (D'Ambrosio, 2020). En outre, une gestion des accouplements est réalisée pour éviter de croiser des reproducteurs trop apparentés grâce à un algorithme d'optimisation de leurs contributions afin de minimiser l'augmentation de consanguinité sous contrainte d'un progrès génétique fixé (Colleau et al., 2009; Chapuis et al., 2016).

Au sein de chaque entreprise de sélection truiticole, il existe au moins deux populations de truites répondant à des finalités différentes : la population de sélection et la population de multiplication. La population (ou noyau) de sélection est conservée au sein de chaque entreprise, et permet d'alimenter la population de multiplication pour fournir suffisamment d'œufs embryonnés améliorés à destination des grossisseurs (Chavanne et al., 2016). Les stratégies à mettre en place au sein des programmes de sélection, pour gérer la diversité génétique et ses effets, pourront donc être variables entre le noyau de sélection et la population multiplicatrice.

5.1. Les objectifs des programmes d'amélioration génétique

5.1.1. Choix de la stratégie d'amélioration génétique

La diversité des programmes d'amélioration génétique des animaux d'élevage dépend non seulement des contraintes biologiques des espèces considérées (en particulier leur niveau de fécondité), mais aussi des types de production (importance des effets d'hétérosis sur les caractères ayant le plus d'importance économique) et des objectifs de sélection (compromis entre caractères de production, de reproduction et santé des animaux). Deux stratégies différentes sont distinguées : le croisement et la sélection. Il y a sur certains points un recouvrement, comme, par exemple, le croisement d'absorption qui s'apparente à un processus de sélection pour introduire de nouvelles aptitudes ou de la variabilité dans une population. L'élevage en race/lignée pure consiste à accoupler des reproducteurs issus de la même population. Tandis que le croisement consiste à accoupler des reproducteurs issus de races/lignées différentes. L'objectif du croisement est de tirer parti de la variabilité génétique entre populations (races ou lignées) disponibles à un moment donné. Dans cette discussion, le croisement est considéré comme une stratégie d'utilisation des reproducteurs (cf 5.4.3. Valorisation de l'hétérosis). Un programme d'amélioration génétique en croisement implique des coûts supplémentaires conséquents, car il nécessite de conduire au moins deux programmes de sélection en race pure sur de races/lignées distinctes. Mais les avantages principaux du croisement sont : 1) de bénéficier de l'effet de complémentarité entre des types génétiques spécialisés pour des fonctions pouvant être antagonistes, comme une meilleure aptitude de reproduction sur la voie maternelle couplée à une meilleure production sélectionnée sur la voie paternelle (Sellier, 1976 ; Roche et al., 2010) ; 2) d'exploiter le phénomène d'hétérosis (ou de vigueur de l'hybride) chez les individus hybrides, se traduisant par des performances moyennes de ces animaux supérieures à la moyenne des performances de leurs parents.

Le croisement constitue également une potentielle protection des améliorations génétiques réalisées par les sélectionneurs, et leur confère donc un avantage compétitif.

5.1.2. La définition de l'objectif de sélection

La définition des objectifs de sélection est l'étape préalable et indispensable à la mise en place de tout programme d'amélioration génétique. Cela permet de connaître et de mettre en relation l'ensemble des aptitudes biologiques hérissables que l'on souhaite améliorer (Nielsen et al., 2014). Cette fonction permet ensuite d'évaluer et donc de classer objectivement les

individus pour y choisir les reproducteurs répondant au mieux à l'objectif de sélection décidé en amont (Phocas et al., 2020). Les objectifs de sélection dans les programmes truiticoles sont majoritairement liés aux caractères de croissance et de production (rendement de découpe et qualité de chair). Cependant, aujourd'hui chez la TAC, de nouveaux caractères, comme des caractères de résistance à des maladies ou de reproduction des femelles, sont intégrés dans les objectifs de sélection (Fraslin, 2018 ; D'Ambrosio, 2020).

5.1.3. Limiter la dépression de consanguinité et l'érosion de la diversité génétique

Au-delà du choix des caractères à inclure dans l'objectif de sélection, la finalité du programme doit être précisée (maximisation de la réponse à la sélection à court versus long terme) car l'importance à accorder au maintien de la variabilité génétique dans l'optimisation du programme de sélection en dépend. Le maintien de la variabilité génétique au-delà des 4-5 premières générations de sélection devient un problème majeur dès lors que l'on s'intéresse à des populations d'effectifs génétiques limités ($N_e < 100$), en raison des problèmes de dérive génétique et de consanguinité. Etant donné les faibles effectifs efficaces calculés dans les populations de truites arc-en-ciel (D'Ambrosio et al., 2019), l'intégration du maintien de la diversité génétique est un enjeu majeur pour les sélectionneurs souhaitant poursuivre leurs programmes d'amélioration génétique en lignées fermées au-delà d'une dizaine d'années.

Deux méthodes alternatives basées sur l'optimisation des contributions des reproducteurs aux générations suivantes ont été proposées pour répondre à cet enjeu. La première développée par Meuwissen (1997) consiste à maximiser le progrès génétique à contrainte fixée pour limiter l'accroissement de la consanguinité à un niveau prédéfini (méthode des contributions optimales (OC)). La seconde méthode mise en place par le SYSAAF chez certains sélectionneurs, vise à minimiser l'accroissement de consanguinité à contrainte fixée sur le progrès génétique souhaité (Colleau et al., 2006). Ces deux méthodes diffèrent dans l'enjeu prioritaire défini pour le programme de sélection. Meuwissen (1997) propose avant tout de limiter l'augmentation de la consanguinité alors que Colleau et al (2006) mettent en priorité l'atteinte d'un gain génétique suffisant. La recommandation faite par la FAO est de limiter l'accroissement de la consanguinité de 0,5 à 1% d'une génération à l'autre (FAO, 2009; Rischkowsky & Pilling, 2007). Cette recommandation est respectée dans les programmes de sélection truiticole du SYSAAF ayant incluent une gestion de la consanguinité (D'Ambrosio, 2020).

D'autres approches sont également proposées pour gérer le choix des reproducteurs et leurs accouplements, telles que l'utilisation d'index d'originalité chez les bovins (Hoze et al., 2018) ou la gestion des anomalies génétiques (Danchin-Burge et al., 2016) pour limiter les effets délétères de la consanguinité dans les populations sous sélection intensive.

Dans mes travaux de thèse, d'importants effets de dépression de consanguinité ont été estimés sur les caractères de reproduction de la truite arc-en-ciel (cf Chapitre 3), ainsi que sur des caractères de production (cf Résultats annexe 3) et de survie (cf Chapitre 2 et Résultats annexe 4). Ces résultats sont cohérents avec ceux de deux méta-analyses répertoriant les différentes estimations de dépression de consanguinité dans plusieurs espèces et sur différents types de caractères (Leroy, 2014 ; Doekes et al., 2021). Ces deux méta-analyses montrent que les effets de dépression de consanguinité peuvent être conséquents pour des caractères de production, et ne sont donc pas uniquement importants à considérer pour les caractères liés à la *fitness* (survie, reproduction), malgré le consensus largement admis dans la littérature, mais vraisemblablement fortement issu des observations en populations sauvages (Lynch & Walsh, 1998 ; DeRose & Roff, 1999). Il est donc essentiel de maintenir une diversité génétique suffisante au sein des populations de truites, pour limiter les phénomènes de dépression de consanguinité sur l'ensemble des caractères d'intérêt en aquaculture.

Au-delà des effets de la dépression de consanguinité, la perte de diversité génétique engendre des risques pour l'adaptation des animaux à des milieux hétérogènes et changeants. Dans le contexte actuel de changement climatique, il est essentiel pour les populations sélectionnées d'avoir de bonnes performances, pas seulement dans un milieu optimal, mais dans différents environnements (Pilling & Hoffmann, 2011). De plus, les entreprises de sélection truiticole françaises exportent une bonne partie de leur production d'œufs à l'international (dans plus de 20 pays d'Europe et du Moyen-Orient) (D'Ambrosio, 2020), où les conditions d'élevage peuvent différer fortement des conditions françaises. Avoir des animaux aux performances constantes dans divers milieux est donc également un enjeu majeur pour l'économie des entreprises de sélection. La résilience des animaux et de leurs performances à des variations environnementales est donc un objectif poursuivi.

Aujourd'hui, il est nécessaire d'avoir une approche plus raisonnée pour poursuivre l'amélioration génétique des lignées commerciales de truites arc-en-ciel, tout en maintenant un réservoir suffisant d'allèles dans ces lignées sélectionnées en populations fermées. A la lumière

des travaux de thèse décrits dans le Chapitre 3, la gestion fine de la consanguinité, au niveau des fragments chromosomiques, nous apparaît comme un enjeu majeur pour ces programmes de sélection, non seulement pour limiter les effets de dépression de consanguinité et donc améliorer les performances des cheptels, mais aussi pour garantir la possibilité pour les lignées de s'adapter aux changements environnementaux, à divers stress abiotiques ou biotiques, ainsi qu'à des objectifs de sélection futurs.

5.2. L'évaluation génétique des reproducteurs

L'évaluation génétique des candidats à la sélection est une étape nécessaire au choix objectif des reproducteurs. Plus l'évaluation des candidats sera précise, plus la sélection sera efficace. Cette évaluation se fait sur la base d'un index synthétique constitué d'une combinaison linéaire de plusieurs index élémentaires établie pour maximiser la réponse sur l'objectif de sélection pré-déterminée (Hazel, 1943).

La méthode et le modèle d'évaluation génétique de référence est depuis trente ans le BLUP-modèle animal (Henderson, 1975) basé sur la prise en compte des relations génétiques entre animaux via une matrice de parenté établie sur pedigree. Classiquement, seuls les effets génétiques additifs sont pris en compte dans le modèle, car les effets non-additifs (dominance, épistasie, interactions génotypes/environnement) sont plus difficiles à estimer précisément sur la base de données pedigree.

L'utilisation de l'information génomique pour construire la matrice de parenté (GBLUP, Habier et al., 2007) permet à présent de considérer efficacement les effets non-additifs dans les modèles d'évaluation génétique. Au sein de ces modèles génomiques, différentes valeurs génotypiques (GV) et génétiques (BV) peuvent être prédites. La valeur BV correspond à l'estimation des effets génétiques additifs, classiquement utilisée en sélection. La valeur GV peut se décomposer en trois composantes : additive, dominance et épistasie (Minvielle, 1992).

5.2.1. Prise en compte des effets de dominance

Jusqu'à maintenant les effets de dominance sont très peu pris en compte dans les évaluations génétiques pour diverses raisons. On s'intéresse en priorité aux effets additifs car ce sont eux qui sont directement héréditaires et transmissibles aux générations futures. De plus, l'estimation des effets de dominance est difficile (manque d'informations sur le pedigree ; peu de grandes familles ; calculs plus complexes, etc.) et n'est pas utile dans tous les programmes d'amélioration génétique. Toutefois, inclure les effets de dominance dans les modèles d'évaluation est conseillé, surtout avec l'utilisation des données génomiques qui facilite leur estimation. Dans tous les cas, ajouter l'effet de la dominance dans le modèle génétique ne détériore pas l'évaluation des reproducteurs, et cela même si le caractère étudié est purement additif (Zeng et al., 2013 ; Duenk et al., 2017). L'évaluation des candidats en considérant les effets de dominance semble indispensable dans les programmes d'amélioration génétique en croisement, notamment pour identifier les meilleurs accouplements à réaliser pour valoriser l'hétérosis (Toro & Varona, 2010 ; Su et al., 2012 ; Sun et al., 2013 ; Vitezica et al., 2013, 2016 ; Xiang et al., 2016 ; Varona et al., 2018a, 2018b ; Fernández et al., 2021). Dans les programmes de sélection en lignée pure, l'intérêt de la prise en compte des effets de dominance est moins évident.

Dans l'étude présentée au Chapitre 3, les valeurs génétiques des individus estimées dans les modèles GBLUP avec ou sans effets de dominance sont corrélées à plus de 99% (Tableau 5.1). Donc, inclure des effets de dominance ne semble pas améliorer la précision de l'estimation des valeurs génétiques additives des reproducteurs. Nos résultats concordent avec ceux obtenus dans plusieurs espèces terrestres, où aucune amélioration des prédictions des valeurs génétiques n'a été observée en considérant les effets de dominance dans les évaluations (poules pondeuses : Heidaritabar et al., 2016 ; mouton : Moghaddar & Van der Werf, 2017 ; bovin : Garcia-Baccino et al., 2020).

Performances	corrélation des EBV
Poids après ponte (PW)	0,999
Taille à la fourche (FL)	0,999
Poids de ponte (SW)	0,997
Nombre d'œufs (EN)	0,998
Poids moyen d'un œuf (EW)	0,998
Semaine de ponte (SD)	0,999
Poids du liquide cœlomique (CL)	0,999

Tableau 5.1. Corrélations entre les valeurs génétiques (BV) estimées par les modèles du chapitre 3 (avec (D_GBLUP) ou sans dominance (GBLUP)) pour 2 performances liées à la taille et 5 caractères de reproduction de truites arc-en-ciel femelles.

Quelques études basées sur des données de simulation ont montré que l'ajout des effets de dominance dans les modèles permet d'augmenter un peu la précision de l'estimation des valeurs génétiques additives (BV) (Toro & Varona, 2010 ; Varona et al., 2018a), mais augmentait surtout la précision de l'estimation des valeurs génotypiques (GV) des individus (Wellmann & Benewitz, 2012). Une étude sur données réelles chez le porc, considérant deux lignées pures (Yorkshire et Landrace) et leur croisement, montre une amélioration de la précision des valeurs génomiques prédites (GEBV) entre 5 et 18% sur la taille de portée en considérant les effets de dominance dans chacune des deux races (Esfandyari et al., 2016).

Par ailleurs, Xiang et al. (2018) ont montré qu'il existait une corrélation entre les effets additifs et les effets de dominance. Cependant la prise en compte de la corrélation entre les effets additifs et de dominance, n'améliore que très peu la précision valeurs génotypiques (GV) qui sont finalement proches de celles estimées par l'approche orthogonale ne prenant pas en compte la corrélation entre les deux effets génétiques (Vitezica et al., 2016). Cette approche orthogonale développée par Vitezica et al. (2016) nous a servi à réaliser les études décrites dans le Chapitres 2 et 3. Elle semble être la méthode la plus fiable et simple à mettre en œuvre.

L'hypothèse de dominance partielle est largement acceptée comme le mécanisme le plus fréquent causant la dépression de consanguinité (Fu & Ritland, 1994 ; Charlesworth & Willis, 2009). Cette hypothèse implique l'existence d'effets génétiques non-additifs (dominance et

épistasie). Sur la base de ces éléments, il nous semble important de considérer, lorsque la structure des données le permet, les effets de dominance dans les modèles génétiques afin d'évaluer correctement les effets de dépression de consanguinité, et ainsi mieux gérer le choix des reproducteurs.

5.2.2. Evaluation des effets de dépression de consanguinité et des anomalies génétiques

Dans le Chapitre 3 de cette thèse, ainsi que dans les résultats annexe 3 et 4, nous avons pu voir que la consanguinité pouvait avoir des effets variables, positifs comme négatifs, sur les performances (taille, production, reproduction femelle et survie) en fonction de sa localisation le long du génome. Il est donc essentiel dans l'évaluation des reproducteurs d'estimer les effets de la consanguinité localement le long du génome pour définir les accouplements à éviter (cf partie 5.4. Les plans d'accouplements), lorsque ces effets sont défavorables.

La dépression de consanguinité résulte le plus souvent de l'expression d'allèles récessifs délétères chez les individus homozygotes (cf Chapitre 1). La prise en compte des anomalies génétiques dans l'évaluation des candidats à la sélection est important afin d'éviter la propagation d'une mutation très délétère dans la population. Dans les programmes de sélection bovine, après avoir identifié les anomalies génétiques en se basant sur les déficits d'haplotypes homozygotes grâce aux données de génotypage et de séquençage (Fritz et al., 2013), les marqueurs associés sont utilisés pour pénaliser les candidats à la sélection plus ou moins fortement dans leur évaluation. Pour cela, les anomalies génétiques sont intégrées avec un poids variable dans l'index de sélection ISU (= Index Synthèse Unique) en fonction de l'importance de l'effet de la mutation d'un point de vue économique (coût souvent lié à la performance et à la survie des descendants), la fréquence de l'allèle délétère dans la population et le génotype des individus (Boichard et al., 2016 ; Michot, 2017).

Varona et al. (2019) proposent de gérer la consanguinité efficacement, en estimant une valeur globale de 'charge de dépression de consanguinité' ou de 'fardeau génétique' pour chaque reproducteur. Chaque individu sera évalué, via une analyse multivariée, sur la base de deux effets génétiques : a (exprimant directement le phénotype individuel) et i (exprimant le phénotype des descendants et correspondant au fardeau génétique, d'après Casellas (2018) qui considère que le fardeau génétique est un trait héritable).

5.2.3. Prise en compte de l'originalité génétique des individus

Une note d'originalité peut être estimée pour chaque candidat à la sélection afin d'avantager les individus génétiquement les plus distants de la population de reproducteurs, permettant ainsi de préserver la diversité génétique de la population (Bordonaro et al., 2012). Pour estimer l'originalité de chaque animal, on peut se baser sur l'apparentement d'un individu i par rapport à la population globale, ou bien sur les fréquences alléliques. Ainsi, dans certaines populations bovines, un indice d'originalité (ORI) est estimé à partir des fréquences alléliques (Danchin-Burge et al., 2016 ; Ganteil et al., 2021) :

$$\text{Index d'originalité } (ORI_i) = -\frac{f_i - f_{pop}}{\sigma_{f_{pop}}}$$

f_i : Fréquence moyenne des marqueurs portés par l'individu i (exemple : pour un seul locus homozygote AA, $f_i = (f_A + f_A)/2$).

f_{pop} : Moyenne des fréquences alléliques pour l'ensemble des marqueurs de la population, pour n individus, $f_{POP} = \sum_{i=1}^n \frac{f_i}{n}$

$\sigma_{f_{pop}}$: Ecart-type des moyennes des fréquences alléliques des marqueurs la population

Cet indice d'originalité permet de mettre en évidence des animaux porteurs d'allèles peu fréquents. Un animal original aura un ORI positif alors qu'un animal avec des allèles communs dans la population aura un ORI négatif. La simplicité des index ORI les rend facile à mettre en place. Danchin-Burge et al. (2016) ont proposé de calculer plusieurs ORI par candidat à la sélection, dans le but de les comparer à la population des animaux contemporains (jeunes), à tous les reproducteurs mâles ou à toutes les femelles en reproduction. Ces différents ORI permettent ensuite de pondérer l'ISU de chaque candidat afin d'avantager les individus originaux, par rapport à la population globale, mais également par rapport aux futurs reproducteurs (contemporains).

Une récente étude chez le porc a mis en évidence l'efficacité des indices d'originalité pour conserver de la diversité génétique dans les programmes de sélection (Ganteil et al., 2021). Cet indicateur pourrait être utilisé dans les programmes de sélection aquacole actuels afin de sélectionner un stock de géniteurs relativement divers. Dans le cas de la truite arc-en-ciel, seul l'index estimé sur la base de la population de candidats à la reproduction nous apparait

intéressant et pertinent. En effet, dans les programmes de sélection truiticole, les reproducteurs ne sont utilisés qu'une seule fois, pour générer la génération suivante, donc un ORI basé sur les anciens/nouveaux géniteurs n'a pas d'intérêt. Le calcul des ORI des candidats à la sélection par rapport à la population globale (50 000 individus avant la sélection massale) nécessiterait bien trop de génotypage à chacune des générations pour les comparer aux 2000 candidats à la sélection.

5.3. Le choix des reproducteurs

Le choix des reproducteurs est décisif car il permet le renouvellement de la population en lui assurant un progrès génétique, en lien avec les objectifs de sélection. On sélectionne donc les animaux ayant les meilleures valeurs génétiques, c'est-à-dire les meilleurs effets additifs des gènes, qui sont transmissibles à la descendance et permettent d'améliorer les performances des générations futures.

5.3.1. Interactions Génotype x Environnement et conservation de la diversité génétique dans le pool de reproducteurs

Dans le contexte actuel de forts changements climatiques et environnementaux, il nous semble essentiel de prendre en compte les effets d'interaction pour choisir les reproducteurs, dans le cas de caractères où des interactions Génotype x Environnement (GxE) sont parfois observées entre milieux contrastés (Phocas et al., 2017). De plus, dans plusieurs espèces de poissons dont la TAC, des niveaux d'interactions GxE modérés à forts ont été observés, en particulier sur des caractères de croissance (Ayles, 1975 ; Fishback et al., 2002 ; Dupont-Nivet et al., 2010 ; Sae-Lim et al., 2013, 2016 ; Li et al., 2019). Pour prendre en compte les interactions GxE, deux options s'offrent à nous : i) sélectionner les individus les plus résilients, c'est-à-dire ayant une valeur génétique stable entre différents environnements, indicateur de la capacité d'adaptation des animaux à différents milieux (Dumont et al., 2014) ; ou ii) choisir les meilleurs candidats pour l'environnement de production ciblé. Dans le contexte d'exportation d'œufs et d'alevins améliorés par les sélectionneurs truiticoles français dans de nombreux pays aux conditions de milieu contrastées, il nous semble important de favoriser la première option,

qui permet également de préserver plus de diversité génétique. En effet, les populations les plus diverses génétiquement et les individus les moins consanguins sont généralement jugés plus résilients face à des modifications de leur environnement (Kristensen et al., 2010). Cette importance de la diversité génétique est à mettre en perspective avec les résultats décrits au Chapitre 4, où quatre régions hétérozygotes (non consanguines) ont été identifiées communes à plusieurs populations de truites arc-en-ciel. Dans ces régions, des gènes ont été identifiés comme ayant un effet essentiel pour l'organisation cellulaire, le développement embryonnaire et les fonctions immunitaires. Nous supposons que cela est lié à la nécessité d'avoir un fort polymorphisme génétique afin de remplir les fonctions nécessaires aux diverses étapes de développement embryonnaire en toutes circonstances, ou encore pour faire face à une variété de pathogènes, virus et bactéries (Gu et al., 2016 ; Durland et al., 2021).

5.3.2. Elimination des candidats à la reproduction porteurs d'allèles délétères

Dans les populations où la gestion des anomalies génétiques est rendue nécessaire, cette dernière doit se faire différemment selon que l'anomalie est létale ou non, dominante ou récessive (Boichard et al., 2016). En effet, il est difficilement envisageable d'éliminer systématiquement les individus porteurs d'anomalies génétiques récessives des groupes de reproducteurs potentiels, contrairement à ce qui a pu être envisagé dans le cas d'anomalies dominantes létales. En races bovines Montbéliarde et Holstein, respectivement 5 et 12 anomalies génétiques létales sont intégrées dans le choix des reproducteurs (Michot, 2017; Hoze et al., 2018). Pour cela, les index ISU des candidats porteurs d'anomalies récessives létales sont pénalisés en fonction de la gravité et de l'impact économique de la mutation (Hoze et al., 2018 ; Bérodiér, 2020 ; cf partie 5.2.2). L'élimination systématique des reproducteurs porteurs d'anomalies récessives aurait pour conséquence de créer un goulot d'étranglement qui ferait chuter la diversité génétique dans la population (Howard et al., 2017). De plus, cette stratégie d'éradication s'applique seulement 'allèle par allèle', car il est impossible d'éliminer tous les allèles délétères du génome de cette manière. Contrairement aux ruminants, la truite arc-en-ciel est une espèce hyper-prolifique, les anomalies létales dans les premiers stades de vie n'ont donc que peu d'impact économique. Dans cette espèce, la pénalisation forte des reproducteurs portant certaines anomalies récessives létales lors du développement embryonnaire ne semble pas être une solution à privilégier car elle conduirait essentiellement à une baisse globale de la diversité génétique.

Une autre stratégie, basée sur le modèle proposé par Varona et al. (2019), permettrait de choisir les individus ayant une estimation de ‘fardeau génétique’ la plus faible possible et donc la plus favorable pour limiter les effets de dépression de consanguinité sur la descendance (Martinez-Castillero et al., 2021). Cette stratégie pourrait être appliquée dans les programmes d’amélioration génétique truiticole afin d’éliminer, non pas ‘allèle par allèle’, mais globalement les individus porteurs d’un fardeau génétique élevé.

5.3.3. Conservation d’un nombre minimal de reproducteurs

Afin de limiter l’érosion de la diversité génétique d’une génération à l’autre, Il faut s’assurer d’avoir un réservoir génétique suffisant, c’est-à-dire avoir un nombre suffisant de reproducteurs génétiquement variés, pour limiter le niveau de consanguinité dans la génération suivante. En effet, plus le nombre d’animaux sélectionnés pour la reproduction est faible, plus la taille efficace de la population décline et donc plus la diversité génétique de la population diminue rapidement au cours des générations.

Sur la base des recommandations faites par le FAO, il est conseillé de limiter l’accroissement de la consanguinité de 0,5 à 1% par génération (FAO, 2009). Cela nous permet d’estimer le nombre de reproducteurs minimal à conserver dans le stock de reproducteurs. Sur la base de la formule classique donné dans le Chapitre 1, on peut estimer un nombre de 25 paires de reproducteurs. Toutefois, Bentsen & Olesen (2002) ont montré, sur la base de simulations stochastiques d’une sélection massale sur 15 générations, qu’un minimum de 50 paires de reproducteurs devait être utilisé par génération afin de limiter l’accroissement de la consanguinité à 1% par génération. Cette différence peut tout simplement s’expliquer par le fait que la formule initiale sur une génération implique que les reproducteurs soient non apparentés. On peut cependant s’attendre à ce que le niveau d’apparentement des reproducteurs augmente au cours des générations, augmentant donc le nombre de paires de reproducteurs indispensable pour limiter l’accroissement de la consanguinité au sein de la population.

5.4. Les plans d'accouplement

Le choix des plans d'accouplement (croisements hiérarchiques ou factoriels) est la partie fondamentale d'un programme d'amélioration génétique pour former les bons croisements pour produire la génération suivante. Dans le programme de sélection d'une lignée pure, la gestion de la diversité génétique se fait majoritairement en limitant l'apparement des reproducteurs utilisés dans le plan d'accouplement. Ainsi, dans les schémas de sélection truiticole, on limite l'accouplement d'animaux trop apparentés (cousins germains) (D'Ambrosio, 2020). La stratégie d'amélioration en croisement permet de valoriser la diversité génétique en optimisant les performances du produit final grâce aux effets d'hétérosis attendus, ainsi qu'à la bonne complémentarité génétique entre les reproducteurs des lignées utilisées dans le croisement.

5.4.1. Choix du plan d'accouplement

Que ce soit pour limiter l'augmentation de la consanguinité ou pour améliorer le gain génétique d'une génération à l'autre, Sonesson & Ødegård (2016) ont montré, en simulant des schémas de sélection aquacole, que l'utilisation d'un plan factoriel 10:10 à la place d'un plan 1:1 ou 1:2 (de type hiérarchique) était bien plus performant. Un plan factoriel 10:10 correspond à un croisement de 10 femelles avec 10 mâles, autrement dit, chaque femelle est croisée avec les 10 mâles et inversement. De plus, les plans factoriels 10 :10, avaient déjà été montrés plus efficaces que les plans d'accouplement hiérarchiques pour limiter l'augmentation de la consanguinité (Dupont-Nivet et al., 2006).

Les plans factoriels partiels constitués d'environ 10 blocs 10 :10 sont les plans d'accouplement utilisés en France depuis plus de vingt ans dans les schémas de sélection de la truite arc-en-ciel pour croiser à chaque génération une centaine de reproducteurs mâles avec une centaine de reproducteurs femelles (D'Ambrosio, 2020). Un plan factoriel complet de 100 reproducteurs de chaque sexe est coûteux à entretenir, et les gains génétiques attendus du plan factoriel complet et du plan factoriel partiel sont similaires, à nombre total de reproducteurs constant (Dupont-Nivet et al., 2002, 2006). Ainsi, l'utilisation des plans factoriels en sélection est une méthode classique, facile à mettre en place chez la truite arc-en-ciel, permettant d'améliorer l'efficacité de la sélection, ainsi que de préserver la variabilité génétique de la population. Qui plus est, au sein des plans d'accouplement factoriels, par rapport aux plans d'accouplement hiérarchiques, un grand nombre d'individus est croisé, permettant d'exploiter

l'information génétique associée à des individus avec des degrés d'apparentement variés et à diverses combinaisons génétiques (pleins-frères et demi-frères) indispensables à l'estimation correcte des effets maternels et de dominance (Haffray et al., 2018).

5.4.2. Gestion de la consanguinité

L'apparentement des reproducteurs, estimé à l'aide de données sur pedigree, peut induire une forte limitation dans la gestion de la diversité génétique au sein des populations de truites arc-en-ciel. En effet, le pedigree parfois très incomplet dans les populations de truites arc-en-ciel peut conduire à une forte sous-estimation des apparentements et donc de la consanguinité réelle dans certaines populations. Il a été démontré que, dans le cadre de sélection génomique, la consanguinité doit être gérée sur la base d'informations génomiques plutôt que sur la base d'informations pedigree (Sonesson et al., 2012 ; Sun et al., 2013 ; Kardos et al., 2015 ; Ceballos et al., 2018; Doekes et al., 2021). En effet, l'apparentement calculé sur le pedigree ne prend pas en compte l'aléa de méiose, qui peut alors différencier deux pleins-frères (Kardos et al., 2015 ; Wang, 2016 ; Howard et al., 2017). Il est donc nécessaire de se baser sur des données d'apparentement génomique afin d'avoir une valeur 'vraie' d'apparentement entre deux individus. De plus, la localisation dans le génome de la consanguinité ne peut pas être identifiée sur pedigree. Or, cela peut grandement limiter la précision de la gestion de la consanguinité et de ses effets.

5.4.2.1. La localisation de la consanguinité dans le génome, une information clef

Les études réalisées dans cette thèse ont démontré que les estimations des effets globaux de la consanguinité ne représentent qu'une moyenne des effets de la consanguinité locale à l'échelle des génomes, avec peu d'effets significatifs malgré des effets parfois importants estimés localement (cf Chapitres 2 et 3, résultats annexe 3 et 4). Les 'vrais' effets de la consanguinité sont donc à estimer localement afin d'avoir une bonne vision des régions où la dépression de consanguinité sévit, ou, *a contrario*, un avantage a été identifié pour l'homozygote porteur d'allèles favorables. En comparant les traces de domestication/sélection observées dans quatre populations de truites, des signatures de sélection positive (avantage à

l'homozygote favorable), mais également de sélection équilibrante (avantage à l'hétérozygote) ont pu être identifiées comme communes à ces populations diverses. L'une étant notamment constituée d'un pool d'individus issus d'écloseries et rivières du Nord-Ouest des USA, et les trois autres, étant des populations françaises (cf Chapitre 4). Il est notable que parmi les 13 zones identifiées sous signature de sélection commune, 18 gènes avaient déjà été mentionnés dans des recherches de traces de sélection dans des espèces variées de vertébrés (bovins, ovins, chiens...). Ces traces de sélection semblent donc être liées à des avantages primordiaux pour assurer la *fitness* (développement embryonnaire, développement cellulaire, etc.) et la domestication (croissance, comportement, reproduction, immunité), qu'il s'agit donc de préserver (cf Chapitre 4).

C'est pourquoi, si l'on veut gérer précisément la consanguinité de nos populations de truites, la valeur globale de l'apparentement entre les reproducteurs n'est pas la seule information à considérer dans les programmes de sélection. La localisation de l'apparentement entre deux individus peut avoir un impact majeur sur la *fitness* de leurs descendants. De plus, l'impact de la consanguinité étant globalement variable entre les caractères, une caractérisation fine multi-caractère des effets de la consanguinité dans le génome de la truite est une étape essentielle avant d'être en mesure de la piloter dans les programmes de sélection.

Pour illustrer l'intérêt et l'impact économique que peut avoir une gestion locale des effets de consanguinité par rapport à une gestion globale, je vous présente différentes stratégies de gestion de la consanguinité au sein d'une population de truite sous sélection. Cet exemple est basé sur le nombre d'œufs produits par femelle, qui est un caractère pouvant être directement valorisé par les sélectionneurs français (la valeur d'un lot de 1000 œufs de truite peut être estimée à 12€, D'Ambrosio, 2020). Considérons ici que le nombre moyen d'œufs par ponte est de 4802 œufs et que les effets locaux de la consanguinité sur le nombre d'œufs sont ceux estimés dans le Chapitre 3. Dans le cas où, seuls l'effet global de la consanguinité est considéré (scénario 1), lorsque le coefficient de consanguinité moyen augmente de 0,10 point, la dépression de consanguinité conduit à une diminution de 49 œufs par ponte, soit une perte économique de 1%, c'est-à-dire à une perte de 0,59€/femelle. Imaginons ensuite deux cas de figures extrêmes : i) scénario 2 : la consanguinité supplémentaire se porte sur les chromosomes où elle est défavorable au nombre d'œufs par femelle ; ii) scénario 3 : elle se porte exclusivement sur les chromosomes où elle induit un effet positif sur le nombre d'œufs par femelle. Dans le scénario 2, une augmentation de 0,10 point de consanguinité induit une perte

de 323 œufs par ponte, alors que dans le scénario 3, l'augmentation de la consanguinité (+ 0,10 points) se traduit par une augmentation de 256 d'œufs par ponte, conduisant soit à une perte économique de 6,7% (perte de 3,88€/femelle) soit à un gain de 5,3% (gain de 3,07€/femelle).

Cela illustre clairement qu'une gestion des seuls effets de la consanguinité globale peut grandement sous-estimer la perte économique associée à la dépression de consanguinité (scénario 1 versus scénario 2). De plus, on peut observer qu'un levier d'amélioration des performances de production des œufs est envisageable grâce aux effets positifs d'une consanguinité localement bien choisie sur le génome. C'est pourquoi, il me semble essentiel de gérer les niveaux de consanguinité locaux dans les programmes d'amélioration génétique, notamment en introduisant le suivi des apparentements génomiques locaux dans le choix des accouplements.

5.4.2.2. Quelles sont les méthodes disponibles pour une gestion de la diversité génétique à l'échelle locale des génomes ?

Varona et al. (2019) proposent de gérer efficacement la consanguinité sans forcément se placer au niveau local en ajoutant un critère d'évaluation des reproducteurs, avec une valeur globale de 'fardeau génétique' estimée pour chacun d'eux (cf partie 5.2.2. Evaluation des effets de la dépression de consanguinité et des anomalies génétiques).

Toutefois, plusieurs approches basées sur l'Optimum Contribution (OC) permettent de gérer la consanguinité localement au niveau des génomes.

Gómez-Romano et al. (2016) proposent une gestion locale de la consanguinité avec une approche d'optimum contribution (Meuwissen, 1997). Pour cela, des coefficients d'apparentement locaux sont utilisés, via des matrices d'apparentement génomique, afin de choisir au mieux les accouplements, minimisant l'apparentement des reproducteurs dans certaines régions génomiques où la diversité génétique est essentielle. Pour chaque locus, la matrice d'apparentement génomique est construite de la manière suivante : si les deux individus partagent le même génotype, la valeur d'apparentement du locus sera de 1, sinon elle sera de 0.

Les matrices d'apparentement génomique peuvent être basées sur divers types d'informations génomiques : 1) l'évolution des fréquences alléliques (Meuwissen et al., 2011 ; Meuwissen et al., 2020), 2) l'état homozygote/hétérozygote d'un locus (Nejati-Javaremi et al.,

1997 ; De Beukelaer et al., 2017), 3) les ROH (De Cara et al., 2013), ou 4) des segments IBD (Meuwissen et al., 2001 ; VanRaden, 2008 ; Meuwissen et al., 2020). Le choix du type d'information utilisée pour construire la matrice d'apparentement est important et pourra avoir un impact sur le résultat final de la gestion de la consanguinité d'une population (Meuwissen et al., 2020).

Meuwissen et al. (2020) proposent, tout comme Gómez-Romano et al. (2016), une méthode basée sur l'utilisation d'informations génomiques et dérivée de l'Optimum Contribution : Genomic OC. Cependant, ils observent que l'utilisation de matrices d'apparentement génomique basée sur l'information IBD est plus performante pour gérer la diversité génétique au cours des générations tout en maintenant un progrès génétique suffisant. De plus, pour gérer la diversité génétique, ils proposent d'utiliser deux matrices génomiques : une première comprenant l'information globale (valeur globale de l'IBD entre individus), et une seconde comprenant l'information des IBD locaux, permettant de gérer localement la diversité génétique. Cela permet donc de garder le contrôle sur l'évolution de la consanguinité globale de la population, tout en gérant également la diversité génétique locale dans les régions où elle a été identifiée comme essentielle.

5.4.3. Valorisation de l'hétérosis

La valorisation des effets génétiques de dominance, à travers le phénomène d'hétérosis (ou overdominance), peut améliorer significativement les performances des descendants. Cette valorisation engendre la création de génotypes hétérozygotes, ce qui réduit le niveau de consanguinité de la nouvelle génération.

5.4.3.1. Prise en compte des effets d'hétérosis dans un programme d'amélioration génétique en lignée pure

Dans le cadre de ma thèse, j'ai pu montrer qu'une part non négligeable de la variabilité phénotypique de la survie précoce des alevins, ainsi que des caractères liés à la taille et la reproduction des femelles d'une population commerciale de truites arc-en-ciel, étaient expliqués par des effets génétique de dominance (cf Chapitre 2 et 3). Ces effets de dominance mériteraient donc d'être exploités dans les plans d'accouplement des populations multiplicatrices afin d'augmenter la performance des descendants en production, mais

également les performances de reproduction des parents. Cette stratégie limiterait également l'érosion de la diversité génétique dans les populations de truites (Wellmann & Bennewitz, 2012).

L'utilisation de la dominance permettrait aussi d'améliorer des caractères peu héritable, comme la survie des poissons (Chapitre 2). Pour rappel, les effets de dominance expliquaient entre 5 et 7% de la variance phénotypique pour des caractères de taille des truites, ce qui était relativement faible par rapport aux estimations obtenues pour les caractères de reproduction et de survie, pour lesquels nous avons estimé des proportions allant jusqu'à 22% de la variance phénotypique.

Plusieurs études, que cela soit pour des espèces terrestres (Toro & Varona, 2010 ; Sun et al., 2013 ; Aliloo et al., 2017) ou bien aquacoles (Fernández et al., 2021), exposent l'intérêt de valoriser les effets d'hétérosis dans les plans d'accouplement au sein des programmes d'amélioration génétique en lignée pure. Toro & Varona (2010) ont démontré par simulation (à partir de performances simulées sur la base d'une héritabilité comprise entre 20 et 40% et d'effets de dominance expliquant entre 5 et 10% de la variance phénotypique) qu'en considérant les effets de dominance lors de la gestion des accouplements, la réponse à la sélection était améliorée (+ 6 à 22%). Sun et al. (2013) montrent également que l'utilisation des effets de dominance, dans les programmes d'amélioration génétique en race pure (bovin), permet d'avoir des descendants avec de meilleures performances et cela malgré de faibles variances de dominance (expliquant seulement 4% de la variance phénotypique).

5.4.3.2. Prise en compte des effets d'hétérosis dans un programme d'amélioration génétique en croisement

Tout comme dans les schémas d'amélioration génétique des porcs et des volailles qui sont réalisés avec des croisements trois ou quatre voies (Sellier, 1976 ; Stock et al., 2020), on pourrait envisager au vu des effets de dominance estimés chez la truite arc-en-ciel, une stratégie de croisement de plusieurs lignées pour les populations multiplicatrices. La sélection sur les lignées paternelles serait sur la croissance, l'efficacité alimentaire et les qualités de découpe et de chair, alors que les lignées maternelles seraient elles-mêmes issues d'un croisement entre deux lignées grand-parentales permettant de valoriser l'hétérosis sur les caractères de

reproduction et de survie des alevins. Cela permettrait d'augmenter le nombre et la survie des juvéniles en évitant les phénomènes de dépression de consanguinité (González-Diéguez et al., 2019, 2020). L'importance des effets de dominance estimés intra-lignée chez la truite arc-en-ciel est un indicateur de l'intérêt potentiel qu'il pourrait y avoir à valoriser les effets d'hétérosis à travers d'accouplements inter-lignées.

Cette stratégie permettrait d'augmenter de manière importante les performances dans les populations de production, c'est-à-dire la performance des alevins exportés par les entreprises françaises, ce qui augmenterait leur compétitivité. Toutefois, pour bénéficier d'effets d'hétérosis forts dans la population de production, il est nécessaire de croiser et donc d'entretenir au moins deux lignées paternelles et maternelles sélectionnées séparément, afin qu'elles soient génétiquement différentes. Il faut, cependant, éviter de croiser des lignées pures trop éloignées, car un croisement hybride entre lignées extrêmes peut conduire à une dépression hybride, c'est-à-dire à une valeur sélective de la descendance inférieure à celle des parents (Edmands, 2007). Ce phénomène, souvent observé dans des populations végétales, a été mis en évidence dans plusieurs espèces de poissons (Gharrett et al., 1999 ; Neff, 2004 ; Granier et al., 2011 ; Huff et al., 2011). Il est généralement dû à une rupture d'association de gènes coadaptés (phénomènes d'épistasie).

5.4.4. Purge du fardeau génétique

La purge du fardeau génétique, c'est-à-dire l'élimination des allèles délétères (Barrett & Charlesworth, 1991 ; Lande et Schemske, 1985), est une théorie selon laquelle, l'obtention de niveaux de consanguinité et de dérives élevés dans une petite population expose les mutations délétères à la sélection naturelle. Cela peut avoir deux conséquences : la diminution de la valeur sélective moyenne de la population par augmentation des phénomènes de dépression de consanguinité, ou la purge des mutations délétères. Dans les espèces hyper-fécondes comme les poissons, on peut imaginer organiser volontairement cette purge génétique en accouplant des reproducteurs très apparentés entre eux. Cette tactique permet de concentrer les allèles favorables dans la population d'alevins survivants. Toutefois, cette stratégie pour limiter les phénomènes de dépression de consanguinité à long terme est très délicate, car cela peut causer une perte de diversité génétique drastique et donc engendrer un risque élevé d'extinction de la population en l'absence de capacité évolutive suffisante (Hedrick, 1994 ;

Frankham & Ralls, 1998). Cette stratégie est donc associée à un risque non négligeable d’aboutir au résultat inverse de ce qui était initialement souhaité.

5.5. Gestion des déséquilibres de contribution des reproducteurs à la génération suivante

A la fin du programme de sélection, après la gestion des accouplements et de la fécondation, des verrous reproductifs généralement présents dans les espèces aquatiques très prolifiques peuvent conduire à une baisse de la taille efficace et donc de la diversité génétique de la population (Hedgewood & Pudovkin, 2011).

Nous avons pu voir dans l’étude du Chapitre 2, que les femelles, les plus petites et les plus fécondes, semblent produire des descendants plus résistants, créant ainsi un déséquilibre de contribution à la génération suivante entre reproducteurs sur la voie maternelle. Ces femelles ont significativement moins de mortalité dans leur descendance que les autres mères, et participent donc plus à la génération suivante, ce qui implique une baisse de la taille efficace de la population. La surmortalité dans certaines descendances maternelles pourrait être liées à des œufs de taille anormalement élevée et un nombre de malformations accrues, mais ce résultat n’est vraiment observé que sur la descendance d’une femelle extrême parmi les 15 femelles du plan d’accouplement expérimental.

Actuellement, dans les schémas de sélection de type PROSPER pratiqués dans les élevages de truites arc-en-ciel français, après la fécondation des ovocytes, un premier regroupement des œufs de taille similaire est effectué avant la sélection sur la taille des descendants qui se fera sur les alevins (Chevassus, 2004). Finalement, cette différence de mortalité des descendants entre mères semble être déjà prise en compte par cette méthode dans les populations de truites arc-en-ciel françaises. Notre étude a été réalisée sur une lignée non-sélectionnée de truites pour laquelle aucun tri sur la taille des œufs n’a été effectué. Cette différence pourrait expliquer la forte baisse de l’effectif efficace (N_e) de la lignée synthétique INRAE (D’Ambrosio et al., 2019), malgré le croisement d’un nombre important de parents et la quasi-absence de sélection volontaire.

Dans le cadre de la conservation de la lignée synthétique, mais également pour limiter l’érosion de sa diversité génétique, il faudrait, dans les premières semaines après la fécondation

des œufs, élever les embryons par pool de taille d'œufs afin de limiter les effets maternels. Cela permettrait d'avoir une contribution similaire entre les femelles concernant le nombre d'alevins après l'éclosion, en lien direct avec le nombre d'individus à la génération suivante. Cependant, il ne faut pas oublier qu'un déséquilibre de contribution parental peut également subvenir bien après la fin de notre expérimentation, c'est-à-dire lors de la croissance des alevins. Une étude complémentaire pourrait être utile pour identifier si des déséquilibres de contribution surviennent ultérieurement dans le développement des alevins, pouvant participer à la perte de diversité génétique de la diversité génétique de la truite arc-en-ciel.

5.6. Que faire en cas de situation critique ?

Afin de s'assurer que les populations sous sélection ne passent pas sous un seuil critique de diversité génétique, pouvant conduire à l'extinction de la population, il est nécessaire de mettre en œuvre une surveillance génomique. Plusieurs indicateurs peuvent être employés à cet effet. Tout d'abord la mesure génomique de la taille efficace de la population, une baisse trop importante devrait alerter les sélectionneurs. D'autres indicateurs classiques de la diversité génétique, comme l'évolution des niveaux d'hétérozygotes observés et attendus (H_o/H_e), les niveaux de dépression de consanguinité observés, ainsi que l'évolution des fréquences alléliques pourraient également être utilisés pour surveiller l'évolution de la consanguinité des populations (Meuwissen, 2009 ; Flanagan et al., 2018). A l'échelle des espèces ou des races/lignées, il existe une mesure de la diversité selon Weitzman dont une étude a montré l'intérêt pour définir les stratégies de conservation de certaines races bovines en danger (d'Arnoldi et al., 1998).

En cas de situation critique, le sauvetage génétique est une des solutions permettant de lutter contre le déclin des petites populations (Hedrick, 2009, 2014 ; Whiteley et al., 2015). Le sauvetage génétique permet de restaurer de la diversité génétique en introduisant de nouveaux allèles favorables, pouvant répondre à de nouveaux objectifs de sélection ou simplement à des changements environnementaux. Un sauvetage génétique réussi permet une augmentation de la diversité génétique d'une population, une baisse de la dépression de consanguinité et induit généralement une augmentation de la taille efficace de la population concernée. Le matériel génétique à utiliser pour ré-introduire de la diversité génétique dans la population peut provenir d'une autre population *in vivo*, c'est-à-dire une population avec des individus actuellement 'vivants'. Il est également possible de réinjecter de la variabilité génétique dans les populations

en utilisant les gamètes cryoconservées d'animaux originaux (Doekes et al., 2018). Ainsi, un plan de sauvegarde génétique a été mis en œuvre il y a plus de vingt ans dans une race bovine française locale à très faible effectif, la Bretonne Pie-Noire, (Colleau et al., 2002).

Le matériel cryoconservé est une ressource précieuse pour gérer la diversité génétique des animaux domestiqués. En effet, si l'on considère du matériel génétique suffisamment éloigné des ressources génétiques des populations 'vivantes', l'apport d'allèles anciens éliminés et donc nouveaux pour la population peut être intéressant. La limite, qui se pose lors de la réintroduction de matériel cryoconservé ancien, est que généralement le niveau génétique de la semence cryoconservée depuis plusieurs années est bien inférieur à la valeur génétique de la population actuelle. Dans ce cas, l'utilisation de matériel cryoconservé peut induire une forte perte de la valeur génétique de la population. C'est pourquoi, le sauvetage génétique reste à ce jour très peu, voire pas utilisé dans les populations en sélection. De plus, il n'existe pas ou peu de recommandations et d'outils pour aider les sélectionneurs et gestionnaires des populations à choisir les meilleurs individus cryoconservés pour répondre à leurs objectifs de sélection. Pour pallier aux limites que pose le sauvetage génétique, que ce soit avec du matériel vivant ou cryoconservé, la génomique pourrait être utilisée en routine afin de choisir au mieux les individus originaux, permettant ainsi un sauvetage génétique efficace, tout en limitant la perte de niveau génétique dans la population. Ainsi, pour restaurer de la diversité génétique dans la lignée synthétique INRAE de truites arc-en-ciel à l'aide de semence cryoconservée, une approche inspirée du sauvetage génétique et de la méthode OCS a récemment été proposée au sein de l'équipe (Colon, 2022) à l'aide d'informations génomiques en quantités limitées. La stratégie de maintien de la diversité génétique sur le long terme proposée peut s'appliquer à faibles coûts humains et financiers aux autres lignées expérimentales, ainsi qu'à toutes populations commerciales où seules des actions ponctuelles de génotypage et cryoconservation de la semence seraient envisageables.

Conclusion

Pour conclure, comprendre l'impact et le rôle de la diversité génomique locale est essentiel afin de mieux la gérer dans les programmes de sélection, et ainsi maîtriser les compromis entre valeurs génétiques des reproducteurs et les effets indésirables de la consanguinité ainsi que de la perte de diversité génétique dans les populations sélectionnées. De plus, en raison de l'importance des effets de dominance observés, concevoir des programmes d'amélioration génétique en croisement serait une perspective intéressante à considérer pour améliorer les performances des cheptels de truite arc-en-ciel en production tout en protégeant la génétique créée par les sélectionneurs.

RÉFÉRENCES

- Adams, S. M., Derks, M. F., Makanjuola, B. O., Marras, G., Wood, B. J., & Baes, C. F. (2021). Investigating inbreeding in the turkey (*Meleagris gallopavo*) genome. *Poultry science*, 100(11), 101366. <https://doi.org/10.1016/j.psj.2021.101366>
- Addo, S., Klingel, S., Thaller, G., & Hinrichs, D. (2021). Genetic diversity and the application of runs of homozygosity-based methods for inbreeding estimation in German White-headed Mutton sheep. *PLoS one*, 16(5), e0250608. <https://doi.org/10.1371/journal.pone.0250608>
- AGRESTE. Enquête Aquaculture (2020). Recensement Pisciculture, culture d'algues et de cyanobactéries 2020. Internet : https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/Chd2116/cd2021-16_Aqua2020.pdf
- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., Goddard, M. E., & Hayes, B. J. (2017). Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *Journal of Dairy Science*, 100(2), 1203-1222. <https://doi.org/10.3168/jds.2016-11261>
- Aulstad, D., & Kittelsen, A. (1971). Abnormal body curvatures of rainbow trout (*Salmo gairdneri*) inbred fry. *Journal of the Fisheries Board of Canada*, 28(12), 1918-1920. <https://doi.org/10.1139/f71-290>
- Ayles, G. B. (1975). Influence of the genotype and the environment on growth and survival of rainbow trout (*Salmo gairdneri*) in central Canadian aquaculture lakes. *Aquaculture*, 6(2), 181-188. [https://doi.org/10.1016/0044-8486\(75\)90069-1](https://doi.org/10.1016/0044-8486(75)90069-1)
- Baes, C., Makanjuola, B., & Schaeffer, L. (2019). Assessing inbreeding and genetic diversity in the Holstein breed using pedigree and genomic approaches. In *Advances in breeding of dairy cattle* (pp. 21-38). Burleigh Dodds Science Publishing.
- Baird, S. F. (1886). Report of the Commissioner, pp. IX–LVII. In: Report of the Commissioner for 1884. Washington: Taylor & Francis
- Balon, E. K. (1995). Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture*, 129(1-4), 3-48. [https://doi.org/10.1016/0044-8486\(94\)00227-F](https://doi.org/10.1016/0044-8486(94)00227-F)
- Balon, E. K. (2004). About the oldest domesticates among fishes. *Journal of fish Biology* 65: 1-27. <https://doi.org/10.1111/j.0022-1112.2004.00563.x>
- Barrett, S. C. H., & Charlesworth, D. (1991). Effects of a change in the level of inbreeding on the genetic load. *Nature*, 352(6335), 522-524. <https://doi.org/10.1038/352522a0>
- Barria, A., López, M. E., Yoshida, G., Carneiro, R., Lhorente, J. P., & Yáñez, J. M. (2018). Population genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon (*Salmo salar* L.) using dense SNP genotypes. *Frontiers in genetics*, 9, 649. <https://doi.org/10.3389/fgene.2018.00649>
- Bateson, W., & Mendel, G. (2013). *Mendel's principles of heredity*. Courier Corporation.
- Baran, P., Delacoste, M., Lascaux, J. M., Bauba, F., & Segura, G. (1995). La compétition interspécifique entre la truite commune (*Salmo trutta* L.) et la truite arc-en-ciel (*Oncorhynchus Mykiss* Walbaum): influence sur les modèles d'habitat. *Bulletin Français de la Pêche et de la Pisciculture*, (337-338-339), 283-290. <https://doi.org/10.1051/kmae:1995032>
- Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in ecology & evolution*, 23(1), 38-44. <https://doi.org/10.1016/j.tree.2007.09.008>
- Bean, T. H. (1895). The rainbow trout. *First annual report of the Commissioners of Fisheries, game and forest of the State of New York*, 135-140.

REFERENCES

- Bed'Hom, B., & Chazara, O. (2011). High resolution genotyping of a very polymorphic region involved in response to pathogens: MHC (Major Histocompatibility Complex). *Actes des 9èmes Journées de la Recherche Avicole, Tours, France, 29 et 30 mars 2011*, 779-783.
- von Behr, Friedrich. (1882). *Five American Salmonidae in Germany*. US Government Printing Office.
- Bentsen, H. B., & Olesen, I. (2002). Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture*, 204(3-4), 349-359. [https://doi.org/10.1016/S0044-8486\(01\)00846-8](https://doi.org/10.1016/S0044-8486(01)00846-8)
- Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M., ... & Phocas, F. (2022). Development of a high-density 665 K SNP array for rainbow trout genome-wide genotyping. *Frontiers in Genetics*. <https://doi.org/10.3389/fgene.2022.941340>
- Bérodier, M. (2020). Utilisation en ferme des données de génotypage pour une gestion optimisée et durable de l'élevage laitier (Doctoral dissertation, université Paris-Saclay). <https://pastel.archives-ouvertes.fr/tel-02940692>
- Bertrand, A. R., Kadri, N. K., Flori, L., Gautier, M., & Druet, T. (2019). RZooRoH: an R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, 10(6), 860-866. <https://doi.org/10.1111/2041-210X.13167>
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., ... & Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications*, 5(1), 1-10. <https://doi.org/10.1038/ncomms4657>
- Birchler, J. A., D. L. Auger, and N. C. Riddle. (2003). In search of the molecular basis of heterosis. *The Plant Cell* 15:2236–2239. <https://doi.org/10.1105/tpc.151030>
- Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. <https://cran.r-project.org/web/packages/detectRUNS/index.html>
- Bobé et al. (2010) Chapitre 2. Sexualité et reproduction, in: La truite arc-en-ciel : De la Biologie à l'élevage. Quae, pp. 39–79.
- Boichard, D., Grohs, C., Michot, P., Danchin-Burge, C., Capitan, A., Genestout, L., ... & Fritz, S. (2016). Prise en compte des anomalies génétiques en sélection: le cas des bovins. *INRA Productions Animales*, 29(5), 351-358. <https://hal.archives-ouvertes.fr/hal-01602349/document>
- Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, 186(1), 241-262. <https://doi.org/10.1534/genetics.110.117275>. <https://doi.org/10.1534/genetics.110.117275>
- Bordonaro, S., Guastella, A. M., Criscione, A., Zuccaro, A., & Marletta, D. (2012). Genetic diversity and variability in endangered Pantesco and two other Sicilian donkey breeds assessed by microsatellite markers. *The Scientific World Journal*, 2012. <https://doi.org/10.1100/2012/648427>
- Botero-Delgado, E., Gilseman, C., Mueller, J. C., & Kempnaers, B. (2020). Negative effects of individual heterozygosity on reproductive success in a wild bird population. *Molecular Ecology*, 29(17), 3196-3216. <https://doi.org/10.1111/mec.15553>
- Breton et al. (2010) Chapitre 10. Reproduction in: La truite arc-en-ciel : De la Biologie à l'élevage. Quae, pp. 213–224.
- Brundtland, G. H. (1987). Rapport de la Commission mondiale de l'environnement et le développement de l'ONU. Accès: http://fr.wikisource.org/wiki/Notre_avenir_à_tous_-_Rapport_Brundtland. Consulté le, 24-09-, 2022.
- Burren, A., Neuditschko, M., Signer-Hasler, H., Frischknecht, M., Reber, I., Menzi, F., ... & Flury, C. (2016). Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Animal genetics*, 47(6), 727-739. <https://doi.org/10.1111/age.12476>

- Caballero, A., and P. D. Keightley. (1998). Inferences on genome-wide deleterious mutation rates in inbred populations of *Drosophila* and mice. *Genetica* 102:229–239. https://doi.org/10.1007/978-94-011-5210-5_19
- Cambray, J. A. (2003). Impact on indigenous species biodiversity caused by the globalisation of alien recreational freshwater fisheries. *Hydrobiologia*, 500(1), 217-230. <https://doi.org/10.1023/A:1024648719995>
- Carrillo, J.A., et Siewerdt, F. (2010). Consequences of long-term inbreeding accumulation on preweaning traits in a closed nucleus Angus herd. *J. Anim. Sci.* 88, 87–95. <https://doi.org/10.2527/jas.2009-1897>
- Casellas, J. (2018). On individual-specific prediction of hidden inbreeding depression load. *Journal of Animal Breeding and Genetics*, 135(1), 37-44. <https://doi.org/10.1111/jbg.12308>
- Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, 19(4), 220-234. <https://doi.org/10.1038/nrg.2017.109>
- Cendron, F., Mastrangelo, S., Tolone, M., Perini, F., Lasagna, E., & Cassandro, M. (2021). Genome-wide analysis reveals the patterns of genetic diversity and population structure of 8 Italian local chicken breeds. *Poultry science*, 100(2), 441-451. <https://doi.org/10.1016/j.psj.2020.10.023>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742-015. <https://doi.org/10.1186/s13742-015-0047-8>
- Chapuis, H., Pincent, C., & Colleau, J. J. (2016). Optimizing selection with several constraints in poultry breeding. *Journal of Animal Breeding and Genetics*, 133(1), 3-12. <https://doi.org/10.1111/jbg.12178>
- Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, 141(4), 1619-1632. <https://doi.org/10.1093/genetics/141.4.1619>
- Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetics Research*, 74(3), 329– 340. <https://doi.org/10.1017/S0016672399004152>
- Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS genetics*, 2(4), e64. <https://doi.org/10.1371/journal.pgen.0020064>
- Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), 783–796. <https://doi.org/10.1038/nrg2664>
- Chen, N., Luo, X., Lu, C., Ke, C., & You, W. (2017). Effects of artificial selection practices on loss of genetic diversity in the Pacific abalone, *Haliotis discus hannai*. *Aquaculture Research*, 48(9), 4923-4933. <https://doi.org/10.1111/are.13311>
- Chavanne, H., Janssen, K., Hofherr, J., Contini, F., Haffray, P., Komen, H., ... & Bargelloni, L. (2016). A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquaculture international*, 24(5), 1287-1307. <https://doi.org/10.1007/s10499-016-9985-0>
- Chevassus, B., Petit, J., & Escaffre, A. M. (1975). Hybridation artificielle entre la truite arc-en-ciel (salmo gairdneri richardson) et le saumon coho (oncorhynchus kisutch walbaum). In *Annales de génétique et de sélection animale* (Vol. 7, No. 1, pp. 1-11). EDP Sciences. https://www.gse-journal.org/articles/gse/pdf/1975/01/AGSE_0003-4002_1975_7_1_ART0001.pdf
- Chevassus, B., Quillet, E., Krieg, F., Hollebecq, M. G., Mambrini, M., Fauré, A., ... & Vandeputte, M. (2004). Enhanced individual selection for selecting fast growing fish: the "PROSPER" method, with application on brown trout (*Salmo trutta fario*). *Genetics Selection Evolution*, 36(6), 1-19. <https://doi.org/10.1186/1297-9686-36-6-643>
- Colleau, J. J., Quéméré, P., Larroque, H., Sergent, J., & Wagner, C. (2002). Gestion génétique de la race bovine Bretonne Pie-Noire: bilan et perspectives. <https://doi.org/10.20870/productions-animales.2002.15.3.3703>

REFERENCES

- Colleau, J. J. J., & Moureaux, S. (2006). Gestion optimisée de la parenté et de la consanguinité dans les programmes de sélection des bovins laitiers. *Productions animales*, 19(1), 3-14. <https://hal.inrae.fr/hal-02661752/document>
- Colleau, J. J., Tual, K., de Preaumont, H., & Regaldo, D. (2009). A mating method accounting for inbreeding and multi-trait selection in dairy cattle populations. *Genetics Selection Evolution*, 41(1), 1-10. <https://doi.org/10.1186/1297-9686-41-7>
- Colon F. (2022). La préservation de la diversité génétique et l'amélioration de l'efficacité alimentaire : Deux grands enjeux pour une truiticulture plus durable. Mémoire de fin d'étude ingénieur AgroParisTech, dominante : Elevage et filières Durables et iNnovants (EDEN)
- Copp, G. H., Bianco, P. G., Bogutskaya, N. G., Erős, T., Falka, I., Ferreira, M. T., ... & Wiesner, C. (2005). To be, or not to be, a non-native freshwater fish?. *Journal of Applied Ichthyology*, 21(4), 242-262. <https://doi.org/10.1111/j.1439-0426.2005.00690.x>
- Crawford, S. S., & Muir, A. M. (2008). Global introductions of salmon and trout in the genus *Oncorhynchus*: 1870–2007. *Reviews in Fish Biology and Fisheries*, 18(3), 313-344. <https://doi.org/10.1007/s11160-007-9079-1>
- Croquet, C., Mayeres, P., Gillon, A., Vanderick, S., et Gengler, N. (2006). Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89, 2257– 2267. [https://doi.org/10.3168/jds.S0022-0302\(06\)72297-4](https://doi.org/10.3168/jds.S0022-0302(06)72297-4)
- Croquet, C., Mayeres, P., Gillon, A., Hammami, H., Soyeurt, H., Vanderick, S., & Gengler, N. (2007). Linear and curvilinear effects of inbreeding on production traits for Walloon Holstein cows. *Journal of Dairy Science*, 90(1), 465-471. [https://doi.org/10.3168/jds.S0022-0302\(07\)72648-6](https://doi.org/10.3168/jds.S0022-0302(07)72648-6)
- Cruden, D. (1949). The computation of inbreeding coefficients: for closed populations. *Journal of heredity*, 40(9), 248-251. <https://doi.org/10.1093/oxfordjournals.jhered.a106039>
- Curik, I., Ferenčaković, M., & Sölkner, J. (2017). Genomic dissection of inbreeding depression: a gate to new opportunities. *Revista Brasileira de Zootecnia*, 46, 773-782. <https://doi.org/10.1590/S1806-92902017000900010>
- D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., ... & Dupont-Nivet, M. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51(1), 1-15. <https://doi.org/10.1186/s12711-019-0468-4>
- D'Ambrosio, J., Morvezzen, R., Brard-Fudulea, S., Bestin, A., Acin Perez, A., Guéméné, D., ... & Phocas, F. (2020). Genetic architecture and genomic selection of female reproduction traits in rainbow trout. *BMC genomics*, 21(1), 1-14. <https://doi.org/10.1186/s12864-020-06955-7>
- D'Ambrosio, J. (2020). Intérêt et optimisation de la sélection génomique chez la truite arc-en-ciel (Doctoral dissertation, Université Paris-Saclay). <https://pastel.archives-ouvertes.fr/tel-03209220/document>
- D'Arnoldi, C. T., Foulley, J. L., & Ollivier, L. (1998). An overview of the Weitzman approach to diversity. *Genetics Selection Evolution*, 30(2), 149-161. https://www.gse-journal.org/articles/gse/pdf/1998/02/GSE_0999-193X_1998_30_2_ART0004.pdf
- Danchin-Burge, C., Moureaux, S., Boichard, D., Baur, A. et Fritz, S. (2016). Un indicateur innovant de variabilité génétique à partir des données moléculaires. Dans Congrès Mondial de la race Brune, Mende, France.
- Danchin-Burge, C. (2020) Indicateurs de variabilité génétique – race bovines – Edition 2020.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Darwin., C. (1859). On the origin of species. *published on*, 24, 1.

- Darwin, C. (1868). The variation of animals and plants under domestication. Volume 2 (London: J. Murray).
- Davenport, C. B. (1908). Degeneration, albinism and inbreeding. *Science*, 28(718), 454-455.
- De Beukelaer, H., Badke, Y., Fack, V., & De Meyer, G. (2017). Moving beyond managing realized genomic relationship in long-term genomic selection. *Genetics*, 206(2), 1127-1138. <https://doi.org/10.1534/genetics.116.194449>
- De Cara, M. Á. R., Villanueva, B., Toro, M. Á., & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular ecology*, 22(24), 6091-6099. <https://doi.org/10.1111/mec.12560>
- Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS biology*, 3(10), e314. <https://doi.org/10.1371/journal.pbio.0030314>
- De Meeûs, T. (2012). Initiation à la génétique des populations naturelles. IRD éditions.
- DeRose, M. A., & Roff, D. A. (1999). A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution*, 53(4), 1288-1292. <https://doi.org/10.1111/j.1558-5646.1999.tb04541.x>
- Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J., & Windig, J. J. (2018). Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. *Genetics Selection Evolution*, 50(1), 1-16. <https://doi.org/10.1186/s12711-018-0385-y>
- Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., & Windig, J. J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genetics Selection Evolution*, 51(1), 1-16. <https://doi.org/10.1186/s12711-019-0497-z>
- Doekes, H. P., Bijma, P., Veerkamp, R. F., de Jong, G., Wientjes, Y. C., & Windig, J. J. (2020). Inbreeding depression across the genome of Dutch Holstein Friesian dairy cattle. *Genetics Selection Evolution*, 52(1), 1-18. <https://doi.org/10.1186/s12711-020-00583-1>
- Doekes, H. P., Bijma, P., & Windig, J. J. (2021). How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes*, 12(6), 926. <https://doi.org/10.3390/genes12060926>
- Dollar, A. M., & Katz, M. (1964). Rainbow trout brood stocks and strains in American hatcheries as factors in the occurrence of hepatoma. *The Progressive Fish-Culturist*, 26(4), 167-174. [https://doi.org/10.1577/1548-8640\(1964\)26\[167:RTBSAS\]2.0.CO;2](https://doi.org/10.1577/1548-8640(1964)26[167:RTBSAS]2.0.CO;2)
- Doublet, A. C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., ... & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, 51(1), 1-13. <https://doi.org/10.1186/s12711-019-0495-1>
- Doublet, A. C. (2020). *La diversité génétique à l'ère de la génomique: évolution de la consanguinité et ses conséquences dans trois races bovines laitières françaises* (Doctoral dissertation, université Paris-Saclay). <https://pastel.archives-ouvertes.fr/tel-03181150/document>
- Doux fils, J., Mathieu, C., Mandiki, S. N., Milla, S., Henrotte, E., Wang, N., ... & Kestemont, P. (2011). Physiological and proteomic evidences that domestication process differentially modulates the immune status of juvenile Eurasian perch (*Perca fluviatilis*) under chronic confinement stress. *Fish & shellfish immunology*, 31(6), 1113-1121. <https://doi.org/10.1016/j.fsi.2011.10.001>
- Druet, T., & Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular ecology*, 26(20), 5820-5841. <https://doi.org/10.1111/mec.14324>
- Duenk, P., Calus, M. P., Wientjes, Y. C., & Bijma, P. (2017). Benefits of dominance over additive models for the estimation of average effects in the presence of dominance. *G3: Genes, Genomes, Genetics*, 7(10), 3405-3414. <https://doi.org/10.1534/g3.117.300113>

REFERENCES

- Dumas, J. (1976). Dynamique et sédentarité d'une population naturalisée de truite arc-en-ciel (*Salmo gairdneri* Richardson) dans un ruisseau de montagne l'Estibère (Hautes-Pyrénées). *Ann. Hydrobiol.*, 7: 115–139
- Dumont, B., González-García, E., Thomas, M., Fortun-Lamothe, L., Ducrot, C., Dourmad, J. Y., & Tichit, M. (2014). Forty research issues for the redesign of animal production systems in the 21st century. *Animal*, 8(8), 1382–1393. <https://doi.org/10.1017/S1751731114001281>
- Dupont-Nivet, M., Vandeputte, M., & Chevassus, B. (2002). Optimization of factorial mating designs for inference on heritability in fish species. *Aquaculture*, 204(3-4), 361-370. [https://doi.org/10.1016/S0044-8486\(01\)00839-0](https://doi.org/10.1016/S0044-8486(01)00839-0)
- Dupont-Nivet, M., Vandeputte, M., Haffray, P., & Chevassus, B. (2006). Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture*, 252(2-4), 161-170. <https://doi.org/10.1016/j.aquaculture.2005.07.005>
- Dupont-Nivet, M., Karahan-Nomm, B., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., ... & Vandeputte, M. (2010). Genotype by environment interactions for growth in European seabass (*Dicentrarchus labrax*) are large when growth rate rather than weight is considered. *Aquaculture*, 306(1-4), 365-368. <https://doi.org/10.1016/j.aquaculture.2010.05.025>
- Durland, E., De Wit, P., & Langdon, C. (2021). Temporally balanced selection during development of larval Pacific oysters (*Crassostrea gigas*) inherently preserves genetic diversity within offspring. *Proceedings of the Royal Society B*, 288(1958), 20203223. <https://doi.org/10.1098/rspb.2020.3223>
- East, E. M. (1908). Inbreeding in corn. *Rep. Conn. Agric. Exp. Stn, 1907*, 419-428.
- East, E. M. (1918). The role of reproduction in evolution. *The American Naturalist*, 52(618/619), 273-289. <https://www.journals.uchicago.edu/doi/pdf/10.1086/279670>
- Edmands, Suzanne. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular ecology* 16.3 : 463-475. <https://doi.org/10.1111/j.1365-294X.2006.03148.x>
- Ehrenbaum, E. (1926). Ueber regenbogenforellen und Steelheadforellen. *Allg. Fisch.-Ztg.*, 51: 288–292. <https://doi.org/10.1080/23308249.2015.1024825>
- Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422-433. <https://doi.org/10.1038/nrg.2016.58>
- Engelsma, K. A., Veerkamp, R. F., Calus, M. P. L., & Windig, J. J. (2011). Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. *Journal of Animal Breeding and Genetics*, 128(6), 473-481. <https://doi.org/10.1111/j.1439-0388.2011.00936.x>
- Escaffre, A. M., & Bergot, P. (1984). Utilization of the yolk in rainbow trout alevins (*Salmo gairdneri* Richardson): effect of egg size. *Reproduction Nutrition Développement*, 24(4), 449-460. https://rnd.edpsciences.org/articles/rnd/pdf/1984/04/RND_0181-1916_1984_24_4_ART0010.pdf
- Esfandyari, H., Bijma, P., Henryon, M., Christensen, O. F., & Sørensen, A. C. (2016). Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genetics Selection Evolution*, 48(1), 1-9. <https://doi.org/10.1186/s12711-016-0220-2>
- Estoup, A., Presa, P., Krieg, F., Vaiman, D., & Guyomard, R. (1993). (CT)_n and (GT)_n microsatellites: a new class of genetic markers for *Salmo trutta* L.(brown trout). *Heredity*, 71(5), 488-496. <https://doi.org/10.1038/hdy.1993.167>
- Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics*, 1, <https://doi.org/10.1177/1177693430500100003>
- Falconer, D.S., Mackay, T.F.C. (1996). Introduction to Quantitative Genetics, Addison We. Ed

- FAO. (2003). Aquaculture production 2001. FAO Yearbook of Fisheries Statistics, 92/2.
- FAO. (2009). Développement de l'aquaculture 3. Gestion des ressources génétiques
- FAO. (2013). In vivo conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. Rome. <https://www.fao.org/3/i3327e/i3327e.pdf>.
- FAO. (2022). La situation mondiale des pêches et de l'aquaculture en 2022. <https://www.fao.org/3/cc0461fr/online/sofia/2022/aquaculture-production.html>
- FAO. (2022). *Oncorhynchus mykiss*. Programme d'Information sur les espèces aquatiques cultivées. Text by Cowx, I. G.. Fisheries and Aquaculture Division [online]. Rome. https://www.fao.org/fishery/en/culturedspecies/oncorhynchus_mykiss_fr?lang=fr
- Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3), 929-941. <https://doi.org/10.1534/genetics.112.147231>
- Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, 155(3), 1405-1413. <https://doi.org/10.1093/genetics/155.3.1405>
- Fernández, J., Villanueva, B., & Toro, M. A. (2021). Optimum mating designs for exploiting dominance in genomic selection schemes for aquaculture species. *Genetics Selection Evolution*, 53(1), 1-13. <https://doi.org/10.1186/s12711-021-00610-9>
- Fishback, A. G., Danzmann, R. G., Ferguson, M. M., & Gibson, J. P. (2002). Estimates of genetic parameters and genotype by environment interactions for growth traits of rainbow trout (*Oncorhynchus mykiss*) as inferred using molecular pedigrees. *Aquaculture*, 206(3-4), 137-150. [https://doi.org/10.1016/S0044-8486\(01\)00707-4](https://doi.org/10.1016/S0044-8486(01)00707-4)
- Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2), 399-433. <https://doi.org/10.1017/S0080456800012163>
- Fisher, R. A. (1930). The genetical theory of natural selection. Clarendon.
- Fisher, R. A. (1958). Polymorphism and natural selection. *Journal of Ecology*, 46(2), 289-293. <https://doi.org/10.2307/2257396>
- Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary applications*, 11(7), 1035-1052. <https://doi.org/10.1111/eva.12569>
- Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180, 977-993. <https://doi.org/10.1534/genetics.108.092221>
- Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F. S., & Sargolzaei, M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC genomics*, 19(1), 1-12. <https://doi.org/10.1186/s12864-018-4453-z>
- Forstmeier, W., Schielzeth, H., Mueller, J. C., Ellegren, H., & Kempnaers, B. (2012). Heterozygosity-fitness correlations in zebra finches: Microsatellite markers can be better than their reputation. *Molecular Ecology*, 21, 3237-3249. <https://doi.org/10.1186/s12864-018-4453-z>
- Frankham, R., & Ralls, K. (1998). Inbreeding leads to extinction. *Nature*, 392(6675), 441-442. <https://doi.org/10.1038/33022>
- Frankham, R., Ballou, S. E. J. D., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*. Cambridge university press.

REFERENCES

- Fraslin, C., Dechamp, N., Bernard, M., Krieg, F., Hervet, C., Guyomard, R., ... & Quillet, E. (2018). Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. *Genetics Selection Evolution*, *50*(1), 1-16. <https://doi.org/10.1186/s12711-018-0431-9>
- Fraslin, C. (2018). Bases génétiques de la réponse à l'infection par *Flavobacterium psychrophilum* chez la truite arc-en-ciel: approche expérimentale et perspectives en sélection (Doctoral dissertation, Université Paris-Saclay (ComUE)). <https://tel.archives-ouvertes.fr/tel-02185038/document>
- Fraslin C., Phocas F., Bestin A., Charles M., Bernard M., Krieg F., Dechamp N., Ciobotaru C., Hozé C., Petitprez F., Milhes M., Lluch J., Bouchez O., Poncet C., Hocdé P., Haffray P., Guiguen Y., Quillet E. (2020). Genetic determinism of spontaneous masculinisation in XX female rainbow trout: new insights using medium throughput genotyping and whole-genome sequencing. *Scientific Reports* *10*:17693. <https://doi.org/10.1038/s41598-020-74757-8>
- Fritz, S., Capitan, A., Djari, A., Rodriguez, S. C., Barbat, A., Baur, A., ... & Boichard, D. (2013). Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS one*, *8*(6), e65550. <https://doi.org/10.1371/journal.pone.0065550>
- Fu, Y. B., & Ritland, K. (1994). Evidence for the partial dominance of viability genes contributing to inbreeding depression in *Mimulus guttatus*. *Genetics*, *136*(1), 323–331. <https://doi.org/10.1093/genetics/136.1.323>
- Galibert, F., Quignon, P., Hitte, C., & André, C. (2011). Toward understanding dog evolutionary and domestication history. *Comptes rendus biologiques*, *334*(3), 190-196. <https://doi.org/10.1016/j.crvi.2010.12.011>
- Gall, G. A., & Crandell, P. A. (1992). The rainbow trout. *Aquaculture*, *100*(1-3), 1-10.
- Ganteil, A., Rodríguez-Ramilo, S. T., Ligonésche, B., & Larzul, C. (2020). Genomic diversity of pigs from a Large White x Pietrain crossbreeding. <https://agris.fao.org/agris-search/search.do?recordID=FR2021035523>
- Ganteil, A., Pook, T., Rodríguez-Ramilo, S. T., Ligonésche, B., & Larzul, C. (2021). Comparison of breeding strategies for the creation of a synthetic pig line. *bioRxiv*. <https://doi.org/10.1101/2021.09.22.461330>
- Gao, G., (2016). A New and Improved Rainbow Trout (*Oncorhynchus mykiss*) Reference Genome Assembly. Presented at the International Conference on Integrative Salmonid Biology.
- Gao, G., Magadan, S., Waldbieser, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., ... & Palti, Y. (2021). A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. *G3*, *11*(4), jkab052. <https://doi.org/10.1093/g3journal/jkab052>
- Garcia, A. L., Bosworth, B., Waldbieser, G., Misztal, I., Tsuruta, S., & Lourenco, D. A. (2018). Development of genomic predictions for harvest and carcass weight in channel catfish. *Genetics Selection Evolution*, *50*(1), 1-12. <https://doi.org/10.1186/s12711-018-0435-5>
- Garcia-Baccino, C. A., Lourenco, D. A., Miller, S., Cantet, R. J., & Vitezica, Z. G. (2020). Estimating dominance genetic variances for growth traits in American Angus males using genomic models. *Journal of Animal Science*, *98*(1). <https://doi.org/10.1093/jas/skz384>
- Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplement of the R package rehh to detect positive selection from haplotype structure. *Molecular ecology resources*, *17*(1), 78-90. <https://doi.org/10.1111/1755-0998.12634>
- Gebregiwergis, G. T., Sørensen, A. C., Henryon, M., & Meuwissen, T. (2020). Controlling coancestry and thereby future inbreeding by optimum-contribution selection using alternative genomic-relationship matrices. *Frontiers in Genetics*, *11*, 345. <https://doi.org/10.3389/fgene.2020.00345>
- Gharrett, A. J., Smoker, W. W., Reisenbichler, R. R., & Taylor, S. G. (1999). Outbreeding depression in hybrids between odd-and even-broodyear pink salmon. *Aquaculture*, *173*(1-4), 117-129. [https://doi.org/10.1016/S0044-8486\(98\)00480-3](https://doi.org/10.1016/S0044-8486(98)00480-3)

- Gjerde, B., Gunnes, K., & Gjedrem, T. (1983). Effect of inbreeding on survival and growth in rainbow trout. *Aquaculture*, 34(3-4), 327-332. [https://doi.org/10.1016/0044-8486\(83\)90212-0](https://doi.org/10.1016/0044-8486(83)90212-0)
- Gjedrem, T., & Baranski, M. (2010). *Selective breeding in aquaculture: an introduction* (Vol. 10). Springer Science & Business Media.
- Glover, K. A., Otterå, H., Olsen, R. E., Slinde, E., Taranger, G. L., & Skaala, Ø. (2009). A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. *Aquaculture*, 286(3-4), 203-210. <https://doi.org/10.1016/j.aquaculture.2008.09.023>
- Gómez-Romano, F., Villanueva, B., Fernández, J., Woolliams, J. A., & Pong-Wong, R. (2016). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution*, 48(1), 1-17. <https://doi.org/10.1186/s12711-015-0172-y>
- González-Diéguéz, D., Tusell, L., Bouquet, A., & Vitezica, Z. (2019). Use of mate allocation in pig crossbreeding schemes: a simulation study.
- González-Diéguéz, D., Tusell, L., Bouquet, A., Legarra, A., & Vitezica, Z. G. (2020). Purebred and crossbred genomic evaluation and mate allocation strategies to exploit dominance in pig crossbreeding schemes. *G3: Genes, Genomes, Genetics*, 10(8), 2829-2841. <https://doi.org/10.1534/g3.120.401376>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Granier, S., Audet, C., & Bernatchez, L. (2011). Heterosis and outbreeding depression between strains of young-of-the-year brook trout (*Salvelinus fontinalis*). *Canadian Journal of Zoology*, 89(3), 190-198. <https://doi.org/10.1139/Z10-108>
- Grossi, D. A., Jafarikia, M., Brito, L. F., Buzanskas, M. E., Sargolzaei, M., & Schenkel, F. S. (2017). Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs. *BMC genetics*, 18(1), 1-13. <https://doi.org/10.1186/s12863-017-0473-y>
- Gu, L., Wu, Y., Hu, S., Chen, Q., Tan, J., Yan, Y., ... & Tang, N. (2016). Analysis of association between MAP2K4 gene polymorphism rs3826392 and IL-1b serum level in southern Chinese Han Ischemic Stroke Patients. *Journal of Stroke and Cerebrovascular Diseases*, 25(5), 1096-1101. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.12.035>
- Guyomard, R., Vandeputte, M. (2010). Chapitre 1. Génétique, in: La truite arc-en-ciel : De la biologie à l'élevage. Quae, pp. 9-33
- Guyomard, R., Boussaha, M., Krieg, F., Hervet, C., & Quillet, E. (2012). A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC genetics*, 13(1), 1-12. <https://doi.org/10.1186/1471-2156-13-15>
- Habier, D., Fernando, R. L., & Dekkers, J. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177(4), 2389-2397. <https://doi.org/10.1534/genetics.107.081190>
- Haffray, P., Enez, F., Bugeon, J., Chapuis, H., Dupont-Nivet, M., Chatain, B., & Vandeputte, M. (2018). Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. *Aquaculture*, 490, 350-354. <https://doi.org/10.1016/j.aquaculture.2018.03.003>
- Haffray, P. (2019). Amélioration d'un programme de sélection massale sur la croissance chez la truite arc-en-ciel par introduction d'une sélection BLUP pour des caractères de qualités grâce aux empreintes génétiques. Génétique animale. Agrocampus Ouest. <https://tel.archives-ouvertes.fr/tel-02145732/document>
- Haldane J. B. S. (1932). *The causes of evolution*. Harper and Brothers, London.
- Hardie, D. C., & Hebert, P. D. (2003). The nucleotypic effects of cellular DNA content in cartilaginous and

REFERENCES

- ray-finned fishes. *Genome*, 46(4), 683-706. <https://doi.org/10.1139/g03-040>
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28(706), 49-50.
- Hazel, Lanoy Nelson. (1943). The genetic basis for constructing selection indexes. *Genetics* 28.6 : 476-490. <https://doi.org/10.1093/genetics/28.6.476>
- Hedgcock, D., Siri, P., & Strong, D. R. (1994). Conservation biology of endangered Pacific salmonids: Introductory remarks. *Conservation Biology*, 8(3), 863-864. <https://doi.org/10.1046/j.1523-1739.1994.08030863.x>
- Hedgcock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, 87(4), 971-1002. <https://doi.org/10.5343/bms.2010.1051>
- Hedrick, P. W. (1994). Purging inbreeding depression and the probability of extinction: full-sib mating. *Heredity*, 73(4), 363-372. <https://doi.org/10.1038/hdy.1994.183>
- Hedrick, P. W. (2009). *Genetics of populations*. Jones & Bartlett Publishers.
- Heidaritabar, M., Wolc, A., Arango, J., Zeng, J., Settar, P., Fulton, J. E., ... & Dekkers, J. C. (2016). Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. *Journal of Animal Breeding and Genetics*, 133(5), 334-346. <https://doi.org/10.1111/jbg.12225>
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. Published by : international biometric society stable. *Biometrics*, 31(2) : 423-447. <https://doi.org/10.2307/2529430>
- Hershberger, W. K. (1992). Genetic variability in rainbow trout populations. *Aquaculture*, 100(1-3), 51-71. [https://doi.org/10.1016/0044-8486\(92\)90339-M](https://doi.org/10.1016/0044-8486(92)90339-M)
- Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium1. *Genetics Research*, 38(3), 209-216. <https://doi.org/10.1017/S0016672300020553>
- Hill, W. G., & Weir, B. S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, 93(1), 47– 64. <https://doi.org/10.1017/S0016672310000480>
- Honsing-Erleburg, (2005). W. Zum Einfluß der Regenbogenforelle und des Bachsaiblings auf Bachforellenpopulationen. Österreichs Fischerei, 58: 286–289.
- Howard, J. T., Pryce, J. E., Baes, C., & Maltecca, C. (2017). Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of dairy science*, 100(8), 6009-6024. <https://doi.org/10.3168/jds.2017-12787>
- Hoze, C., Fritz, S., Baur, A., Grohs, C., Danchin-Burge, C., & Boichard, D. (2018). Prise en compte des gènes d'intérêt dans les objectifs de sélection en bovins laitiers. In 24. *Rencontres autour des Recherches sur les Ruminants (3R)*. <https://hal.inrae.fr/hal-02735939/document>
- Huff, D. D., Miller, L. M., Chizinski, C. J., & Vondracek, B. (2011). Mixed-source reintroductions lead to outbreeding depression in second-generation descendants of a native North American fish. *Molecular Ecology*, 20(20), 4246-4258. <https://doi.org/10.1111/j.1365-294X.2011.05271.x>
- Huneman, P. (2011). Sélection in : Les mondes darwiniens, L'évolution de l'évolution. *Matériologiques* p 59-110
- Huxley, J. (1942). Evolution. The modern synthesis. *Evolution. The Modern Synthesis*.
- Jain, S. K., & Allard, R. W. (1966). The effects of linkage, epistasis, and inbreeding on population changes under selection. *Genetics*, 53(4), 633. <https://doi.org/10.1093/genetics/53.4.633>

- Jinadasa, J., Kothalawala, A. B., & Herath, H. M. B. K. (2005). Status of rainbow trout (*Oncorhynchus mykiss*-Walbaum 1792) population after the cessation of stocking in waterways of Horton Plains, Sri Lanka. <http://dr.lib.sjp.ac.lk/handle/123456789/1041>
- Kardos, M., Luikart, G., & Allendorf, F. W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity*, *115*(1), 63-72. <https://doi.org/10.1038/hdy.2015.17>
- Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary applications*, *9*(10), 1205-1218. <https://doi.org/10.1111/eva.12414>
- Kaspar, V., Vandeputte, M., Kohlmann, K., Hulak, M., Rodina, M., Gela, D., ... & Linhart, O. (2008). A proposal and case study towards a conceptual approach of validating sperm competition in common carp (*Cyprinus carpio* L.), with practical implications for hatchery procedures. *Journal of Applied Ichthyology*, *24*(4), 406-409. <https://doi.org/10.1111/j.1439-0426.2008.01148.x>
- Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, *189*(1), 237-249. <https://doi.org/10.1093/genetics/190.1.283>
- Kendall, N. W., McMillan, J. R., Sloat, M. R., Buehrens, T. W., Quinn, T. P., Pess, G. R., ... & Zabel, R. W. (2015). Anadromy and residency in steelhead and rainbow trout (*Oncorhynchus mykiss*): a review of the processes and patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(3), 319-342. <https://doi.org/10.1139/cjfas-2014-0192>
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624-626. <https://authors.library.caltech.edu/5456/1/hrst.mit.edu/hrs/evolution/public/papers/kimura1968/kimura1968.pdf>
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.
- Kincaid, H. L. (1976). Effects of inbreeding on rainbow trout populations. *Transactions of the American Fisheries Society*, *105*(2), 273– 280. [https://doi.org/10.1577/1548-8659\(1976\)105<273:EOIORT>2.0.CO;2](https://doi.org/10.1577/1548-8659(1976)105<273:EOIORT>2.0.CO;2)
- Kincaid, H. L. (1983). Inbreeding in fish populations used for aquaculture. *Aquaculture*, *33*(1-4), 215-227. [https://doi.org/10.1016/0044-8486\(83\)90402-7](https://doi.org/10.1016/0044-8486(83)90402-7)
- Kristensen, T. N., & Sørensen, A. C. (2005). Inbreeding—lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science*, *80*(2), 121–133. <https://doi.org/10.1079/ASC41960121>
- Kristensen, T.N., Pedersen, K.S., Vermeulen, C.J., et Loeschcke, V. (2010). Research on inbreeding in the “omic” era. *Trends Ecol. Evol.* *25*, 44–52. <https://doi.org/10.1016/j.tree.2009.06.014>
- Lagarde, H., Phocas, F., Pouil, S., Goardon, L., Bideau, M., Guyvarc'h, F., ... & Lallias, D. (2023). Are resistances to acute hyperthermia or hypoxia stress similar and consistent between early and late ages in rainbow trout using isogenic lines?. *Aquaculture*, 738800. <https://doi.org/10.1016/j.aquaculture.2022.738800>
- Lallias, D., Bernard, M., Ciobotaru, C., Dechamp, N., Labbé, L., Goardon, L., ... & Dupont-Nivet, M. (2021). Sources of variation of DNA methylation in rainbow trout: combined effects of temperature and genetic background. *Epigenetics*, *16*(9), 1031-1052. <https://doi.org/10.1080/15592294.2020.1834924>
- Lande, R., & Schemske, D. W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution*, *39*(1), 24-40. <https://doi.org/10.1111/j.1558-5646.1985.tb04077.x>
- Larson, G., & Fuller, D. Q. (2014). The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 115-136. <https://palaeobarn.com/sites/domestication.org.uk/files/downloads/112.pdf>
- Leach, G. C. (1924). Propagation and distribution of food fishes, fiscal year 1923, Appendix XIII. In: *Report of the United States Commissioner of Fisheries for the fiscal year 1923 with Appendixes*. Washington: Taylor &

REFERENCES

Francis.

Lefort-Buson, M., & Damerval, C. (1988). Les outils d'évaluation de la diversité génétique et phénotypique. *Agronomie*, 8(3), 173-178. <https://hal.inrae.fr/hal-02727075/document>

Leroy, G. (2014). Inbreeding depression in livestock species: review and meta-analysis. *Animal genetics*, 45(5), 618-628. <https://doi.org/10.1111/age.12178>

Leutenegger, A. L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., & Thompson, E. A. (2003). Estimation of the inbreeding coefficient through use of genomic data. *The American Journal of Human Genetics*, 73(3), 516-523. <https://doi.org/10.1086/378207>

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87(836), 331-333. <https://doi.org/10.1086/281792>

Levins, R., & Lewontin, R. (1985). *The dialectical biologist*. Harvard University Press.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74(1), 175-195. <https://doi.org/10.1093/genetics/74.1.175>

Li, Y., Yang, Y., Zheng, W., & Cheng, J. (2019). Genetic parameters and genotype by environment interactions for growth traits and survival of olive flounder (*Paralichthys olivaceus*) in recirculating aquaculture system and flow-through system. *Aquaculture*, 510, 56-60. <https://doi.org/10.1016/j.aquaculture.2019.05.043>

Lozada-Soto, E. A., Tiezzi, F., Cole, J. B., VanRaden, P. M., & Maltecca, C. (2022). Patterns of inbreeding and selection using runs of homozygosity in North American dairy cattle. In *World Congress on Genetics Applied to Livestock Production (WCGALP 2022)*. https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/20_006.pdf

Lukhtanov, V. A., Dincă, V., Talavera, G., & Vila, R. (2011). Unprecedented within-species chromosome number cline in the Wood White butterfly *Leptidea sinapis* and its significance for karyotype evolution and speciation. *BMC evolutionary biology*, 11(1), 1-11. <https://doi.org/10.1186/1471-2148-11-109>

Lush, J.L. (1937). *Animal breeding plans*. Iowa State University press

Luo, L. J., Z.-K. Li, H. W. Mei, Q. Y. Shu, R. Tabien, D. B. Zhong, C. S. Ying, J. W. Stansel, G. S. Khush, and A. H. Paterson. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. *Genetics* 158:1755–1771. <https://doi.org/10.1093/genetics/158.4.1755>

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1, pp. 535-557). Sunderland, MA: Sinauer. https://www.invemar.org.co/redcostera1/invemar/docs/RinconLiterario/2011/febrero/AG_8.pdf

MacLean, C. (2014). Hapbin: An Efficient Method for Calculating the EHH and iHS of Genetic Datasets. <https://static.epcc.ed.ac.uk/dissertations/hpc-msc/2013-2014/Hapbin:%20An%20Efficient%20Method%20for%20Calculating%20EHH%20and%20iHS%20of%20Genetic%20Datasets.pdf>

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778), 20132881. <https://doi.org/10.1098/rspb.2013.2881>

Makanjuola, B. O., Maltecca, C., Miglior, F., Schenkel, F. S., & Baes, C. F. (2020b). Effect of recent and ancient inbreeding on production and fertility traits in Canadian Holsteins. *BMC Genomics*, 21(1), 1–15. <https://doi.org/10.1186/s12864-020-07031-w>

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., & Baes, C. F. (2020a). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of dairy science*, 103(6), 5183-5199. <https://doi.org/10.3168/jds.2019-18013>

- Malhado, C. H. M., Malhado, A. C. M., Carneiro, P. L. S., Ramos, A. A., Carrillo, J. A., & Pala, A. (2013). Inbreeding depression on production and reproduction traits of buffaloes from Brazil. *Animal Science Journal*, 84(4), 289-295. <https://doi.org/10.1111/asj.12006>
- Mandiki, S. N. M., Milla, S., Wang, N., Blanchard, G., Djonkack, T., Tanascaux, S., & Kestemont, P. (2011). Effects of probiotic bacteria on growth parameters and immune defence in Eurasian perch *Perca fluviatilis* L. larvae under intensive culture conditions. *Aquaculture Research*, 42(5), 693-703. <https://doi.org/10.1111/j.1365-2109.2010.02744.x>
- Marchinko, K. B., Matthews, B., Arnegard, M. E., Rogers, S. M., & Schluter, D. (2014). Maintenance of a genetic polymorphism with disruptive natural selection in stickleback. *Current Biology*, 24(11), 1289-1292. <https://doi.org/10.1016/j.cub.2014.04.026>
- Marras G, Rossoni A, Schwarzenbacher H, Biffani S, Biscarini F, Nicolazzi EL. (2017). Zanardi: an open-source pipeline for multiple-species genomic analysis of SNP array data. *Anim Genet*. 48:121. https://www.researchgate.net/profile/Stefano-Biffani/publication/307915049_zanardi_An_open-source_pipeline_for_multiple-species_genomic_analysis_of_SNP_array_data/links/5b85a5d3299bf1d5a72e9f39/zanardi-An-open-source-pipeline-for-multiple-species-genomic-analysis-of-SNP-array-data.pdf
- Martikainen, K., Tyrisevä, A. M., Matilainen, K., Pösö, J., & Uimari, P. (2017). Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. *Journal of Animal Breeding and Genetics*, 134(5), 383-392. <https://doi.org/10.1111/jbg.12285>
- Martikainen, K., Sironen, A., & Uimari, P. (2018). Estimation of intrachromosomal inbreeding depression on female fertility using runs of homozygosity in Finnish Ayrshire cattle. *Journal of dairy science*, 101(12), 11097-11107. <https://doi.org/10.3168/jds.2018-14805>
- Martinez, V., Dettleff, P. J., Galarce, N., Bravo, C., Dorner, J., Iwamoto, R. N., & Naish, K. (2022). Estimates of Effective Population Size in Commercial and Hatchery Strains of Coho Salmon (*Oncorhynchus kisutch* (Walbaum, 1792)). *Animals*, 12(5), 647. <https://doi.org/10.3390/ani12050647>
- Martinez-Castillero, M., Varona, L., Pegolo, S., Rossoni, A., & Cecchinato, A. (2021). Bayesian inference of the inbreeding load variance for fertility traits in Brown Swiss cattle. *Journal of Dairy Science*, 104(9), 10040-10048. <https://doi.org/10.3168/jds.2020-20087>
- Masel, J. (2011). Genetic drift. *Current Biology*, 21(20), R837-R838. [https://www.cell.com/current-biology/pdf/S0960-9822\(11\)00882-7.pdf](https://www.cell.com/current-biology/pdf/S0960-9822(11)00882-7.pdf)
- Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., & Portolano, B. (2017a). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution*, 49(1), 1-10. <https://doi.org/10.1186/s12711-017-0360-z>
- Mastrangelo, S., Portolano, B., Di Gerlando, R., Ciampolini, R., Tolone, M., Sardina, M. T., & International Sheep Genomics Consortium. (2017b). Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. *Animal*, 11(7), 1107-1116. <https://doi.org/10.1017/S1751731116002780>
- Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., Portolano, B., & Bi. Ov. Ita Consortium. (2018). Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. *Animal genetics*, 49(1), 71-81. <https://doi.org/10.1111/age.12634>
- Mather, F. (1886). Account of eggs repacked at Cold Spring Harbor N. Y., and shipped to foreign countries, under the Direction of the U. S. Fish Commission, during the Winter of 1884-1885, pp. 147-151. In: *Report of the Commissioner for 1884*. Washington: Taylor & Francis.
- Mayr, E. (1954). Change of genetic environment and evolution. In : Huxley J. (ed.), *Evolution a process*, 157-180, Allen et Unwin, London. https://www2.hawaii.edu/~khayes/Journal_Club/fall2006/Mayr_1954_in_Evol_as_a_Proc_157-180.pdf

REFERENCES

- Mayr, E. (1963). *Animal species and evolution*. 797 p., Harvard University Press, Cambridge, Mass. <https://doi.org/10.4159/harvard.9780674865327>
- McDonald, M. (1893). Report of the United States Commissioner of Fish and Fisheries for the fiscal years 1889–90 and 1890–91, pp. 1–96. In: *Report of the Commissioner for 1889 to 1891*. Washington: Taylor & Francis
- McPhail, J. D. (2007). *The freshwater fishes of British Columbia*. University of Alberta.
- McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., ... & Wilson, J. F. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83(3), 359-372. <https://doi.org/10.1016/j.ajhg.2008.08.007>
- Mendel, G. (1965). *Experiments in plant hybridisation*. Harvard University Press.
- Metian, M., Troell, M., Christensen, V., Steenbeek, J., & Pouil, S. (2020). Mapping diversity of species in global aquaculture. *Reviews in Aquaculture*, 12(2), 1090-1100. <https://doi.org/10.1111/raq.12374>
- Meuwissen, T. H. E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of animal science*, 75(4), 934-940. <https://doi.org/10.2527/1997.754934x>
- Meuwissen, T. H., & Goddard, M. E. (2001). Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution*, 33(6), 1-30. <https://doi.org/10.1186/1297-9686-33-6-605>
- Meuwissen, T. (2009). Towards consensus on how to measure neutral genetic diversity?. *Journal of Animal Breeding and Genetics*, 126(5), 333-334. <https://doi.org/10.1111/j.1439-0388.2009.00839.x>
- Meuwissen, T. H. E., Luan, T., & Woolliams, J. A. (2011). The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *Journal of Animal Breeding and Genetics*, 128(6), 429-439. <https://doi.org/10.1111/j.1439-0388.2011.00966.x>
- Meuwissen, T. H., Sonesson, A. K., Gebregiwerigis, G., & Woolliams, J. A. (2020). Management of genetic diversity in the era of genomics. *Frontiers in genetics*, 11, 880. <https://doi.org/10.3389/fgene.2020.00880>
- Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC genomics*, 21(1), 1-14. <https://doi.org/10.1186/s12864-020-6463-x>
- Michot, P. (2017). Utilisation des séquences de génome complet pour l'identification de mutations délétères responsables d'anomalies génétiques récessives chez le bovin (Doctoral dissertation, Université Paris-Saclay (ComUE)). <https://tel.archives-ouvertes.fr/tel-02068853/document>
- Milla, S., Pasquet, A., El Mohajer, L., & Fontaine, P. (2021). How domestication alters fish phenotypes. *Reviews in Aquaculture*, 13(1), 388-405. <https://doi.org/10.1111/raq.12480>
- Minvielle, F. (1992). Les modèles à un et deux locus. *INRA Productions Animales*, (hs), 61-67. <https://hal.archives-ouvertes.fr/hal-00895998/document>
- Misztal, I., Lawlor, T. J., & Gengler, N. (1997). Relationships among estimates of inbreeding depression, dominance and additive variance for linear traits in Holsteins. *Genetics Selection Evolution*, 29(3), 319-326. https://www.gse-journal.org/articles/gse/pdf/1997/03/GSE_0999-193X_1997_29_3_ART0002.pdf
- Moghaddar, N., & Van der Werf, J. H. J. (2017). Genomic estimation of additive and dominance effects and impact of accounting for dominance on accuracy of genomic evaluation in sheep populations. *Journal of Animal Breeding and Genetics*, 134(6), 453-462. <https://doi.org/10.1111/jbg.12287>
- Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., & Quinn, T. P. (2013). Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. *Molecular Ecology*, 22(5), 1295–1309. <https://doi.org/10.1111/mec.12185>

- Neff, B. D. (2004). Stabilizing selection on genomic divergence in a wild fish population. *Proceedings of the National Academy of Sciences*, 101(8), 2381-2385. <https://doi.org/10.1073/pnas.0307522100>
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the national academy of sciences*, 70(12), 3321-3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Nei, M., & Maruyama, T. (1975). Lewontin-Krakauertest for neutral genes. *Genetics*, 80(2), 395. <https://doi.org/10.1093/genetics/80.2.395>
- Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of human genetics*, 47(3), 253-259. <https://doi.org/10.1111/j.1469-1809.1983.tb00993.x>
- Nejati-Javaremi, A., Smith, C., & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science*, 75(7), 1738-1745. <https://doi.org/10.2527/1997.7571738x>
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome research*, 15(11), 1566-1575. <https://doi.org/10.1101/gr.4252305>
- Nielsen, H. M., Amer, P. R., & Byrne, T. J. (2014). Approaches to formulating practical breeding objectives for animal production systems. *Acta Agriculturae Scandinavica, Section A—Animal Science*, 64(1), 2-12. <https://doi.org/10.1080/09064702.2013.827237>
- Nosrati, M., Nanaei, H. A., Javanmard, A., & Esmailzadeh, A. (2021). The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. *Genomics*, 113(3), 1407-1415. <https://doi.org/10.1016/j.ygeno.2021.03.005>
- Ohno, S. (1970). *Evolution by gene duplication*. Springer Science & Business Media.
- Okazaki, T. (1984). Genetic Divergence and Its Zoogeographic Implications in Closely Related Species *Salmo gairdneri* and *Salmo mykiss*. *Japanese Journal of Ichthyology*, 31(3), 297-311. <https://doi.org/10.11369/jji1950.31.297>
- O'Malley, H. (1927). Report of the Commissioner of fisheries for the fiscal year ended June 30, 1926, pp. pp. I–XLVI. In: *Report of the United States Commissioner of Fisheries for the fiscal 1926 with Appendix*. Washington: Taylor & Francis.
- Palińska-Żarska, K., Król, J., Woźny, M., Kamaszewski, M., Szudrowicz, H., Wiechetek, W., ... & Żarski, D. (2021). Domestication affected stress and immune response markers in *Perca fluviatilis* in the early larval stage. *Fish & Shellfish Immunology*, 114, 184-198. <https://doi.org/10.1016/j.fsi.2021.04.028>
- Palti, Y., Genet, C., Gao, G., Hu, Y., You, F. M., Boussaha, M., ... & Luo, M. C. (2012). A second generation integrated map of the rainbow trout (*Oncorhynchus mykiss*) genome: analysis of conserved synteny with model fish genomes. *Marine biotechnology*, 14(3), 343-357. <https://doi.org/10.1007/s10126-011-9418-z>
- Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., ... & Rexroad III, C. E. (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular ecology resources*, 14(3), 588-596. <https://doi.org/10.1111/1755-0998.12204>
- Pante, M. J. R., Gjerde, B., & McMillan, I. (2001). Effect of inbreeding on body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 192(2–4), 201–211. [https://doi.org/10.1016/S0044-8486\(00\)00467-1](https://doi.org/10.1016/S0044-8486(00)00467-1)
- Parker, G.A., (1998). 1 - Sperm Competition and the Evolution of Ejaculates: Towards a Theory Base, in: Birkhead, T.R., Møller, A.P. (Eds.), *Sperm Competition and Sexual Selection*. Academic Press, San Diego, pp. 3–54. <https://doi.org/10.1016/B978-012100543-6/50026-X>
- Pasquet, A. (2018). Effects of domestication on fish behaviour. In *Animal domestication*. IntechOpen. <https://doi.org/10.5772/intechopen.78752>

REFERENCES

- Pavlidis P, Živković D, Stamatakis A, Alachiotis N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol.* 30:2224–34. <https://doi.org/10.1093/molbev/mst112>
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., ... & Wegmann, M. (2013). Essential biodiversity variables. *Science*, 339(6117), 277-278. <https://doi.org/10.1126/science.1229931>
- Pilling, D., & Hoffmann, I. (2011). Climate change and animal genetic resources for food and agriculture: state of knowledge, risks and opportunities. *FAO CGRFA Background Study Paper*, (53), 28.
- Phocas, F. (2011). L'optimisation des programmes de sélection. Numéro spécial : Amélioration génétique 341-356, (24), no4. <https://doi.org/10.20870/productions-animales.2011.24.4.3266>.
- Phocas, F., Belloc, C., Bidanel, J., Delaby, L., Dourmad, J. Y., Dumont, B., ... & Brochard, M. (2017). Quels programmes d'amélioration génétique des animaux pour des systèmes d'élevage agro-écologiques ?. <https://doi.org/10.20870/productions-animales.2017.30.1.2232>
- Phocas, F. (2020). La définition des objectifs de sélection. *Génétique des Animaux d'élevage. Diversité et adaptation dans un monde changeant.*, Quae. <https://hal.inrae.fr/hal-03153836/document>
- Price, E. O. (1984). Behavioral aspects of animal domestication. *The quarterly review of biology*, 59(1), 1-32. <https://doi.org/10.1086/413673>
- Price, E. O. (1999). Behavioral development in animals undergoing domestication. *Applied Animal Behaviour Science*, 65(3), 245-271. [https://doi.org/10.1016/S0168-1591\(99\)00087-8](https://doi.org/10.1016/S0168-1591(99)00087-8)
- Pryce, J. E., Haile-Mariam, M., Goddard, M. E., & Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution*, 46(1), 1-14. <https://doi.org/10.1186/s12711-014-0071-7>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>
- Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC genetics*, 13(1), 1-11. <https://doi.org/10.1186/1471-2156-13-70>
- Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock science*, 166, 133-143. <https://doi.org/10.1016/j.livsci.2014.05.003>
- Quillet, E., Aubard, G., & Queau, I. (2002). Mutation in a sex-determining gene in rainbow trout: detection and genetic analysis. *Journal of Heredity*, 93(2), 91-99. <https://doi.org/10.1093/jhered/93.2.91>
- Ravenel, W. C. (1899). Report on the propagation and distribution of food-fishes, pp. XXXI–CXXII. In: *Report of the Commissioner for the year ending June 30, 1898*. Washington: Taylor & Francis. https://penbay.org/cof/COF_1895_3.pdf
- Restoux, G., Rognon, X., Vieaud, A., Guemene, D., Petitjean, F., Rouger, R., ... & Tixier-Boichard, M. (2022). Managing genetic diversity in breeding programs of small populations: the case of French local chicken breeds. *Genetics Selection Evolution*, 54(1), 1-17. <https://doi.org/10.1186/s12711-022-00746-2>
- Rischkowsky, B., & Pilling, D. (2007). *The state of the world's animal genetic resources for food and agriculture*. Food & Agriculture Org.
- Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research*, 67(2), 175-185. <https://doi.org/10.1017/S0016672300033620>
- Riva Rossi, C. M., Lessa, E. P., & Pascual, M. A. (2004). The origin of introduced rainbow trout (*Oncorhynchus mykiss*) in the Santa Cruz River, Patagonia, Argentina, as inferred from mitochondrial DNA. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(7), 1095-1101. <https://doi.org/10.1139/f04-056>

- Rodríguez-Ramilo, S. T., Elsen, J. M., & Legarra, A. (2019). Inbreeding and effective population size in French dairy sheep: Comparison between genomic and pedigree estimates. *Journal of dairy science*, *102*(5), 4227-4237. <https://doi.org/10.3168/jds.2018-15405>
- Roehe, R., Shrestha, N. P., Mekki, W., Baxter, E. M., Knap, P. W., Smurthwaite, K. M., ... & Edwards, S. A. (2010). Genetic parameters of piglet survival and birth weight from a two-generation crossbreeding experiment under outdoor conditions designed to disentangle direct and maternal effects. *Journal of Animal Science*, *88*(4), 1276-1285. <https://doi.org/10.2527/jas.2009-2287>
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* *34*:299-3302. <https://doi.org/10.1093/molbev/msx248>
- Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., ... & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, *464*(7288), 587-591. <https://doi.org/10.1038/nature08832>
- Rye, M. A., & Mao, I. L. (1998). Non additive genetic effects and inbreeding depression for body weight in Atlantic salmon (*Salmo salar L.*). *Livestock Production Science*, *57*(1), 15– 22. [https://doi.org/10.1016/S0301-6226\(98\)00165-1](https://doi.org/10.1016/S0301-6226(98)00165-1)
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., ... & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832-837. <https://doi.org/10.1038/nature01140>
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913-918. <https://doi.org/10.1038/nature06250>
- Sae-Lim, P., Kause, A., Mulder, H. A., Martin, K. E., Barfoot, A. J., Parsons, J. E., ... & Komen, H. (2013). Genotype-by-environment interaction of growth traits in rainbow trout (*Oncorhynchus mykiss*): a continental scale study. *Journal of Animal Science*, *91*(12), 5572-5581. <https://doi.org/10.2527/jas.2012-5949>
- Sae-Lim, P., Gjerde, B., Nielsen, H. M., Mulder, H., & Kause, A. (2016). A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Reviews in Aquaculture*, *8*(4), 369-393. <https://doi.org/10.1111/raq.12098>
- Sakai, K., Nomura, M., Takasima, F., & Oto, H., (1975). The over-ripening phenomenon of rainbow trout, 2: Changes in the percentage of eyed eggs, hatching rate and incidence of abnormal alevins during the process of over-ripening. *Bulletin of the Japanese Society of Scientific Fisheries*.
- Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science*, *241*, 104257. <https://doi.org/10.1016/j.livsci.2020.104257>
- Schiavo, G., Bovo, S., Bertolini, F., Tinarelli, S., Dall'Olio, S., Costa, L. N., ... & Fontanesi, L. (2020). Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal*, *14*(5), 910-920. <https://doi.org/10.1017/S175173111900332X>
- Schiavo, G., Bovo, S., Muñoz, M., Ribani, A., Alves, E., Araújo, J. P., ... & Fontanesi, L. (2021). Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. *Animal Genetics*, *52*(2), 155-170. <https://doi.org/10.1111/age.13045>
- Schiavo, G., Bovo, S., Ribani, A., Tinarelli, S., Cappelloni, M., Gallo, M., & Fontanesi, L. (2022) The genomic inbreeding trend over the last 25 years in three Italian pig breeds. In *World Congress on Genetics Applied to Livestock Production (WCGALP 2022)*. https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/67_008.pdf
- Scott, D., Hewitson, J., & Fraser, J. C. (1978). Origins of rainbow trout, *salmo-gairdneri richardson*, in new-

REFERENCES

zealand. *California fish and game*, 64(3), 210-218

Šegvić-Bubić, T., Grubišić, L., Trumbić, Ž., Stanić, R., Ljubković, J., Maršić-Lučić, J., & Katavić, I. (2017). Genetic characterization of wild and farmed European seabass in the Adriatic sea: assessment of farmed escapees using a Bayesian approach. *ICES Journal of Marine Science*, 74(1), 369-378. <https://doi.org/10.1093/icesjms/fsw155>

Sellier, P. (1976). The basis of crossbreeding in pigs; a review. *Livestock Production Science*, 3(3), 203-226. [https://doi.org/10.1016/0301-6226\(76\)90016-6](https://doi.org/10.1016/0301-6226(76)90016-6)

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., ... & Flury, C. (2017). Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genetics Selection Evolution*, 49(1), 1-13. <https://doi.org/10.1186/s12711-017-0358-6>

Signer-Hasler, H., Burren, A., Ammann, P., Drögemüller, C., & Flury, C. (2019). Runs of homozygosity and signatures of selection: A comparison among eight local Swiss sheep breeds. *Animal genetics*, 50(5), 512-525. <https://doi.org/10.1111/age.12828>

Signer-Hasler, H., Henkel, J., Bangerter, E., Bulut, Z., Drögemüller, C., Leeb, T., & Flury, C. (2022). Runs of homozygosity in Swiss goats reveal genetic changes associated with domestication and modern selection. *Genetics Selection Evolution*, 54(1), 1-11. <https://doi.org/10.1186/s12711-022-00695-w>

Smith, G. R., & Stearley, R. F. (1989). The classification and scientific names of rainbow and cutthroat trouts. *Fisheries*, 14(1), 4-10. [https://doi.org/10.1577/1548-8446\(1989\)014<0004:TCASNO>2.0.CO;2](https://doi.org/10.1577/1548-8446(1989)014<0004:TCASNO>2.0.CO;2)

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35. <https://doi.org/10.1017/S0016672308009579>

Sober, E. (2001). The two faces of fitness. <https://joelvelasco.net/teaching/167win10/Sober%20-%20two%20faces%20of%20fitness.pdf>

Sonesson, A. K., Woolliams, J. A., & Meuwissen, T. H. (2012). Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution*, 44(1), 1-10. <https://doi.org/10.1186/1297-9686-44-27>

Sonesson, A. K., & Ødegård, J. (2016). Mating structures for genomic selection breeding programs in aquaculture. *Genetics Selection Evolution*, 48(1), 1-7. <https://doi.org/10.1186/s12711-016-0224-y>

Spencer, H. (1864). *The Principles of Biology: Volume 1*.

Springate, J. R. C., Bromage, N. R., Elliott, J. A. K., & Hudson, D. L. (1984). The timing of ovulation and stripping and their effects on the rates of fertilization and survival to eying, hatch and swim-up in the rainbow trout (*Salmo gairdneri* R.). *Aquaculture*, 43(1-3), 313-322. [https://doi.org/10.1016/0044-8486\(84\)90032-2](https://doi.org/10.1016/0044-8486(84)90032-2)

Srikanth, K., Neupane, M., Jaafar, M. A., Wolfe, C. W., Huson, H. J., & Van Tassell, C. P. (2022). Genetic diversity and inbreeding in US Jersey population and germplasm collection. In *World Congress on Genetics Applied to Livestock Production (WCGALP 2022)*. https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/72_004.pdf

Stanković, D., Crivelli, A. J., & Snoj, A. (2015). Rainbow trout in Europe: introduction, naturalization, and impacts. *Reviews in Fisheries Science & Aquaculture*, 23(1), 39-71. <https://doi.org/10.1080/23308249.2015.1024825>

Stock, J., Bennowitz, J., Hinrichs, D., & Wellmann, R. (2020). A review of genomic models for the analysis of livestock crossbred data. *Frontiers in genetics*, 11, 568. <https://doi.org/10.3389/fgene.2020.00568>

Su, G. S., Liljedahl, L. E., & Gall, G. A. E. (1996). Effects of inbreeding on growth and reproductive traits in rainbow trout *Oncorhynchus mykiss*. *Aquaculture*, 142, 139-148. [https://doi.org/10.1016/0044-8486\(96\)01255-0](https://doi.org/10.1016/0044-8486(96)01255-0)

Su, G., Christensen, O. F., Ostensen, T., Henryon, M., & Lund, M. S. (2012). Estimating additive and non-

additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. <https://doi.org/10.1371/journal.pone.0045293>

Sun, C., VanRaden, P. M., O'connell, J. R., Weigel, K. A., & Gianola, D. (2013). Mating programs including genomic relationships and dominance effects. *Journal of dairy science*, *96*(12), 8014-8023. <https://doi.org/10.3168/jds.2013-6969>

Suzuki, R., & Fukuda, (1971). Growth and survival of F1 hybrids among Salmonid fishes. *Bull. Freshwater Fisk. Res. Lab.*, *21*, 117-138.

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical population biology*, *2*(2), 125-141. [https://doi.org/10.1016/0040-5809\(71\)90011-6](https://doi.org/10.1016/0040-5809(71)90011-6)

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular biology and evolution*, *31*(10), 2824-2827. <https://doi.org/10.1093/molbev/msu211>

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595. <https://doi.org/10.1093/genetics/123.3.585>

Talebi, R., Szmatoła, T., Mészáros, G., & Qanbari, S. (2020). Runs of homozygosity in modern chicken revealed by sequence data. *G3: Genes, Genomes, Genetics*, *10*(12), 4615-4623. <https://doi.org/10.1534/g3.120.401860>

Teletchea, F., & Fontaine, P. (2014). Levels of domestication in fish: implications for the sustainable future of aquaculture. *Fish and fisheries*, *15*(2), 181-195. <https://doi.org/10.1111/faf.12006>

Teletchea, F. (2018). Fish domestication: an overview. *Animal domestication*. <https://doi.org/10.5772/intechopen.79628>

Thorgaard, G. H. (1983). Chromosomal differences among rainbow trout populations. *Copeia*, 650-662. <https://doi.org/10.2307/1444329>

Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L., Ristow, S. S., ... & Palti, Y. (2002). Status and opportunities for genomics research with rainbow trout. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *133*(4), 609-646. [https://doi.org/10.1016/S1096-4959\(02\)00167-7](https://doi.org/10.1016/S1096-4959(02)00167-7)

Toro, M. A., & Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genetics Selection Evolution*, *42*(1), 1-9. <https://doi.org/10.1186/1297-9686-42-33>

Valdivia, K., Jouanno, E., Volff, J. N., Galiana-Arnoux, D., Guyomard, R., Helary, L., ... & Guiguen, Y. (2014). High temperature increases the masculinization rate of the all-female (XX) rainbow trout “Mal” population. *PLoS one*, *9*(12), e113355. <https://doi.org/10.1371/journal.pone.0113355>

Vandeputte, M., & Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics*, *5*, 432. <https://doi.org/10.3389/fgene.2014.00432>

VanRaden, P. M. (2007). Genomic measures of relationship and inbreeding. *INTERBULL bulletin*, (37), 33-33.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414-4423. <https://doi.org/10.3168/jds.2007-0980>

Varona, L., Legarra, A., Toro, M. A., & Vitezica, Z. G. (2018a). Non-additive effects in genomic selection. *Frontiers in genetics*, *9*, 78. <https://doi.org/10.3389/fgene.2018.00078>

Varona, L., Legarra, A., Herring, W., & Vitezica, Z. G. (2018b). Genomic selection models for directional

REFERENCES

- dominance: an example for litter size in pigs. *Genetics Selection Evolution*, 50(1), 1-13. <https://doi.org/10.1186/s12711-018-0374-1>
- Varona, L., Altarriba, J., Moreno, C., Martínez-Castillero, M., & Casellas, J. (2019). A multivariate analysis with direct additive and inbreeding depression load effects. *Genetics Selection Evolution*, 51(1), 1-12. <https://doi.org/10.1186/s12711-019-0521-3>
- Vernier, J. M. (1969). Table chronologique du developement embryonnaire de la truite arc-en-ciel, Salmo gairdneri Rich. *Ann. Embryol. Morphogenet.*, 2, 495-520.
- Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*. 21: 2791–2793. <https://doi.org/10.1093/bioinformatics/bti403>
- Villanueva, B., Fernández, A., Peiró-Pastor, R., Peñaloza, C., Houston, R. D., Sonesson, A. K., ... & Saura, M. (2022). Population structure and genetic variability in wild and farmed Mediterranean populations of gilthead seabream and European seabass inferred from a 60K combined species SNP array. *Aquaculture Reports*, 24, 101145. <https://doi.org/10.1016/j.aqrep.2022.101145>
- De Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, 6(11), 1248-1258. <https://doi.org/10.1111/2041-210X.12418>
- Vitezica, Z. G., Varona, L., & Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, 195(4), 1223-1230. <https://doi.org/10.1534/genetics.113.155176>
- Vitezica, Z. G., Varona, L., Elsen, J. M., Misztal, I., Herring, W., & Legarra, A. (2016). Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genetics Selection Evolution*, 48(1), 1-8. <https://doi.org/10.1186/s12711-016-0185-1>
- Vitezica, Z. G., Reverter, A., Herring, W., & Legarra, A. (2018). Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genetics Selection Evolution*, 50(1), 1-8. <https://doi.org/10.1186/s12711-018-0437-3>
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual review of genetics*, 47, 97-120. <https://doi.org/10.1146/annurev-genet-111212-133526>
- Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS biology*, 4(3), e72. <https://doi.org/10.1371/journal.pbio.0040072>
- Wallace, R. A., & Selman, K. (1978). Oogenesis in *Fundulus heteroclitus*: I. Preliminary observations on oocyte maturation in vivo and in vitro. *Developmental biology*, 62(2), 354-369. [https://doi.org/10.1016/0012-1606\(78\)90222-1](https://doi.org/10.1016/0012-1606(78)90222-1)
- Walker, A. F. (2004). *An investigation of escaped rainbow trout in the upper River Earn and Loch Earn during 2002/03*. Fisheries Research Services. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.610.9143&rep=rep1&type=pdf>
- Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient?. *Theoretical population biology*, 107, 4-13. <https://doi.org/10.1016/j.tpb.2015.08.006>
- Wang, J., Hill, W. G., Charlesworth, D., & Charlesworth, B. (1999). Dynamics of inbreeding depression due to deleterious mutations in small populations: Mutation parameters and inbreeding rate. *Genetics Research*, 74(2), 165–178. <https://doi.org/10.1017/S0016672399003900>
- Wang, S., Hard, J. J., & Utter, F. (2002). Salmonid inbreeding: A review. *Reviews in Fish Biology and Fisheries*, 11(4), 301–319. <https://doi.org/10.1023/A:1021330500365>

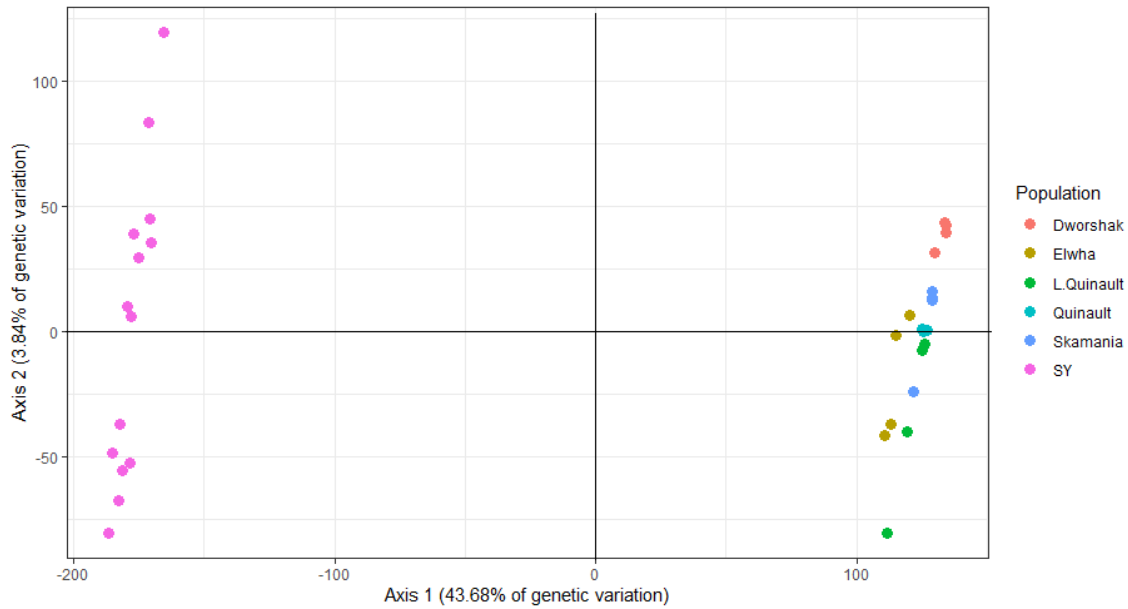
- Waters, C. D., Hard, J. J., Fast, D. E., Knudsen, C. M., Bosch, W. J., & Naish, K. A. (2020). Genomic and phenotypic effects of inbreeding across two different hatchery management regimes in Chinook salmon. *Molecular Ecology*, 29(4), 658–672. <https://doi.org/10.1111/mec.15356>
- Weigand, H., & Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*, 184(2), 528-583. <https://doi.org/10.1093/zoolinnean/zly007>
- Weinberg, W. (1908). Über den nachweis der vererbung beim menschen. *Jh. Ver. vaterl. Naturk. Württemb.*, 64, 369-382.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370. <https://doi.org/10.2307/2408641>
- Wellmann, R., & Bennewitz, J. (2012). Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genetics Research*, 94(1), 21-37. <https://doi.org/10.1017/S0016672312000018>
- Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in ecology & evolution*, 30(1), 42-49. <https://doi.org/10.1016/j.tree.2014.10.009>
- Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *The American Naturalist*, 186, S24–S36.
- Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, 56(645), 330-338.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97-159. <https://doi.org/10.1093/genetics/16.2.97>
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. <http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p356-wright.pdf>
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 395-420. <https://doi.org/10.2307/2406450>
- Wright, D. (2015). Article commentary: the genetic architecture of domestication in animals. *Bioinformatics and biology insights*, 9, BBI-S28902. <https://doi.org/10.4137/BBI.S28902>
- Xiang, T., Christensen, O. F., Vitezica, Z. G., & Legarra, A. (2016). Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genetics Selection Evolution*, 48(1), 1-14. <https://doi.org/10.1186/s12711-016-0271-4>
- Xiang, T., Christensen, O. F., Vitezica, Z. G., & Legarra, A. (2018). Genomic model with correlation between additive and dominance effects. *Genetics*, 209(3), 711-723. <https://doi.org/10.1534/genetics.118.301015>
- Yáñez, J. M., Bassini, L. N., Filp, M., Lhorente, J. P., Ponzoni, R. W., & Neira, R. (2014). Inbreeding and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding nucleus in Chile. *Aquaculture*, 420, S15-S19. <https://doi.org/10.1016/j.aquaculture.2013.05.028>
- Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R. P., Zinovieva, N., ... & Megens, H. J. (2017). Genome-wide SNP data unveils the globalization of domesticated pigs. *Genetics Selection Evolution*, 49(1), 1-15. <https://doi.org/10.1186/s12711-017-0345-y>
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., ... & Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423-1428. <https://doi.org/10.1016/j.cub.2012.05.045>
- Yoshida, G. M., Cáceres, P., Marín-Nahuelpi, R., Koop, B. F., & Yáñez, J. M. (2020). Estimates of autozygosity through runs of homozygosity in farmed coho salmon. *Genes*, 11(5), 490. <https://doi.org/10.3390/genes11050490>

REFERENCES

- Yu, X., Setyawan, P., Bastiaansen, J. W., Liu, L., Imron, I., Groenen, M. A., ... & Megens, H. J. (2022). Genomic analysis of a Nile tilapia strain selected for salinity tolerance shows signatures of selection and hybridization with blue tilapia (*Oreochromis aureus*). *Aquaculture*, *560*, 738527. <https://doi.org/10.1016/j.aquaculture.2022.738527>
- Zanella, R., Peixoto, J. O., Cardoso, F. F., Cardoso, L. L., Biegelmeyer, P., Cantão, M. E., ... & Ledur, M. C. (2016). Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genetics Selection Evolution*, *48*(1), 1-10. <https://doi.org/10.1186/s12711-016-0203-3>
- Zeder, M. A. (2012). The domestication of animals. *Journal of anthropological research*, *68*(2), 161-190.
- Zeng, J., Toosi, A., Fernando, R. L., Dekkers, J., & Garrick, D. J. (2013). Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics Selection Evolution*, *45*(1), 1-17. <https://doi.org/10.1186/1297-9686-45-11>
- Zhang, M., Han, W., Tang, H., Li, G., Zhang, M., Xu, R., ... & Wu, K. (2018). Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. *BMC genomics*, *19*(1), 1-12. <https://doi.org/10.1186/s12864-018-4973-6>
- Zhang, J., Nie, C., Li, X., Ning, Z., Chen, Y., Jia, Y., ... & Qu, L. (2020). Genome-wide population genetic analysis of commercial, indigenous, game, and wild chickens using 600K SNP microarray data. *Frontiers in genetics*, *11*, 543294. <https://doi.org/10.3389/fgene.2020.543294>
- Zhao, Y., Zhu, X., & Sun, X. (2008). Microsatellite diversity in cultured populations of rainbow trout *Oncorhynchus mykiss* in China. *Journal of Fish Biology*, *73*(5), 1249-1255. <https://doi.org/10.1111/j.1095-8649.2008.01894.x>
- Zhao, F., McParland, S., Kearney, F., Du, L., & Berry, D. P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution*, *47*(1), 1-12. <https://doi.org/10.1186/s12711-015-0127-3>

Liste des Annexes

<u>Annexe 1</u> : Annexes du chapitre 4	222
<u>Annexe 2</u> : Article Bernard et al., 2022 (Associé au Chapitre 4)	253
<u>Résultat annexe 3</u> : Effets de la consanguinité chromosomique sur les performances de production d'une lignée de truite arc-en-ciel françaises	269
<u>Résultat annexe 4</u> : Effets locaux de la consanguinité sur la survie du stade oeillé jusqu'à trois semaines après la première alimentation dans la lignée synthétique INRAE	273

Annexe 1 : Annexes du chapitre 4*Figure*

Supplementary figure 1. Principal component analysis (PCA) of the genetic diversity of SY, and HA sub-populations based on 546,903 SNPs.

Tables

Supplementary information S1. Commons signatures of selection between French lines (LB, LC and SY) and American pooled population (HA) detected by XP-EHH statistic.

Chromosome	Start (Mb)	End (Mb)	Size (kb)	p-value (LB vs HA)	p-value (LC vs HA)	p-value (SY vs HA)
1	76.14	77.13	990	4.34	4.52	5.15
1	83.16	84.15	990	5.27	5.02	4.66
2	47.98	48.97	990	6.66	7.33	7.55
3	39.55	40.54	990	5.34	6.13	7.17
3	53.53	54.51	980	5.64	5.81	5.94
3	55.88	56.87	990	5.76	6.17	5.90
4	31.44	32.43	990	5.58	5.78	5.83
5	72.47	73.46	990	7.80	7.84	7.48
6	8.48	9.27	790	4.32	4.95	5.33
7	41.39	42.38	990	5.61	5.72	5.40
8	17.03	18.02	990	4.26	4.64	4.49
8	35.86	36.88	1020	5.14	4.75	5.11
8	43.30	44.29	990	4.07	4.10	4.48
9	22.83	23.82	990	5.55	6.02	5.63
9	32.09	32.96	870	6.32	6.37	7.79
10	38.29	39.28	990	6.50	7.22	7.29
10	50.69	51.63	940	5.31	5.02	5.67
10	53.15	53.98	830	5.01	4.78	5.66
11	30.36	31.35	990	4.94	5.49	6.41
11	70.97	71.59	620	5.00	4.48	5.34
15	25.33	26.32	990	4.25	4.63	4.54
17	21.04	22.03	990	4.74	4.94	5.86
17	45.28	45.94	660	5.35	5.98	5.99
17	48.57	49.55	980	4.32	4.28	4.81
18	21.38	22.37	990	6.98	6.15	7.60
18	30.32	31.31	990	5.78	5.73	6.42
19	60.04	60.57	530	5.13	5.11	5.47
20	11.29	12.86	1570	5.31	5.40	4.55
21	36.39	37.38	990	6.27	7.09	7.97
21	43.71	44.42	710	5.73	4.68	4.73
22	29.95	30.94	990	5.35	5.44	5.06
23	40.32	41.31	990	6.80	6.67	7.84
28	26.39	27.36	970	5.88	5.98	7.04
32	14.60	15.59	990	4.79	5.12	6.20

ANNEXES

Supplementary Informations S2. Listing of ROH islands in LB population.

Population	Region	Chromosome	Start (bp)	End (bp)	minimal number of individual with a ROH	maximum number of individual with a ROH	Size (kb)	Number of SNP
LB	1	1	37915678	38653572	37	45	737.894	165
LB	2	1	64617073	67098816	34	51	2481.743	875
LB	3	2	23868185	32219683	31	42	8351.498	2708
LB	4	2	35239000	39193247	30	41	3954.247	1436
LB	5	2	42888093	46664330	30	41	3776.237	1375
LB	6	2	51143497	54700249	31	37	3556.752	1079
LB	7	2	63944992	65345099	31	37	1400.107	502
LB	8	2	88464243	91540049	30	49	3075.806	1111
LB	9	3	45560871	48398441	36	45	2837.57	1067
LB	10	5	10663782	12345014	31	64	1681.232	485
LB	11	5	19169399	21484242	31	41	2314.843	520
LB	12	5	29736388	30493510	30	43	757.122	294
LB	13	5	64653370	66620509	37	43	1967.139	929
LB	14	5	74156520	74930479	37	38	773.959	225
LB	15	5	75227937	76772085	36	38	1544.148	633
LB	16	5	77637068	78580579	38	39	943.511	289
LB	17	5	78908668	79959615	37	39	1050.947	322
LB	18	5	80749779	82975103	30	40	2225.324	453
LB	19	5	84432804	86254702	31	38	1821.898	535
LB	20	6	53274844	53813409	31	35	538.565	130
LB	21	6	65163997	76578486	30	59	11414.489	4065
LB	22	6	77041039	78317185	32	38	1276.146	253
LB	23	7	29841587	31209238	30	37	1367.651	333
LB	24	8	32451587	32587340	30	35	135.753	47
LB	25	8	32761898	34956105	31	37	2194.207	683
LB	26	8	62506977	63014982	40	47	508.005	234
LB	27	9	8252372	13489877	33	43	5237.505	692
LB	28	9	29076329	29510373	33	35	434.044	120
LB	29	9	31109664	31862453	55	59	752.789	247
LB	30	9	34949838	35704590	34	47	754.752	228
LB	31	9	52582826	53032303	31	36	449.477	160
LB	32	10	5451635	6139194	31	36	687.559	55
LB	33	10	50776636	53365759	30	36	2589.123	883
LB	34	11	43123000	48711819	30	49	5588.819	1675
LB	35	11	61716395	70261960	30	42	8545.565	2640
LB	36	12	57966286	59311318	32	43	1345.032	356
LB	37	14	15375155	19190398	36	44	3815.243	1351
LB	38	14	19466883	19976523	40	41	509.64	222
LB	39	15	15087848	16449713	31	35	1361.865	509
LB	40	15	32049708	32580608	42	43	530.9	98
LB	41	15	38695870	39569031	40	41	873.161	72
LB	42	16	1504245	3678601	30	39	2174.356	95
LB	43	16	42985788	45785712	30	40	2799.924	998
LB	44	16	49865814	50582277	35	36	716.463	247
LB	45	16	63726217	64314707	34	37	588.49	93
LB	46	16	66926376	67673082	30	57	746.706	167
LB	47	17	24215328	25322318	30	45	1106.99	328
LB	48	17	82153140	82704497	31	40	551.357	152
LB	49	17	87983251	88139239	30	36	155.988	11
LB	50	18	18826706	20180178	31	38	1353.472	332
LB	51	18	25177973	26162836	30	39	984.863	333
LB	52	18	29549416	31410069	35	41	1860.653	476
LB	53	20	373366	1072011	32	38	698.645	159
LB	54	20	2151687	2804146	33	39	652.459	92
LB	55	21	19855594	21023455	30	39	1167.861	463
LB	56	21	31909814	32413419	43	44	503.605	164
LB	57	21	37885508	39928412	31	38	2042.904	642
LB	58	21	43027431	43882400	31	35	854.969	254
LB	59	21	45487498	49870381	30	42	4382.883	1552
LB	60	22	21804075	22982060	32	47	1177.985	275
LB	61	23	39969222	40528756	33	36	559.534	182
LB	62	24	10444152	12297872	30	39	1853.72	628
LB	63	24	24661089	25610409	30	41	949.32	243
LB	64	24	31124998	34541321	30	45	3416.323	371
LB	65	25	33619515	37004465	30	36	3384.95	891
LB	66	27	1001200	2456970	34	61	1455.77	346
LB	67	27	3699748	8547295	30	35	4847.547	1786
LB	68	27	40111377	40708890	30	36	597.513	79
LB	69	28	98155	4706304	30	37	4608.149	1229

Supplementary Informations S2 (continued)

LB	70	28	18883939	20780653	33	46	1896.714	727
LB	71	31	19400510	21982123	30	44	2581.613	731
LB	72	32	10404596	11307753	36	48	903.157	88

Supplementary Informations S3. Listing of ROH islands in LC population.

Population	Region	Chromosome	Start (bp)	End (bp)	minimal number of individual with a ROH	maximum number of individual with a ROH	Size (kb)	Number of SNP
LC	1	1	18469187	18556036	23	31	86.849	29
LC	2	1	54787315	55126490	26	27	339.175	91
LC	3	1	55590744	57183606	23	28	1592.862	370
LC	4	2	23970894	26816036	24	34	2845.142	957
LC	5	2	46241786	48995566	25	29	2753.78	835
LC	6	2	54520562	57077054	24	29	2556.492	788
LC	7	2	63332310	64160328	23	28	828.018	267
LC	8	2	70751696	71349680	32	45	597.984	259
LC	9	2	86992445	89342463	26	29	2350.018	740
LC	10	2	101439423	102308529	29	34	869.106	60
LC	11	3	7539334	8248546	25	33	709.212	68
LC	12	3	10830965	11638872	23	31	807.907	166
LC	13	3	11951130	12556060	23	27	604.93	170
LC	14	5	9110690	11424181	24	34	2313.491	642
LC	15	5	29287381	32988836	23	42	3701.455	1352
LC	16	5	83099381	83690115	27	30	590.734	106
LC	17	5	85849888	87377080	23	38	1527.192	377
LC	18	6	18416503	19016941	27	29	600.438	224
LC	19	6	51864705	52926644	25	37	1061.939	245
LC	20	6	68381523	70715726	25	31	2334.203	749
LC	21	6	71516911	73909056	23	28	2392.145	790
LC	22	7	81047629	81961623	24	30	913.994	112
LC	23	8	25298392	27085347	24	30	1786.955	538
LC	24	8	36963617	37700315	24	29	736.698	350
LC	25	9	10691053	11297442	24	27	606.389	67
LC	26	9	17567012	17583148	26	27	16.136	4
LC	27	9	34949838	35680914	28	36	731.076	218
LC	28	9	53323436	53997981	34	38	674.545	194
LC	29	10	28944930	29510233	39	41	565.303	169
LC	30	10	31990087	33246920	23	44	1256.833	433
LC	31	11	22545973	24072616	24	29	1526.643	517
LC	32	11	24829129	25450315	29	32	621.186	281
LC	33	11	26668393	27926206	23	32	1257.813	314
LC	34	12	6865277	7531151	23	28	665.874	40
LC	35	12	25698140	27021366	24	27	1323.226	576
LC	36	12	56481044	59482979	23	27	3001.935	977
LC	37	12	63357961	63898222	23	27	540.261	196
LC	38	12	83932712	85216318	23	28	1283.606	120
LC	39	14	14543801	15756807	23	42	1213.006	520
LC	40	15	65174	907980	26	28	842.806	77
LC	41	15	36088756	42530378	23	49	6441.622	1198
LC	42	16	34508344	35807654	25	37	1299.31	357
LC	43	16	41886939	43022516	25	30	1135.577	475
LC	44	16	60469657	61015644	23	27	545.987	178
LC	45	16	63207327	64133566	23	30	926.239	89
LC	46	16	65006965	66602606	26	30	1595.641	309
LC	47	17	26041970	27178342	24	27	1136.372	342
LC	48	17	80387883	80906573	24	28	518.69	153
LC	49	20	18757371	19826284	23	27	1068.913	734
LC	50	20	31703904	33296665	24	28	1592.761	286
LC	51	20	35397378	36585363	24	31	1187.985	263
LC	52	20	37055019	38346960	23	29	1291.941	96
LC	53	20	39349819	40137287	23	33	787.468	81
LC	54	23	18063822	19340194	23	27	1276.372	432
LC	55	23	29507059	30042661	23	27	535.602	99
LC	56	23	35795558	36391079	34	34	595.521	124
LC	57	24	171798	1063549	29	32	891.751	328
LC	58	25	68959	1242984	23	29	1174.025	187
LC	59	27	41679916	42408578	23	44	728.662	78
LC	60	28	10534860	12236153	23	28	1701.293	630

Supplementary Informations S3 (continued)

LC	61	28	22405654	30421569	23	51	8015.915	2267
LC	62	29	3863667	5342193	26	34	1478.526	347
LC	63	29	7986739	9001851	26	34	1015.112	300
LC	64	30	10089175	10198495	24	29	109.32	34
LC	65	30	37285302	37901325	24	27	616.023	177
LC	66	31	8959147	9930253	23	33	971.106	138
LC	67	32	22259341	35747505	24	47	13488.164	4242
LC	68	32	36554175	41756527	23	48	5202.352	445

Supplementary Informations S4. Listing of ROH islands in SY population.

Population	Region	Chromosome	Start (bp)	End (bp)	minimal number of individual with a ROH	maximum number of individual with a ROH	Size (kb)	Number of SNP
SY	1	1	23895660	32299299	3	6	8403.639	3260
SY	2	1	35808355	39950479	3	6	4142.124	1147
SY	3	1	45482533	49510181	3	6	4027.648	1216
SY	4	1	54017951	57303744	3	6	3285.793	810
SY	5	1	58942450	61925147	3	6	2982.697	915
SY	6	2	21182038	28710841	3	5	7528.803	2380
SY	7	2	41451015	42393208	3	5	942.193	371
SY	8	2	43225731	44717729	3	5	1491.998	556
SY	9	2	46241786	50906815	3	5	4665.029	1175
SY	10	2	85891565	86780092	4	5	888.527	263
SY	11	2	88464243	89603730	3	5	1139.487	374
SY	12	2	101432468	102733482	3	8	1301.014	84
SY	13	3	11175849	15517816	3	5	4341.967	1124
SY	14	3	64957197	69579063	3	5	4621.866	1313
SY	15	4	40360460	42434297	3	5	2073.837	140
SY	16	5	14805124	15503745	3	5	698.621	166
SY	17	5	28983031	34965161	3	5	5982.13	2318
SY	18	5	47622053	57519989	3	6	9897.936	5020
SY	19	5	59636687	64690829	3	5	5054.142	2473
SY	20	5	90605360	91442046	3	5	836.686	100
SY	21	6	51430708	52432183	3	5	1001.475	206
SY	22	6	70848486	78233333	3	8	7384.847	2281
SY	23	7	32384619	33973127	3	7	1588.508	294
SY	24	7	74888980	78923440	3	8	4034.46	1075
SY	25	7	79641828	84848798	3	5	5206.97	550
SY	26	8	20285974	26029716	3	5	5743.742	2096
SY	27	8	34140673	41290630	3	5	7149.957	2587
SY	28	8	42291532	50786152	3	6	8494.62	3072
SY	29	8	52087401	65601888	3	7	13514.487	5243
SY	30	9	29987835	30587507	4	6	599.672	193
SY	31	10	6853184	7485838	3	6	632.654	51
SY	32	10	8877026	10508018	3	5	1630.992	283
SY	33	10	22918523	24879369	3	6	1960.846	695
SY	34	10	44695272	47686380	3	5	2991.108	1026
SY	35	10	51966084	53804053	3	5	1837.969	610
SY	36	11	86726	4901641	3	5	4814.915	269
SY	37	11	26532001	30308576	3	5	3776.575	1109
SY	38	11	44489537	60798881	3	5	16309.344	6106
SY	39	11	63431857	67919727	3	5	4487.87	1408
SY	40	11	70138980	81333617	3	7	11194.637	1793
SY	41	12	20008729	31986993	3	5	11978.264	4608
SY	42	12	57365548	59104178	3	5	1738.63	449
SY	43	12	69516187	76526432	3	7	7010.245	2624
SY	44	13	29905169	31339595	3	7	1434.426	451
SY	45	15	30438939	32222297	3	6	1783.358	543
SY	46	15	38959362	41200349	3	5	2240.987	250
SY	47	15	51305330	52125686	4	5	820.356	187
SY	48	16	36350628	37003116	5	6	652.488	107
SY	49	16	77181969	78523820	3	5	1341.851	112
SY	50	17	22120475	29648519	3	8	7528.044	2212
SY	51	19	19019082	20918714	3	5	1899.632	537
SY	52	20	16186428	17452113	3	5	1265.685	858
SY	53	20	18757371	20293219	3	5	1535.848	1018
SY	54	21	28948774	37816640	3	5	8867.866	2699
SY	55	22	21748790	23146080	3	5	1397.29	373

Supplementary Informations S4 (continued)

SY	56	22	37016900	39652050	3	5	2635.15	528
SY	57	23	16328186	19957220	3	5	3629.034	1161
SY	58	23	38984762	40528756	3	5	1543.994	497
SY	59	23	40899442	44105946	3	6	3206.504	889
SY	60	23	54317250	56040751	3	6	1723.501	337
SY	61	23	57777070	58647026	5	9	869.956	55
SY	62	24	476990	1420878	3	5	943.888	482
SY	63	24	9662109	10496574	3	5	834.465	301
SY	64	25	68959	6563063	3	8	6494.104	1824
SY	65	25	9865509	24852669	3	7	14987.16	4786
SY	66	27	210674	6149283	3	6	5938.609	1734
SY	67	27	30192769	31260897	3	5	1068.128	182
SY	68	28	38517	2751235	4	8	2712.718	687
SY	69	28	5844860	7273036	3	5	1428.176	331
SY	70	29	56625	16568506	3	7	16511.881	4712
SY	71	30	24902128	46284262	3	6	21382.134	6368
SY	72	31	20141654	32686976	3	5	12545.322	3176
SY	73	31	32948668	35680134	3	6	2731.466	1143
SY	74	31	39279227	44004600	3	9	4725.373	1184
SY	75	32	45405	1658695	3	5	1613.29	137
SY	76	32	12150296	13844481	3	5	1694.185	256

Supplementary Informations S5. Listing of ROH islands in HA population.

Population	Region	Chromosome	Start (bp)	End (bp)	minimal number of individual with a ROH	maximum number of individual with a ROH	Size (kb)	Number of SNP
HA	1	1	37988601	38642023	8	14	653.422	143
HA	2	1	53916827	58055002	8	15	4138.175	1094
HA	3	2	12832830	14875197	9	13	2042.367	72
HA	4	2	43376979	46489122	8	19	3112.143	1094
HA	5	2	47110687	48088345	8	18	977.658	291
HA	6	2	48401649	49696799	8	13	1295.15	301
HA	7	3	1301525	6150920	8	20	4849.395	223
HA	8	3	38383412	39266896	9	14	883.484	212
HA	9	4	65560	2087943	9	19	2022.383	531
HA	10	4	3634878	4810490	8	14	1175.612	148
HA	11	5	72370519	73022603	8	13	652.084	166
HA	12	5	79977317	81414409	8	15	1437.092	264
HA	13	5	81863412	83661029	8	14	1797.617	327
HA	14	6	41160543	45917413	8	17	4756.87	1595
HA	15	7	5214022	6720549	12	12	1506.527	40
HA	16	7	33169838	33973127	9	14	803.289	73
HA	17	7	36511838	37538941	9	15	1027.103	278
HA	18	8	38894006	41290630	9	19	2396.624	895
HA	19	8	42291532	45158017	8	17	2866.485	927
HA	20	9	30770572	31684404	9	14	913.832	308
HA	21	9	31951193	33408418	8	19	1457.225	386
HA	22	9	35854387	36685434	8	17	831.047	199
HA	23	10	50464943	52604109	8	20	2139.166	920
HA	24	11	38532676	39272112	8	14	739.436	132
HA	25	12	57214342	60069285	8	19	2854.943	907
HA	26	13	30346169	31268540	8	12	922.371	261
HA	27	13	52277423	65387397	8	20	13109.974	486
HA	28	14	239379	2366736	8	16	2127.357	489
HA	29	15	38645655	41818766	8	16	3173.111	366
HA	30	16	29701986	32368999	8	13	2667.013	927
HA	31	17	14454618	21290705	11	20	6836.087	270
HA	32	17	35959531	39693611	8	17	3734.08	1352
HA	33	17	43066965	44631467	8	15	1564.502	301
HA	34	17	45039983	46086370	8	14	1046.387	547
HA	35	18	5443588	13449480	9	20	8005.892	283
HA	36	18	29849523	31219788	8	14	1370.265	320
HA	37	19	3343276	7734257	17	20	4390.981	164
HA	38	19	8511147	10171555	12	20	1660.408	98
HA	39	19	23019183	24710996	8	16	1691.813	632
HA	40	20	40489189	41092944	8	13	603.755	34

Supplementary Informations S5 (continued)

HA	41	21	10582777	17462095	8	20	6879.318	289
HA	42	21	33646369	36275074	8	18	2628.705	531
HA	43	22	21813326	22946247	8	17	1132.921	255
HA	44	23	10202503	12528117	8	17	2325.614	354
HA	45	25	68959	2064779	8	13	1995.82	435
HA	46	25	2448285	3759251	8	15	1310.966	385
HA	47	26	6899534	10859628	8	14	3960.094	236
HA	48	26	39757066	50852419	9	20	11095.353	514
HA	49	27	210674	1501648	8	17	1290.974	122
HA	50	28	38517	1473516	8	20	1434.999	241
HA	51	29	2166230	2825468	8	12	659.238	80
HA	52	30	43460572	46284262	8	20	2823.69	919
HA	53	31	41362245	42229571	9	15	867.326	188
HA	54	32	36854957	41756527	9	20	4901.57	433

Supplementary Informations S6. Listing of significant signals of selection identified by iHS in LB population.

Population	Chromosome	Start (bp)	End (bp)	SNP number	mean(log(p-value))	max(log(p-value))	number of extreme SNP	mean(log(p-value)) of extreme SNP
LB	1	31100000	32000000	334	0.44	4.082	1	4.082
LB	1	69800000	70700000	254	0.476	4.529	1	4.529
LB	2	23000000	25000000	592	0.561	4.562	7	3.624
LB	2	25600000	29300000	956	0.762	8.593	22	3.751
LB	2	30700000	32300000	372	0.935	4.402	16	3.208
LB	2	64200000	66400000	585	0.649	5.366	4	3.679
LB	2	69700000	72000000	770	0.608	4.083	3	3.064
LB	2	88900000	89800000	191	0.553	4.817	1	4.817
LB	3	23400000	24300000	421	0.772	4.36	7	3.826
LB	3	30500000	31400000	275	0.46	5.797	1	5.797
LB	3	71500000	72400000	151	0.418	4.041	1	4.041
LB	4	11000000	22000000	259	0.319	6.444	4	4.715
LB	4	28900000	30300000	334	0.474	6.415	3	4.433
LB	5	10700000	11600000	146	0.463	4.216	1	3.216
LB	5	27300000	28200000	189	0.449	7.784	1	7.784
LB	5	69300000	70600000	553	0.297	5.681	2	4.703
LB	5	84300000	85200000	274	0.274	4.522	1	4.522
LB	6	40200000	41100000	251	0.47	5.05	1	5.05
LB	6	62300000	64900000	762	0.595	5.647	13	3.563
LB	6	67800000	69900000	631	0.801	5.044	17	3.642
LB	6	70400000	72100000	452	0.699	4.96	8	3.848
LB	8	14500000	15400000	252	0.53	5.093	1	5.093
LB	8	30700000	32900000	644	0.857	4.102	30	3.54
LB	8	35000000	36400000	391	0.873	4.211	12	3.431
LB	8	45200000	46400000	428	0.674	4.234	2	3.757
LB	9	50700000	51600000	257	0.309	4.357	1	4.357
LB	10	29500000	30400000	242	0.339	5.795	1	5.795
LB	11	24900000	25900000	328	0.721	4.361	13	3.713
LB	11	49600000	52700000	1065	0.645	4.264	11	3.403
LB	11	60100000	61000000	295	0.587	4.652	1	4.652
LB	11	67200000	68100000	206	0.672	5.163	1	5.163
LB	15	17500000	18400000	271	0.601	4.402	1	4.402
LB	16	44100000	47000000	697	0.341	9.835	16	4.22
LB	16	47800000	48700000	295	0.302	4.235	1	4.235
LB	17	24000000	24900000	248	0.61	4.185	1	4.185
LB	20	6100000	9400000	1067	1.55	4.569	84	3.444
LB	20	10500000	16500000	2590	1.722	6.439	506	3.749
LB	20	19100000	20100000	770	1.462	4.452	50	3.315
LB	20	34100000	35100000	303	0.721	4.217	3	3.755
LB	21	29500000	30500000	321	0.474	7.5	2	5.374
LB	22	16000000	17700000	620	0.737	8.028	14	3.968
LB	23	17100000	18900000	504	0.867	4.16	20	3.402
LB	23	30500000	32400000	559	0.784	5.559	11	3.529
LB	24	7000000	7900000	221	0.334	5.444	1	5.444
LB	24	11300000	12800000	314	0.652	18.532	2	11.361
LB	26	34400000	36300000	591	0.71	4.326	4	3.493
LB	27	5400000	6400000	359	0.749	10.307	2	6.678
LB	27	6600000	10300000	1215	0.795	4.006	26	3.362
LB	27	12100000	15900000	1222	0.77	5.761	57	3.658

Supplementary Informations S6 (continued)

LB	27	34200000	35100000	84	0.232	6.427	1	6.427
LB	28	30200000	31100000	293	0.384	6.502	1	6.502
LB	29	19200000	20100000	298	0.315	4.433	1	4.433
LB	30	31500000	32400000	227	0.511	5.874	1	5.874
LB	31	16100000	17000000	174	0.572	4.006	1	4.006
LB	31	39900000	40800000	152	0.249	4.035	1	4.035

Supplementary Informations S7. Listing of significant signals of selection identified by iHS in LC population.

Population	Chromosome	Start (bp)	End (bp)	SNP number	mean(log(p-value))	max(log(p-value))	number of extreme SNP	mean(log(p-value)) of extreme SNP
LC	1	17900000	18800000	276	0.489	8.041	1	8.041
LC	1	34200000	35100000	211	0.321	4.113	1	4.113
LC	1	66100000	67000000	214	0.645	6.645	1	6.645
LC	2	23200000	24500000	391	0.558	4.492	4	3.633
LC	2	27000000	27900000	225	0.256	4.501	1	4.501
LC	2	33400000	34300000	345	0.544	4.161	1	4.161
LC	2	46000000	46900000	312	0.347	12.352	1	12.352
LC	2	70000000	71300000	362	0.474	5.604	3	4.774
LC	2	84200000	85100000	246	0.357	4.473	1	4.473
LC	3	31400000	33400000	646	0.604	4.409	15	3.618
LC	3	60500000	62400000	704	0.469	5.097	7	4.21
LC	4	2100000	3500000	424	0.349	8.641	4	5.636
LC	5	9300000	10200000	211	0.737	7.728	1	7.728
LC	5	10300000	11200000	184	0.628	4.612	1	4.612
LC	5	33700000	34600000	325	0.789	4.179	1	4.179
LC	5	49900000	51000000	609	0.256	4.712	2	4.147
LC	6	41600000	42500000	245	0.469	6.843	1	6.843
LC	6	48800000	49600000	103	0.353	4.655	1	4.655
LC	6	65200000	67200000	703	0.823	4.556	26	3.575
LC	7	46500000	47400000	296	0.471	9.596	1	9.596
LC	7	64800000	65700000	257	0.257	5.069	1	5.069
LC	7	74700000	75600000	185	0.416	8.922	1	8.922
LC	8	15300000	16200000	233	0.447	4.13	1	4.13
LC	8	53500000	55400000	746	1.031	4.793	18	3.439
LC	10	59500000	60700000	303	0.767	4.21	4	3.712
LC	11	11600000	12600000	249	0.667	4.155	6	3.414
LC	11	46000000	48100000	506	0.786	4.293	7	3.508
LC	11	48900000	50100000	395	0.616	4.391	7	3.469
LC	11	51000000	52700000	601	0.747	5.116	12	3.83
LC	11	64300000	65700000	349	0.286	6	2	4.765
LC	12	20900000	21800000	303	0.372	5.502	1	5.502
LC	12	52200000	53100000	253	0.543	6.151	1	6.151
LC	12	55600000	56500000	385	0.255	4.502	1	4.502
LC	12	62800000	64400000	401	0.464	4.354	3	3.676
LC	13	41200000	42100000	296	0.399	4.176	1	4.176
LC	14	18700000	21000000	744	1.042	4.801	56	3.631
LC	15	26300000	27200000	308	0.571	4.338	1	4.338
LC	15	27400000	28400000	395	0.679	4.296	5	3.618
LC	15	30300000	32100000	446	0.793	4.364	10	3.784
LC	15	48300000	53100000	1261	0.786	4.056	26	3.326
LC	15	53300000	57700000	1180	0.891	5.69	60	3.943
LC	16	30900000	31800000	321	0.913	4.202	1	4.202
LC	16	45800000	47000000	504	0.826	4.063	5	3.544
LC	16	47600000	48800000	471	0.835	4.554	10	3.56
LC	17	56600000	58500000	698	1.068	4.002	23	3.203
LC	17	69200000	71100000	505	0.879	5.623	41	4.02
LC	18	29500000	30400000	259	0.803	8.224	1	8.224
LC	19	39400000	40300000	292	0.465	4.84	1	4.84
LC	19	41300000	42200000	275	0.75	4.239	3	4.12
LC	20	11200000	13300000	745	1.197	4.787	11	3.357
LC	20	27100000	28700000	536	0.877	4.147	6	3.556
LC	21	25800000	26700000	269	0.249	4.09	1	4.09
LC	22	10200000	12000000	524	0.669	4.05	15	3.346
LC	22	15600000	19000000	1140	0.966	5.644	49	3.504
LC	22	39000000	39900000	106	0.247	4.9	1	4.9
LC	23	17200000	18400000	316	0.491	5.713	4	3.955

Supplementary Informations S7 (continued)

LC	23	28100000	29400000	464	0.457	5.492	2	4.486
LC	24	30000000	30900000	150	0.337	4.496	1	4.496
LC	26	12400000	13300000	277	0.277	7.099	1	7.099
LC	27	2200000	3700000	376	0.629	5.312	2	4.247
LC	27	7200000	10000000	900	0.801	4.025	9	3.341
LC	28	29500000	31400000	578	0.67	5.672	14	3.677
LC	29	9700000	10600000	266	0.559	5.569	1	5.569
LC	29	25900000	26800000	272	0.466	5.967	1	5.967
LC	30	18700000	19800000	316	0.592	4.674	3	4.168
LC	30	30700000	32400000	401	0.805	9.441	21	3.834
LC	30	36700000	38200000	377	0.607	5.734	3	4.253
LC	30	38700000	39600000	221	0.543	4.848	2	4.645
LC	32	27400000	29700000	576	0.94	4.561	12	3.67

Supplementary Informations S8. Listing of significant signals of selection identified by iHS in SY population.

Population	Chromosome	Start (bp)	End (bp)	SNP number	mean(log(p-value))	max(log(p-value))	number of extreme SNP	mean(log(p-value)) of extreme SNP
SY	1	61000000	62000000	230	0.727	6.776	5	4.483
SY	1	62700000	63600000	209	0.481	5.234	1	5.234
SY	1	66500000	67800000	374	0.777	4.805	3	3.833
SY	1	74400000	75800000	356	0.455	6.595	5	4.144
SY	2	31600000	32500000	237	0.456	4.074	1	3.074
SY	2	46300000	47300000	291	0.754	4.551	2	3.777
SY	2	70200000	71200000	366	0.611	4.25	2	3.759
SY	4	1400000	3800000	559	0.503	4.673	4	3.554
SY	4	46300000	46800000	31	0.362	5.603	1	5.603
SY	5	48900000	49800000	414	0.469	7.34	1	7.34
SY	5	63700000	64600000	442	0.497	5.821	1	5.821
SY	5	66500000	68100000	721	0.378	5.118	2	4.575
SY	5	68700000	71300000	1160	0.741	6.967	14	4.065
SY	5	90300000	91200000	123	0.59	5.989	1	5.989
SY	6	21900000	22800000	247	0.662	4.335	1	4.335
SY	6	27400000	28300000	232	0.367	5.247	1	5.247
SY	6	60300000	61200000	291	0.608	4.923	1	4.923
SY	7	39700000	40600000	212	0.424	6.529	1	6.529
SY	7	71300000	72200000	153	0.557	4.103	1	4.103
SY	7	72500000	73900000	282	0.398	4.58	2	3.797
SY	8	54500000	57700000	1003	0.835	5.673	13	3.52
SY	8	60200000	61100000	296	0.654	4.116	1	4.116
SY	8	61600000	64300000	813	0.831	5.394	13	3.577
SY	9	31800000	33400000	316	0.669	5.531	5	4.442
SY	9	35400000	36800000	280	0.574	5.412	6	3.995
SY	9	42800000	43700000	249	0.388	5.202	3	4.903
SY	10	49300000	50700000	468	0.644	7.63	2	5.332
SY	11	9100000	10000000	149	0.339	4.377	1	4.377
SY	11	40900000	42600000	408	0.375	5.076	3	4.081
SY	11	54100000	56600000	814	1.138	4.95	62	3.624
SY	11	64100000	65200000	305	1.15	4.339	9	3.523
SY	12	26000000	26900000	327	0.711	4.208	1	4.208
SY	12	32900000	34700000	417	0.54	4.178	2	3.604
SY	12	36000000	38900000	748	0.868	5.165	34	3.539
SY	12	69700000	72500000	831	0.928	5.221	13	3.772
SY	12	74300000	76300000	472	0.949	6.743	15	3.834
SY	13	22600000	23500000	261	0.692	7.843	2	5.786
SY	14	15900000	18000000	522	0.64	6.14	7	3.93
SY	15	29000000	29900000	181	0.312	4.01	1	4.01
SY	16	14600000	15500000	286	0.468	5.901	1	5.901
SY	16	32700000	33600000	251	0.706	4.16	1	4.16
SY	16	44700000	46100000	426	0.608	4.273	2	3.674
SY	16	57300000	58200000	153	0.377	6.268	1	6.268
SY	17	27000000	28800000	476	0.831	6.391	4	4.267
SY	17	55300000	56800000	456	0.905	5.933	15	3.558
SY	17	59500000	60400000	216	0.556	4.774	1	4.774
SY	17	72200000	73100000	209	0.522	4.302	3	3.917
SY	18	11900000	12400000	40	0.422	5.675	1	5.675
SY	19	19700000	20900000	322	0.805	4.075	2	3.646
SY	19	39400000	41200000	524	0.407	4.875	5	3.689
SY	19	48800000	50600000	455	0.48	4.1	2	3.736

Supplementary Informations S8 (continued)

SY	20	6200000	7100000	259	0.395	4.345	1	4.345
SY	20	13000000	14200000	445	0.956	5.369	2	4.978
SY	20	16400000	17300000	582	0.225	4.533	1	4.533
SY	21	21400000	22300000	238	0.396	4.099	1	4.099
SY	21	29600000	31900000	653	0.582	4.06	4	3.408
SY	23	13600000	16400000	943	0.844	4.109	9	3.416
SY	24	9700000	11000000	406	0.631	4.522	2	4.308
SY	24	21900000	23100000	290	0.639	4.515	2	3.814
SY	26	11300000	12800000	477	0.487	6.094	3	4.492
SY	27	15800000	16700000	178	0.423	5.122	1	5.122
SY	27	50000000	50500000	30	0.366	4.145	1	4.145
SY	28	9200000	10100000	212	0.335	5.389	1	5.389
SY	28	15600000	16500000	240	0.37	4.044	1	4.044
SY	28	22100000	24900000	858	0.83	4.374	7	3.39
SY	28	30800000	32600000	450	0.451	5.007	3	4.896
SY	29	3800000	4700000	165	0.608	7.131	1	7.131
SY	29	20900000	21800000	261	0.501	4.022	1	4.022
SY	29	22400000	23300000	261	0.398	4.688	1	4.688
SY	30	12500000	13400000	274	0.434	6.137	1	6.137
SY	31	22800000	23700000	102	0.477	4.046	1	4.046
SY	31	33100000	35400000	764	0.947	5.433	31	3.787
SY	32	22400000	23300000	195	0.318	4.652	1	4.652

Supplementary Informations S9. Listing of significant signals of selection identified by iHS in HA population.

Population	Chromosome	Start (bp)	End (bp)	SNP number	mean(log(p-value))	max(log(p-value))	number of extreme SNP	mean(log(p-value)) of extreme SNP
HA	1	1000000	11500000	202	0.342	4.299	2	3.717
HA	1	16700000	20300000	729	0.533	5.298	10	4.254
HA	1	24000000	25900000	429	0.515	6.291	7	3.995
HA	1	29900000	31900000	402	0.467	4.942	3	4.217
HA	1	32500000	35100000	464	0.485	4.09	9	3.423
HA	1	37000000	38400000	165	0.401	5.857	2	4.454
HA	1	40600000	41900000	256	0.431	5.401	2	4.473
HA	1	42400000	43300000	179	0.41	5.633	2	4.367
HA	1	43900000	46700000	517	0.477	4.435	4	3.972
HA	1	50600000	52000000	254	0.431	4.094	2	3.983
HA	1	52100000	53100000	167	0.482	7.127	2	5.291
HA	1	53700000	54400000	51	0.379	5.963	1	5.963
HA	1	54800000	55700000	245	0.481	4.371	5	4.363
HA	1	57400000	59500000	304	0.574	4.975	6	3.916
HA	1	61500000	62400000	222	0.505	5.527	1	5.527
HA	1	63200000	64100000	135	0.6	4.363	1	4.363
HA	1	65100000	66000000	168	0.502	5.764	2	4.725
HA	1	66200000	68000000	351	0.502	8.041	7	4.883
HA	1	68300000	69400000	204	0.581	6.374	6	5.113
HA	1	69700000	72600000	437	0.339	4.113	4	3.476
HA	1	80200000	83000000	299	0.338	4.152	4	3.591
HA	2	18200000	19100000	79	0.324	4.696	1	4.696
HA	2	25400000	26300000	132	0.355	4.006	1	3.006
HA	2	29800000	30800000	198	0.449	5.69	2	4.568
HA	2	31600000	34400000	611	0.372	5.321	4	4.405
HA	2	34500000	36800000	484	0.479	5.53	9	4.227
HA	2	37200000	39800000	555	0.513	8.311	9	4.652
HA	2	41700000	43900000	321	0.564	10.848	5	6.599
HA	2	46800000	47700000	127	0.534	4.677	3	4.324
HA	2	47800000	48900000	89	0.709	13.587	4	9.445
HA	2	51400000	54100000	293	0.466	7.047	4	4.846
HA	2	55700000	57400000	296	0.363	6.111	3	4.862
HA	2	61900000	63700000	273	0.448	5.003	5	4.4
HA	2	69500000	72500000	583	0.439	8.26	8	4.233
HA	2	80300000	81500000	145	0.457	5.248	3	4.089
HA	2	85300000	86200000	171	0.489	4.485	1	4.485
HA	2	87000000	87900000	142	0.408	4.031	1	4.031
HA	2	89500000	90400000	141	0.396	5.369	1	5.369
HA	2	1,03E+08	1,04E+08	34	0.446	9.163	1	9.163
HA	3	9600000	10700000	105	0.412	6.058	2	4.797
HA	3	11200000	13200000	263	0.393	6.778	3	4.871

ANNEXES

Supplementary information S9 (continued)

HA	3	18100000	19000000	153	0.353	7.648	1	7.648
HA	3	22100000	26000000	934	0.528	8.317	14	4.055
HA	3	26200000	29400000	629	0.452	5.34	6	4.001
HA	3	32100000	33000000	177	0.399	4.471	1	4.471
HA	3	35900000	39400000	609	0.598	7.012	10	4.489
HA	3	48400000	51400000	707	0.413	10.205	8	4.818
HA	3	55500000	56400000	138	0.385	5.031	2	4.131
HA	3	57500000	59800000	493	0.509	5.244	8	4.168
HA	3	65400000	66300000	178	0.399	5.563	1	5.563
HA	3	66400000	67300000	121	0.345	4.364	1	4.364
HA	3	71400000	73100000	192	0.451	17.892	2	12.317
HA	3	75400000	77200000	242	0.353	4.149	2	3.819
HA	4	1500000	2100000	56	0.862	5.669	2	5.287
HA	4	12700000	13600000	151	0.482	6.103	3	5.028
HA	4	14200000	15600000	268	0.536	6.196	6	4.828
HA	4	15800000	17300000	269	0.346	5.205	2	4.368
HA	4	17500000	21900000	786	0.415	4.245	10	3.506
HA	4	22900000	25600000	475	0.508	19.023	10	5.115
HA	4	27500000	29400000	238	0.363	4.097	4	3.543
HA	4	36400000	37300000	84	0.315	5.469	1	5.469
HA	4	37700000	38300000	38	0.328	5.823	1	5.823
HA	5	17200000	18100000	142	0.451	4.666	3	3.867
HA	5	19000000	20700000	241	0.472	9.489	4	4.983
HA	5	23800000	26000000	384	0.351	5.368	5	3.795
HA	5	26300000	30200000	769	0.477	5.563	7	3.537
HA	5	31600000	33200000	341	0.642	8.643	6	4.973
HA	5	33800000	38800000	1226	0.645	6.979	12	4.132
HA	5	38900000	40400000	453	0.681	6.695	7	4.408
HA	5	44400000	47000000	646	0.686	5.021	8	3.667
HA	5	47100000	48800000	534	0.695	5.487	5	3.928
HA	5	49100000	60500000	3520	0.732	14.552	73	4.659
HA	5	61000000	63000000	531	0.801	6.667	14	4.651
HA	5	63400000	70200000	1796	0.653	6.75	29	4.321
HA	5	71400000	72600000	222	0.653	7.68	4	4.818
HA	5	74500000	76700000	484	0.614	4.521	7	3.618
HA	5	77900000	79500000	297	0.501	6.962	2	4.986
HA	5	89400000	90900000	210	0.37	4.102	3	3.849
HA	6	700000	1400000	34	0.471	4.389	1	4.389
HA	6	9000000	12300000	408	0.418	11.995	6	5.788
HA	6	24500000	26800000	445	0.421	4.58	4	4.035
HA	6	28100000	29200000	222	0.524	5.505	4	3.887
HA	6	29600000	34700000	875	0.522	6.476	24	4.026
HA	6	36300000	38400000	375	0.422	4.46	3	3.754
HA	6	40200000	43300000	320	0.995	10.432	29	4.937
HA	6	43400000	47200000	503	1.033	13.657	41	5.055
HA	6	48100000	50400000	323	0.616	8.29	3	5.193
HA	6	54000000	54900000	142	0.343	4.87	1	4.87
HA	6	56700000	57900000	273	0.434	4.16	2	3.882
HA	6	58200000	59100000	136	0.382	5.719	1	5.719
HA	6	67800000	70000000	441	0.624	7.683	10	4.544
HA	6	70800000	72500000	300	0.532	4.394	5	3.534
HA	6	73900000	74800000	162	0.376	6.241	1	6.241
HA	6	76200000	77100000	91	0.332	4.11	1	4.11
HA	7	16500000	17900000	202	0.377	6.679	2	4.844
HA	7	21700000	22600000	184	0.355	4.654	1	4.654
HA	7	30100000	33100000	349	0.432	10.906	7	5.183
HA	7	37800000	39400000	367	0.43	11.145	3	6.92
HA	7	39700000	41200000	205	0.45	5.295	3	4.087
HA	7	41900000	43200000	246	0.468	4.79	2	4.535
HA	7	48800000	49700000	152	0.388	5.875	1	5.875
HA	7	50600000	54300000	600	0.508	5.57	15	4.332
HA	7	54400000	55300000	96	0.321	4.37	1	4.37
HA	7	56300000	57200000	94	0.375	5.001	1	5.001
HA	7	58300000	60600000	348	0.423	6.705	3	4.608
HA	7	61500000	63000000	241	0.339	7.11	3	4.627
HA	7	64900000	66700000	326	0.493	4.192	6	3.503
HA	7	66800000	67700000	148	0.402	6.497	1	6.497
HA	7	68400000	69300000	125	0.383	4.482	1	4.482
HA	7	70700000	72300000	247	0.396	6.113	4	4.34
HA	7	75800000	76700000	126	0.366	4.813	2	4.258
HA	7	77800000	78700000	124	0.371	5.99	1	5.99
HA	8	9400000	10300000	137	0.434	6.107	2	6.107
HA	8	11000000	13300000	331	0.515	5.054	8	4.035
HA	8	15000000	19200000	793	0.433	4.928	8	3.523
HA	8	19300000	21100000	336	0.396	6.055	5	3.853

Supplementary information S9 (continued)

HA	8	21200000	24300000	633	0.451	7.256	7	4.825
HA	8	26100000	27700000	291	0.318	4.317	2	3.748
HA	8	30900000	31800000	172	0.408	4.298	1	4.298
HA	8	32200000	33100000	134	0.385	7.149	1	7.149
HA	8	36200000	39200000	406	0.646	15.149	10	5.6
HA	8	39500000	40200000	40	1.066	11.796	5	7.14
HA	8	44900000	46100000	139	0.483	6.857	2	5.033
HA	8	46900000	48200000	273	0.536	5.027	3	3.998
HA	8	49800000	50700000	182	0.485	5.163	2	4.916
HA	8	53100000	61700000	1617	0.473	7.999	25	3.863
HA	8	62100000	64700000	589	0.445	4.708	7	3.774
HA	8	64900000	68100000	639	0.416	5.081	6	3.725
HA	8	70300000	72000000	326	0.373	4.335	2	3.691
HA	8	72900000	74300000	233	0.519	5.036	5	4.222
HA	8	81400000	82100000	60	0.353	8.051	1	8.051
HA	9	13400000	16000000	307	0.414	5.629	7	4.068
HA	9	22300000	23200000	178	0.368	6.609	1	6.609
HA	9	24900000	27800000	633	0.457	6.625	4	4.969
HA	9	28100000	32500000	521	0.787	14.201	34	5.019
HA	9	33100000	36300000	438	0.72	10.161	11	5.029
HA	9	36400000	39000000	413	0.47	4.335	7	3.641
HA	9	42200000	47600000	1003	0.463	5.797	11	4.211
HA	9	52100000	54000000	326	0.458	10.858	3	7.151
HA	9	54900000	56900000	355	0.517	7.105	8	5.179
HA	9	65400000	67400000	309	0.368	4.373	4	3.753
HA	10	8700000	9600000	96	0.369	4.731	1	4.731
HA	10	20700000	21900000	218	0.537	5.102	3	3.918
HA	10	25500000	27300000	386	0.648	5.867	15	4.156
HA	10	27400000	31000000	659	0.462	5.347	7	4.115
HA	10	31500000	33800000	395	0.379	4.719	5	4.081
HA	10	34800000	36600000	316	0.568	4.665	6	4.272
HA	10	39000000	41400000	457	0.561	9.359	10	5.572
HA	10	44600000	46300000	321	0.505	4.179	3	3.835
HA	10	54900000	55800000	121	0.412	4.853	1	4.853
HA	10	56900000	59000000	235	0.453	5.446	4	4.308
HA	10	59600000	61400000	290	0.431	5.185	6	3.952
HA	10	66300000	68000000	203	0.311	4.87	3	3.97
HA	11	8200000	9100000	77	0.473	4.025	1	4.025
HA	11	9400000	11600000	293	0.325	4.109	3	3.715
HA	11	13800000	15100000	281	0.437	6.284	2	5.723
HA	11	18200000	19800000	235	0.392	4.527	3	3.693
HA	11	20000000	21800000	319	0.374	4.953	3	4.178
HA	11	23600000	28300000	829	0.426	5.581	11	3.599
HA	11	29900000	30800000	122	0.534	7.563	3	5.257
HA	11	32300000	33200000	159	0.317	4.718	1	4.718
HA	11	34300000	36700000	345	0.467	9.494	3	5.658
HA	11	37600000	38600000	73	0.774	18.022	4	8.054
HA	11	41300000	46400000	725	0.443	10.812	11	4.642
HA	11	46700000	50200000	542	0.435	6.435	6	4.218
HA	11	51500000	53600000	464	0.453	6.072	6	3.784
HA	11	57800000	59400000	376	0.48	6.376	4	4.176
HA	11	71300000	73200000	211	0.368	8.992	3	5.718
HA	11	7500000	75700000	51	0.355	5.04	1	5.04
HA	11	76700000	77600000	69	0.255	4.41	1	4.41
HA	12	7800000	8700000	82	0.272	4.119	1	4.119
HA	12	12600000	15300000	344	0.465	8.714	7	4.6
HA	12	21300000	23000000	324	0.463	6.738	3	4.552
HA	12	24100000	29100000	1029	0.496	5.421	18	3.627
HA	12	30900000	31800000	139	0.39	4.909	1	4.909
HA	12	38600000	40300000	313	0.467	6.559	9	4.02
HA	12	43500000	46800000	528	0.412	4.676	6	3.799
HA	12	49600000	52100000	401	0.491	4.207	4	3.713
HA	12	53700000	56400000	653	0.564	5.204	10	4.078
HA	12	56800000	57700000	137	0.513	4.267	1	4.267
HA	12	58700000	60500000	181	0.842	18.282	9	8.175
HA	12	64300000	65900000	258	0.452	5.385	2	4.249
HA	12	68300000	70600000	428	0.409	6.642	8	4.468
HA	12	71300000	72400000	229	0.52	6.412	4	5.448
HA	12	72700000	76000000	587	0.431	11.94	9	5.141
HA	12	76400000	77800000	290	0.35	5.342	3	3.985
HA	12	81500000	82400000	106	0.409	4.615	1	4.615
HA	13	20500000	21400000	166	0.398	4.272	2	3.936
HA	13	22400000	25600000	614	0.402	4.56	7	3.739
HA	13	33700000	35900000	511	0.448	7.302	6	4.39
HA	13	39700000	43100000	777	0.476	6.536	15	3.874

ANNEXES

Supplementary information S9 (continued)

HA	13	43600000	46800000	610	0.529	7.005	15	4.835
HA	14	3100000	5300000	307	0.537	6.542	4	5.18
HA	14	5900000	7000000	209	0.424	4.18	2	3.921
HA	14	7900000	8800000	136	0.487	4.698	2	3.87
HA	14	9300000	10200000	151	0.413	8.093	1	8.093
HA	14	10500000	11500000	189	0.402	5.836	2	4.469
HA	14	12800000	13700000	156	0.276	4.926	1	4.926
HA	14	17000000	18800000	337	0.4	6.613	4	4.123
HA	14	19800000	22100000	367	0.499	6.123	6	4.286
HA	14	23100000	24700000	269	0.431	4.76	2	4.062
HA	14	24800000	27400000	345	0.49	7.996	7	4.859
HA	15	18300000	22200000	785	0.402	5.782	9	3.816
HA	15	22500000	24800000	503	0.512	10.461	7	4.609
HA	15	25600000	28400000	578	0.504	5.76	11	4.064
HA	15	37200000	38700000	176	0.452	9.945	3	6.085
HA	15	42200000	43100000	76	0.376	5.126	1	5.126
HA	15	45300000	46700000	204	0.442	5.553	4	3.871
HA	15	48700000	50000000	251	0.372	5.684	3	4.659
HA	15	51800000	52700000	104	0.335	4.249	1	4.249
HA	15	54100000	55500000	196	0.401	4.781	3	4.072
HA	15	56400000	57700000	235	0.31	5.292	2	4.197
HA	16	13200000	17900000	855	0.5	6.989	13	3.937
HA	16	24600000	25500000	162	0.462	4.556	1	4.556
HA	16	29000000	30100000	135	0.35	5.59	3	4.329
HA	16	32500000	34900000	401	0.566	5.74	9	3.639
HA	16	36700000	37600000	78	0.383	9.012	1	9.012
HA	16	39100000	42300000	616	0.536	6.481	12	4.313
HA	16	45100000	47500000	498	0.479	7.168	13	4.205
HA	16	49900000	51700000	337	0.454	4.311	3	3.895
HA	16	52300000	54300000	463	0.544	9.909	9	4.846
HA	16	54900000	55800000	173	0.358	5.644	1	5.644
HA	17	26000000	26900000	151	0.412	7.388	1	7.388
HA	17	27700000	28600000	181	0.41	4.99	1	4.99
HA	17	30800000	31700000	160	0.414	10.476	1	10.476
HA	17	32700000	33600000	151	0.339	5.035	1	5.035
HA	17	33800000	35100000	244	0.365	12.542	2	8.299
HA	17	40400000	41300000	108	0.523	13.752	1	13.752
HA	17	46800000	47700000	128	0.589	4.521	1	4.521
HA	17	48100000	49100000	175	0.705	10.956	3	8.395
HA	17	53400000	54300000	115	0.306	4.188	1	4.188
HA	17	59900000	61900000	385	0.377	5.621	6	4.185
HA	17	66700000	70200000	672	0.407	5.139	6	3.81
HA	17	73500000	74400000	134	0.31	4.785	1	4.785
HA	17	76400000	77300000	192	0.491	4.331	1	4.331
HA	17	78700000	79600000	128	0.325	5.357	1	5.357
HA	17	81500000	82700000	172	0.515	12.131	2	7.684
HA	17	84600000	85300000	57	0.262	4.247	1	4.247
HA	18	17800000	19800000	195	0.333	4.184	3	3.697
HA	18	22400000	24100000	298	0.371	4.297	2	4.147
HA	18	25600000	28400000	518	0.477	5.243	9	3.833
HA	18	28900000	30000000	164	0.496	4.316	2	3.772
HA	18	32200000	33100000	164	0.493	9.684	2	8.512
HA	18	35900000	36800000	168	0.424	5.95	2	4.688
HA	18	37700000	38700000	237	0.493	4.943	2	4.457
HA	18	40600000	43100000	511	0.507	5.957	12	3.993
HA	18	46200000	48000000	249	0.353	8.005	2	6.121
HA	18	49700000	50600000	146	0.465	10.297	1	10.297
HA	18	52400000	53500000	153	0.395	4.138	2	3.805
HA	18	53800000	55000000	150	0.412	4.672	2	4.129
HA	18	58400000	59300000	148	0.477	4.024	1	4.024
HA	18	62700000	63600000	81	0.324	5.102	1	5.102
HA	19	16300000	17200000	139	0.317	4.05	1	4.05
HA	19	18400000	19300000	111	0.427	11.91	1	11.91
HA	19	22900000	24800000	206	0.608	6.144	4	3.98
HA	19	25700000	27100000	245	0.528	4.653	6	3.951
HA	19	28600000	30800000	347	0.548	5.971	8	4.338
HA	19	31400000	33000000	316	0.453	4.194	6	3.649
HA	19	33400000	34700000	264	0.45	4.678	3	4.343
HA	19	35900000	38800000	539	0.462	4.243	9	3.656
HA	19	38900000	41400000	520	0.445	5.853	8	4.032
HA	19	42700000	43800000	197	0.465	4.24	3	3.726
HA	19	44600000	47700000	629	0.484	11.624	10	4.54
HA	19	48800000	50400000	266	0.374	5.298	3	4.436
HA	19	50500000	51400000	84	0.461	14.858	1	14.858
HA	19	54300000	55900000	274	0.446	4.734	2	3.951

Supplementary information S9 (continued)

HA	20	12300000	13200000	98	0.366	6.817	1	6.817
HA	20	14000000	17500000	453	0.631	9.512	14	6.094
HA	20	18600000	20600000	377	0.956	15.648	16	6.465
HA	20	21600000	22500000	206	0.387	5.263	1	5.263
HA	20	22800000	24500000	363	0.403	4.235	2	3.892
HA	20	28500000	30800000	364	0.453	6.786	7	4.179
HA	20	33200000	35100000	311	0.445	6.558	5	4.986
HA	20	39400000	40000000	39	0.564	15.71	1	15.71
HA	21	22800000	24600000	424	0.431	6.912	3	4.443
HA	21	25000000	27200000	344	0.405	8.727	5	5.183
HA	21	30500000	33000000	515	0.501	15.722	14	5.043
HA	21	35600000	37200000	164	0.503	13.862	2	10.432
HA	21	44400000	45900000	237	0.381	5.242	2	4.415
HA	21	46500000	47700000	222	0.476	4.312	5	3.619
HA	21	49200000	50400000	217	0.398	4.249	2	3.724
HA	22	9600000	10500000	146	0.277	4.868	1	4.868
HA	22	10700000	11600000	160	0.321	7.07	1	7.07
HA	22	12000000	13000000	198	0.486	6.116	3	5.726
HA	22	13800000	18300000	895	0.439	7.297	11	4.511
HA	22	19200000	20200000	185	0.478	4.662	3	4.014
HA	22	22700000	25000000	292	0.492	8.431	6	4.797
HA	22	26900000	27900000	176	0.445	4.371	4	3.525
HA	22	29700000	30900000	95	0.463	8.973	2	6.978
HA	23	12100000	14400000	330	0.613	6.96	5	4.737
HA	23	22800000	24300000	199	0.366	4.289	2	4.185
HA	23	27500000	29300000	343	0.51	5.408	8	3.713
HA	23	30000000	33000000	510	0.536	4.579	14	3.811
HA	23	37700000	38600000	121	0.328	4.261	1	4.261
HA	23	39000000	40400000	213	0.283	4.525	2	3.91
HA	23	45300000	46900000	160	0.358	5.742	3	4.66
HA	24	0	4200000	768	0.814	9.609	53	4.048
HA	24	9000000	11100000	391	0.418	4.328	4	3.72
HA	24	15000000	16800000	310	0.42	5.593	4	4.525
HA	24	17900000	19200000	280	0.376	4.723	4	4.401
HA	24	22000000	23000000	160	0.459	5.929	2	5.403
HA	24	24100000	25300000	180	0.392	4.201	3	3.731
HA	24	27200000	28100000	75	0.347	8.355	1	8.355
HA	24	33600000	34100000	30	0.328	5.602	1	5.602
HA	25	1200000	2100000	152	0.371	7.662	1	7.662
HA	25	8400000	9900000	270	0.451	6.098	2	5.196
HA	25	16800000	18700000	318	0.491	7.245	4	4.826
HA	25	19400000	20500000	204	0.397	4.865	2	4.349
HA	25	21400000	22300000	187	0.492	4.242	2	3.759
HA	25	23100000	25100000	366	0.388	8.965	3	6.084
HA	25	28800000	30500000	284	0.464	4.602	5	3.701
HA	26	10900000	12200000	304	0.661	8.883	10	5.835
HA	26	17100000	20800000	749	0.512	7.6	12	4.564
HA	26	25400000	26500000	262	0.423	7.598	3	5.188
HA	26	27100000	28700000	314	0.462	4.546	5	3.612
HA	26	33900000	35000000	272	0.378	4.305	2	3.962
HA	26	35400000	36500000	211	0.462	5.432	2	4.589
HA	27	1500000	2500000	136	0.759	15.056	6	9.081
HA	27	4000000	5100000	195	0.475	6.906	4	4.206
HA	27	5600000	6900000	266	0.444	4.466	2	4.04
HA	27	8100000	9000000	147	0.416	8.119	1	8.119
HA	27	10900000	13800000	551	0.438	4.538	8	3.527
HA	27	14900000	16800000	330	0.406	9.626	4	4.756
HA	27	22100000	24500000	554	0.444	7.147	4	4.351
HA	27	36300000	36900000	46	0.329	5.219	1	5.219
HA	28	1100000	2000000	87	0.473	6.016	1	6.016
HA	28	4000000	4900000	105	0.412	4.436	1	4.436
HA	28	7400000	9300000	309	0.482	4.915	5	4.417
HA	28	12200000	15900000	609	0.43	4.303	11	3.633
HA	28	18000000	19600000	323	0.564	5.422	7	3.767
HA	28	21100000	24000000	534	0.521	7.391	10	4.519
HA	28	27400000	29000000	302	0.481	6.604	4	4.861
HA	28	29300000	31100000	297	0.555	6.543	11	4.044
HA	28	32200000	34000000	219	0.349	5.359	3	3.898
HA	29	7300000	8200000	124	0.436	4.677	2	4.677
HA	29	12000000	12900000	143	0.512	5.692	1	5.692
HA	29	13600000	14500000	204	0.416	4.391	2	4.218
HA	29	15500000	16400000	120	0.336	4.333	1	4.333
HA	29	17900000	26300000	1561	0.487	7.902	29	3.851
HA	29	26700000	28300000	323	0.458	6.477	2	4.915
HA	29	28500000	30400000	416	0.457	5.475	4	4.682

ANNEXES

Supplementary information S9 (continued)

HA	29	31200000	36000000	811	0.422	6.683	14	4.139
HA	30	10700000	12100000	192	0.49	14.365	2	9.504
HA	30	12300000	13200000	138	0.418	6.706	1	6.706
HA	30	13800000	14800000	169	0.374	4.588	2	3.905
HA	30	16700000	17700000	150	0.541	6.238	2	5.492
HA	30	17900000	18800000	126	0.521	4.537	1	4.537
HA	30	20000000	23900000	803	0.486	4.73	14	3.636
HA	30	28200000	31000000	400	0.353	5.302	5	3.981
HA	30	32400000	33300000	125	0.309	4.644	1	4.644
HA	31	17300000	20000000	370	0.374	4.841	4	3.548
HA	31	20500000	22500000	293	0.484	4.495	3	3.666
HA	31	29900000	32200000	352	0.38	4.38	4	3.468
HA	31	32400000	36300000	810	0.484	5.953	18	3.829
HA	31	36800000	37700000	128	0.479	4.47	1	4.47
HA	31	40100000	41500000	196	0.494	4.552	2	4.049
HA	32	13500000	14200000	53	0.427	7.914	1	7.914
HA	32	19300000	20400000	174	0.449	4.664	3	3.592
HA	32	20800000	23500000	452	0.457	13.801	4	6.535
HA	32	24600000	25500000	119	0.413	5.138	1	5.138
HA	32	29100000	30000000	93	0.35	5.636	1	5.636
HA	32	32200000	33800000	362	0.414	5.125	3	4.688

Supplementary Information S10. iHS and ROH values per population within the 9 overlapping regions detecting in the four populations.

Region	Chromosome	Start (bp)	End (bp)	Size (kb)	Population	iHS statistic			ROH islands	
						min_iHS	max_iHS	number of SNP with iHS >2.5	minimal number of individual with a ROH	maximum number of individual with a ROH
chr2_a	2	25400000	26300000	900	LB	-0.48	4.002	26	37	41
					LC	-0.813	2.055	0	34	34
					SY	-2.186	2.678	2	4	4
					HA	-1.449	3.295	3	0	0
chr2_b	2	31600000	34200000	2600	LB	-3.115	3.542	85	22	38
					LC	-2.944	3.98	10	6	16
					SY	-3.338	3.403	20	0	1
					HA	-4.575	4.477	9	0	0
chr2_c	2	46000000	46664330	664.3	LB	-0.693	1.207	0	30	41
					LC	-0.986	7.241	1	22	27
					SY	-1.799	2.522	1	2	4
					HA	-0.788	2.794	1	5	19
chr2_d	2	69700000	71200000	1500	LB	-2.408	3.344	15	15	24
					LC	-0.502	4.709	8	18	45
					SY	-4.028	1.844	13	1	3
					HA	-3.47	4.401	15	0	1
chr2_e	2	88464243	89342463	878.2	LB	0	3.278	2	42	49
					LC	-1.059	2.523	1	27	29
					SY	-1.412	3.388	1	3	5
					HA	-2.804	2.845	2	0	0
chr12_a	12	57966286	59104178	1137.9	LB	-1.311	2.413	0	32	43
					LC	-0.689	0.824	0	26	27
					SY	-1.521	1.33	0	4	5
					HA	-0.293	8.682	9	12	19
chr15_a	15	38959362	39569031	609.7	LB	0	0.772	0	41	41
					LC	0.382	1.11	0	43	49
					SY	-0.377	1.024	0	3	5
					HA	-0.493	0.2	0	15	16
chr16_a	16	45800000	47000000	1200	LB	-0.618	4.59	16	21	28
					LC	-3.177	3.926	35	5	11
					SY	-3.194	1.83	10	0	0
					HA	-5.397	2.695	17	1	1
chr20_a	20	19100000	19826284	726.3	LB	-1.773	3.573	340	28	28
					LC	-1.487	1.804	0	23	27
					SY	-0.986	2.372	0	5	5
					HA	-3.261	8.208	24	4	7

Supplementary Information S11. Listing of the 253 genes which composed the nine shared genomic regions. In bold, genes identified linked to domestication in the literature.

Region	Chromosome	Start (bp)	End (bp)	GeneID_NCBI	Gene name	Protein_name
chr2_a	2	25404261	25413252	LOC110539045	<i>mrp2</i>	melanocortin-2 receptor accessory protein 2A
	2	25429983	25439975	LOC110539072	<i>cep162</i>	centrosomal protein of 162 kDa
	2	25459885	25460412	si:ch211-198c19.1	<i>si:ch211-198c19.1</i>	uncharacterized protein si:ch211-198c19.1
	2	25670709	25731769	LOC110539098	<i>adgrb1</i>	adhesion G protein-coupled receptor B1
	2	25865381	25942747	tsnare1	<i>tsnare1</i>	t-SNARE domain-containing protein 1
	2	26099353	26113491	LOC110539113	<i>pttg1IP</i>	pituitary tumor-transforming gene 1 protein-interacting protein
	2	26124100	26273005	cdk14	<i>cdk14</i>	cyclin-dependent kinase 14
chr2_b	2	31152535	31153718	LOC110486775	<i>cd83</i>	CD83 antigen
	2	31181413	31188053	LOC110486838	<i>col9a2</i>	collagen alpha-2(IX) chain
	2	31192780	31196175	LOC110486852	<i>snrnp48</i>	LOW QUALITY PROTEIN: U11/U12 small nuclear ribonucleoprotein 48 kDa protein
	2	31199645	31201994	sri	<i>SRI</i>	sorcin
	2	31231836	31247754	herpud2	<i>herpud1</i>	homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 2 protein
	2	31259759	31260705	mxtx2	<i>mxtx2</i>	mix-type homeobox gene 2
	2	31264146	31292609	LOC110534710	<i>bms1</i>	LOW QUALITY PROTEIN: ribosome biogenesis protein BMS1 homolog
	2	31298565	31299524	pkdc	<i>pkdc</i>	uncharacterized protein pkdc
	2	31300296	31315596	LOC110486933	<i>capn7</i>	calpain-7
	2	31319545	31351681	LOC110486940	<i>sh3bp5</i>	SH3 domain-binding protein 5
	2	31357998	31359956	LOC110486970	<i>mettl6</i>	tRNA N(3)-methylcytidine methyltransferase METTL6
	2	31360570	31362443	eaf1	<i>eaf1</i>	ELL-associated factor 1
	2	31367384	31387352	LOC110486958	<i>bmper</i>	BMP-binding endothelial regulator protein
	2	31578536	31590319	LOC110486984	<i>nt5c</i>	cytosolic 5'-nucleotidase 3
	2	31592526	31595674	dctn3	<i>dctn3</i>	dynactin subunit 3
	2	31596953	31618883	LOC110486995	<i>atg5</i>	autophagy protein 5
	2	31633434	31641196	LOC110486999	<i>prdm1</i>	PR domain zinc finger protein 1
	2	31766295	31777509	LOC110487017	<i>tbx20</i>	T-box transcription factor TBX20
	2	31841920	31856046	LOC110534719	<i>dpy19l1</i>	probable C-mannosyltransferase DPY19L1
	2	31857942	31866115	LOC100136303	<i>fbpl4</i>	FBPL4
	2	31868189	31869340	LOC118939268	<i>LOC118939268</i>	putative per-hexamer repeat protein 5
	2	31913752	31931554	LOC110487037	<i>SMARCC1 sm</i>	SWI/SNF complex subunit SMARCC1
	2	31942520	31972532	LOC110487045	<i>chst5</i>	chondroitin sulfate proteoglycan 5
	2	32000093	32024459	LOC110534747	<i>scap</i>	sterol regulatory element-binding protein cleavage-activating protein
	2	32027899	32056089	LOC110487053	<i>ptpn23</i>	tyrosine-protein phosphatase non-receptor type 23
	2	32057504	32062277	LOC110487062	<i>gabpb1</i>	GA-binding protein subunit beta-1
	2	32097058	32132329	LOC110487106	<i>sema6d</i>	semaphorin-6D
2	32333153	32348060	LOC110487115	<i>cucyl1b1</i>	soluble guanylate cyclase gcy-31	
2	32363786	32364509	tnfaip8l2b	<i>tnf- a - ip8l2b</i>	tumor necrosis factor, alpha-induced protein 8-like protein 2 B	
2	32368959	32370920	lysmd1	<i>lysmd1</i>	lysM and putative peptidoglycan-binding domain-containing protein 1	
2	32371292	32373051	scnm1	<i>scnm1</i>	sodium channel modifier 1	
2	32375013	32380552	LOC110487143	<i>tmod4</i>	tropomodulin-4	
2	32428673	32429827	LOC110487180	<i>mcl1</i>	induced myeloid leukemia cell differentiation protein Mcl-1 homolog	
2	32434714	32441843	ensab	<i>ensab</i>	endosulfine alpha b	
2	32442802	32454733	hormad1	<i>hormad1</i>	HORMA domain-containing protein 1	

Supplementary information S11 (continued)

2	32478786	32485523	onecutl	<i>onecutl</i>	one cut domain, family member, like
2	32489987	32497084	LOC110487241	<i>ctsk</i>	cathepsin S
2	32499316	32504873	ctsk	<i>ctsk</i>	cathepsin K
2	32540574	32542412	lingo4b	<i>lingo4b</i>	leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 4b
2	32550989	32564064	si:ch211-113g11.6	<i>sema7b</i>	semaphorin-7A
2	32578905	32581987	tlr18	<i>tlr18</i>	toll-like receptor 18
2	32590666	32591396	LOC110487342	<i>s100a1</i>	protein S100-A1
2	32593905	32596968	rrnad1	<i>rrnad1</i>	protein RRNAD1
2	32597851	32605204	isg20l2	<i>isg20l2</i>	interferon-stimulated 20 kDa exonuclease-like 2
2	32613407	32631768	LOC110534753	<i>bcan</i>	brevican core protein
2	32633341	32636147	LOC110487305	<i>hapln2</i>	hyaluronan and proteoglycan link protein 2-like
2	32673814	32683648	rhbg2a	<i>rhbg2a</i>	ammonium transporter Rh type B
2	32811832	32848876	LOC110487398	<i>mef2d</i>	myocyte-specific enhancer factor 2D homolog
2	32856648	32868058	LOC110487370	<i>tanabin</i>	tanabin
2	32873278	32875512	LOC110487428	<i>clec12a</i>	C-type lectin domain family 12 member B
2	32878443	32881626	LOC110487411	<i>asgr2</i>	asialoglycoprotein receptor 2
2	32878443	32881216	LOC110487411	<i>cd209</i>	CD209 antigen-like protein B
2	32926581	32938324	LOC110487447	<i>pear1</i>	platelet endothelial aggregation receptor 1
2	32941392	32960787	LOC110487461	<i>lma</i>	lamin-A
2	32988968	32994158	LOC110487479	<i>mex3a</i>	RNA-binding protein MEX3A
2	33003748	33010644	LOC110487493	<i>rab11a</i>	ras-related protein Rab-11A
2	33019162	33020144	LOC110487507	<i>lamtor2</i>	regulator complex protein LAMTOR2
2	33022576	33026896	LOC110487525	<i>ubqln4</i>	ubiquilin-4
2	33044669	33045436	LOC110487536	<i>plcd1</i>	1-phosphatidylinositol phosphodiesterase
2	33146782	33147465	LOC110487587	<i>LOC110487587</i>	protein rapunzel
2	33211470	33224395	LOC110487618	<i>nr2f5</i>	nuclear receptor subfamily 2 group F member 5
2	33255043	33255495	bola1	<i>bola1</i>	bolA-like protein 1
2	33260547	33290072	sv2a	<i>sv2a</i>	synaptic vesicle glycoprotein 2A
2	33328714	33351350	LOC110534759	<i>mtme11</i>	myotubularin-related protein 11
2	33362775	33372155	LOC110487659	<i>OTUD7B</i>	OTU domain-containing protein 7B
2	33401266	33404426	rbm8a	<i>rbm8a</i>	RNA-binding protein 8A
2	33407301	33408369	LOC110487688	<i>LOC110487688</i>	type-4 ice-structuring protein LS-12
2	33434944	33460206	ntrk1	<i>ntrk1</i>	high affinity nerve growth factor receptor
2	33471154	33476762	LOC110487724	<i>paqr6</i>	membrane progesterin receptor delta
2	33490557	33503096	smg5	<i>smg5</i>	protein SMG5
2	33611066	33643616	LOC110487763	<i>syngap1</i>	ras/Rap GTPase-activating protein SynGAP-like
2	33648123	33662526	LOC110487773	<i>phf20l1</i>	PHD finger protein 1
2	33670182	33692897	LOC110487797	<i>rgl2</i>	ral guanine nucleotide dissociation stimulator-like 2
2	33695470	33704008	LOC110487810	<i>dnase2</i>	deoxyribonuclease-2-alpha
2	33720096	33725595	LOC110487856	<i>pbx1</i>	pre-B-cell leukemia transcription factor 1
2	33726513	33732154	LOC110487833	<i>notch3</i>	neurogenic locus notch homolog protein 3
2	33734250	33745953	tap1	<i>tap1</i>	antigen peptide transporter 1
2	33750187	33757005	LOC110487876	<i>brd2</i>	bromodomain-containing protein 2
2	33782498	33840056	LOC110487899	<i>rftn1</i>	raftlin
2	33874323	33879452	LOC110487915	<i>cysteine-rich venom</i>	cysteine-rich venom protein TEL1

Supplementary information S11 (continued)

	2	33889008	33944053	LOC110487926	<i>plcl2</i>	inactive phospholipase C-like protein 2
	2	34076040	34121602	LOC110487942	<i>satb1</i>	DNA-binding protein SATB1
	2	34164780	34168823	LOC118939520	<i>maip1</i>	m-AAA protease-interacting protein 1, mitochondrial-like
	2	34181013	34184551	LOC110488005	<i>rab5a</i>	ras-related protein Rab-5A
	2	34186054	34188011	LOC110534780	<i>pp2c1</i>	protein phosphatase 2C-like domain-containing protein 1
	2	34195019	34234227	LOC110488012	<i>znf385d</i>	zinc finger protein 385D
chr2_c	2	45596070	46131467	brsk2a	<i>brsk2a</i>	serine/threonine-protein kinase BRSK2
	2	46140455	46288787	abtb2b	<i>abtb2b</i>	ankyrin repeat and BTB/POZ domain-containing protein 2
	2	46321492	46490963	b4galnt4a	<i>b4galnt4a</i>	N-acetyl-beta-glucosaminyl-glycoprotein 4-beta-N-acetylglactosaminyltransferase 1
chr2_d	2	69650570	69766876	LOC110494902	<i>bcar1</i>	breast cancer anti-estrogen resistance protein 1
	2	69776967	69827147	LOC110494926	<i>cfdp1</i>	craniofacial development protein 1
	2	69827848	69831977	LOC110494936	<i>tmem170a</i>	transmembrane protein 170A
	2	69834778	69870527	LOC110501833	<i>carmil3</i>	capping protein, Arp2/3 and myosin-I linker protein 3
	2	69872387	69890582	LOC110494911	<i>ctcf</i>	transcriptional repressor CTCF
	2	69914840	69920645	LOC110494959	<i>map11</i>	microtubule-associated protein 11
	2	69960808	69979564	zgc:152891	<i>alox15B</i>	polyunsaturated fatty acid lipoxygenase ALOX15B
	2	69996569	70038308	LOC110494967	<i>ripor1</i>	rho family-interacting cell polarization regulator 1
	2	70109361	70111616	LOC118941470	<i>LOC118941470</i>	C-factor-like
	2	70134549	70167339	LOC110495023	<i>pard6a</i>	partitioning defective 6 homolog alpha
	2	70183642	70210082	LOC110495034	<i>ranbp10</i>	ran-binding protein 10
	2	70221012	70235940	LOC110495070	<i>nutf2</i>	nuclear transport factor 2
	2	70252156	70291045	LOC110495093	<i>edc4</i>	enhancer of mRNA-decapping protein 4
	2	70294687	70320596	LOC110495151	<i>nrn1</i>	neuritin
	2	70355083	70363343	LOC110495161	<i>pskh1</i>	serine/threonine-protein kinase H1 homolog
	2	70372940	70377931	LOC110495172	<i>mbtps1</i>	membrane-bound transcription factor site-1 protease
	2	70395483	70402130	slc38a8a	<i>slc38a8a</i>	putative sodium-coupled neutral amino acid transporter 8
	2	70405038	70546678	LOC110495228	<i>necab2</i>	N-terminal EF-hand calcium-binding protein 2
	2	70648771	70656952	LOC110495269	<i>pnp</i>	LOW QUALITY PROTEIN: purine nucleoside phosphorylase
	2	70666820	70679267	LOC110495285	<i>calb2</i>	calretinin
	2	70685893	70689370	got2a	<i>got2a</i>	glutamic-oxaloacetic transaminase 2a, mitochondrial
	2	70693645	70697009	si:dkey-12e7.4	<i>si:dkey-12e7.4</i>	C-factor
	2	70699710	70701617	syt19	<i>syt19</i>	synaptotagmin 1
	2	70715509	70732105	chka	<i>chka</i>	choline kinase alpha
	2	70735105	70739752	kmt5b	<i>kmt5b</i>	histone-lysine N-methyltransferase KMT5B
	2	70750174	70765049	anpepa	<i>anpepa</i>	alanyl (membrane) aminopeptidase a
	2	70766184	70767921	arpin	<i>arpin</i>	arpin
	2	70769802	70773077	fam169b	<i>fam169b</i>	protein FAM169B
	2	70795398	70902345	igf1ra	<i>igf1a</i>	insulin-like growth factor 1a receptor
	2	70911618	70923623	synm	<i>synm</i>	synemin
	2	70926970	70944684	ttc23	<i>ttc23</i>	tetratricopeptide repeat protein 23
	2	71034885	71088499	LOC110495445	<i>mej2a</i>	myocyte-specific enhancer factor 2A
2	71110153	71148367	LOC110495464	<i>mctp2</i>	multiple C2 and transmembrane domain-containing protein 2	
2	71424383	71431788	LOC110495486	<i>nr2f2</i>	COUP transcription factor 2	
2	71705733	71712750	LOC110495511	<i>arrdc4</i>	arrestin domain-containing protein 4	
2	71766655	71770738	LOC110495533	<i>lysmd4</i>	lysM and putative peptidoglycan-binding domain-containing protein 4	

Supplementary information S11 (continued)

	2	71996319	72005735	LOC110495562	<i>rgma</i>	repulsive guidance molecule A
	2	88401782	88566003	LOC110499041	<i>nav2</i>	neuron navigator 2
	2	88569808	88574053	LOC110499068	<i>dbx1</i>	homeobox protein DBX1-B-like
	2	88626108	88628087	LOC110499106	<i>rnf141</i>	RING finger protein 141
	2	88632318	88638134	LOC110499075	<i>ampd3</i>	AMP deaminase 3
	2	88662471	88664105	LOC110499113	<i>znf135</i>	gastrula zinc finger protein XICGF26.1
	2	88691792	88725368	LOC110499119	<i>lamb1</i>	laminin subunit beta-1
	2	88738982	88747708	<i>bmt2</i>	<i>bmt2</i>	S-adenosylmethionine sensor upstream of mTORC1
chr2_e	2	88751842	88757030	<i>tmem168b</i>	<i>tmem168b</i>	transmembrane protein 168
	2	88763933	88770041	<i>znf277</i>	<i>znf277</i>	zinc finger protein 277
	2	88815803	88980439	LOC110502274	<i>immp2l</i>	mitochondrial inner membrane protease subunit 2-like
	2	88961466	88963583	LOC110499152	<i>lrrn3</i>	leucine-rich repeat neuronal protein 3
	2	89054222	89068369	<i>amfra</i>	<i>amfra</i>	autocrine motility factor receptor a
	2	89081398	89099738	<i>nfatc3b</i>	<i>nfatc3b</i>	nuclear factor of activated T-cells, cytoplasmic 3
	2	89101074	89106577	<i>dus2</i>	<i>dus2</i>	tRNA-dihydrouridine(20) synthase
	2	89108542	89122783	<i>slc7a10b</i>	<i>slc7a10b</i>	asc-type amino acid transporter 1
	2	89125571	89127080	LOC110499212	<i>faap24</i>	Fanconi anemia core complex-associated protein 24
	2	89132966	89160451	<i>tdrd12</i>	<i>tdrd12</i>	putative ATP-dependent RNA helicase TDRD12
	2	89176737	89201371	<i>ankrd27</i>	<i>ankrd27</i>	ankyrin repeat domain-containing protein 27
	2	89206172	89208254	LOC110499228	<i>irx5</i>	iroquois-class homeodomain protein IRX-5
	12	57928364	58070475	LOC110537841	<i>mtmr4</i>	myotubularin-related protein 4
	12	58077014	58099778	<i>ca4a</i>	<i>ca4a</i>	carbonic anhydrase 4
	12	58160036	58183948	LOC110537844	<i>gusb</i>	beta-glucuronidase
	12	58202471	58217899	LOC110537845	<i>vkorc11l</i>	vitamin K epoxide reductase complex subunit 1-like protein 1
	12	58226058	58272243	LOC110537846	<i>ggnbp2</i>	gametogenetin-binding protein 2
	12	58274489	58284851	LOC110537848	<i>sap30l</i>	histone deacetylase complex subunit SAP30L
chr12_a	12	58289152	58290329	LOC110537849	<i>hand1</i>	heart- and neural crest derivatives-expressed protein 1
	12	58515077	58526170	LOC110539152	<i>cops6</i>	COP9 signalosome complex subunit 6
	12	58576295	58623621	<i>rab28</i>	<i>rab28</i>	ras-related protein Rab-28
	12	58641808	58658331	LOC118937820	<i>grxcr1</i>	glutaredoxin domain-containing cysteine-rich protein 1-like
	12	58705256	58824206	LOC110538919	<i>atp8a1</i>	phospholipid-transporting ATPase IA
	12	58829891	58832988	LOC110537866	<i>shisa3</i>	protein shisa-3 homolog
	12	58876007	58915967	LOC110537867	<i>rab11FIP5</i>	rab11 family-interacting protein 5
	12	58935911	58956990	LOC110537868	<i>sfxn5</i>	sideroflexin-5
	12	58961794	58974179	LOC110537870	<i>smyd5</i>	SET and MYND domain-containing protein 5
	12	59052329	59104678	LOC110537871	<i>gfra4</i>	GDNF family receptor alpha-4-like
	15	38957118	39015668	LOC110490041	<i>chn1</i>	N-chimaerin
	15	39018648	39020549	LOC110490042	<i>atp5mc1</i>	ATP synthase lipid-binding protein, mitochondrial
chr15_a	15	39033211	39061162	LOC110490043	<i>zc3h15</i>	zinc finger CCCH domain-containing protein 15
	15	39082814	39083994	LOC110490726	<i>ZP4</i>	zona pellucida sperm-binding protein 4-like
	15			LOC110490841	<i>LOC110490841</i>	uncharacterized protein LOC110490841
	15	39114220	39115842	LOC110490842	<i>nid2</i>	nidogen-2-like
	15	39562174	39588605	<i>brca2</i>	<i>brca2</i>	breast cancer type 2 susceptibility protein

Supplementary information S11 (continued)

	16	45791987	45805978	cd63	<i>cd63</i>	CD63 antigen
	16	45818302	45831783	nemp1	<i>nemp1</i>	nuclear envelope integral membrane protein 1
	16	45837473	45842758	LOC110492119	<i>orm2</i>	ORM1-like protein 2
	16	45852109	45867009	pan2	<i>pan2</i>	PAN2-PAN3 deadenylation complex catalytic subunit PAN2
	16	45887779	45896918	LOC110492121	<i>spats2</i>	spermatogenesis-associated serine-rich protein 2
	16	45922380	45931301	LOC110492123	<i>pym1</i>	partner of Y14 and mago B
	16	45932961	45941812	LOC110492122	<i>mmp19</i>	matrix metalloproteinase-19
	16	45944572	45947261	LOC110492124	<i>tmem198</i>	transmembrane protein 198
	16	45962257	45963989	LOC110492125	<i>trhr</i>	thyrotropin-releasing hormone receptor
	16	45975689	45977866	ocstamp	<i>ocstamp</i>	osteoclast stimulatory transmembrane protein
	16	45980078	45991542	slc13a3	<i>slc13a3</i>	solute carrier family 13 member 3
	16	46000033	46001046	LOC110491071	<i>mc3r</i>	melanocortin receptor 3-like, partial
	16	46026070	46027508	tcf15	<i>tcf15</i>	transcription factor 15
	16	46039031	46039995	srxn1	<i>srxn1</i>	sulfiredoxin-1
	16	46055386	46060369	scrt2	<i>scrt2</i>	transcriptional repressor scratch 2
	16	46103518	46106135	slc52a3	<i>slc52a3</i>	solute carrier family 52, riboflavin transporter, member 3-A
	16	46116439	46122783	LOC110492133	<i>ube2v1</i>	ubiquitin-conjugating enzyme E2 variant 1
	16	46131166	46134522	peds1	<i>peds1</i>	transmembrane protein 189
	16	46137263	46138138	cebpb	<i>cebpb</i>	CCAAT enhancer binding protein beta
	16	46153582	46165363	LOC110492134	<i>ptpn1</i>	tyrosine-protein phosphatase non-receptor type 1
	16	46171316	46184400	LOC110492135	<i>klhl12</i>	kelch-like protein 12
	16	46189750	46192483	tuba5	<i>tuba5</i>	tubulin alpha 5
chr16_a	16	46194311	46229676	kdm5bb	<i>kdm5bb</i>	lysine-specific demethylase 5B-B
	16	46254179	46255285	LOC110492737	<i>rap1a</i>	ras-related protein O-Krev
	16	46260285	46264827	zgc:198371	<i>zgc:198371</i>	XK-related protein 7
	16	46267810	46277748	LOC110492739	<i>cables2</i>	CDK5 and ABL1 enzyme substrate 2
	16	46280748	46283694	LOC110492740	<i>adrm1</i>	proteasomal ubiquitin receptor ADRM1
	16	46288507	46292847	LOC110492742	<i>osbp2</i>	oxysterol-binding protein-related protein 2
	16	46294628	46316126	LOC110492743	<i>SOGA1</i>	protein SOGA1
	16	46317915	46329345	LOC110492744	<i>wrap73</i>	WD repeat-containing protein on Y chromosome
	16	46331570	46336851	LOC110492745	<i>mapre14</i>	microtubule-associated protein RP/EB family member 1
	16	46382835	46407767	LOC110492747	<i>nol4</i>	nucleolar protein 4-like
	16	46417647	46418260	LOC110492748	<i>ppdpf</i>	pancreatic progenitor cell differentiation and proliferation factor
	16	46423689	46439366	LOC110492750	<i>tpd54</i>	tumor protein D54
	16	46443478	46445975	LOC110492749	<i>dnajC5</i>	dnaJ homolog subfamily C member 5-like
	16	46457962	46459730	LOC110492751	<i>gid8</i>	glucose-induced degradation protein 8-B homolog
	16	46479850	46486073	LOC110492754	<i>ythdf1</i>	YTH domain-containing family protein 1
	16	46490812	46493922	LOC110492757	<i>atp1b3</i>	sodium/potassium-transporting ATPase subunit beta-1-interacting protein 3
	16	46524604	46530269	LOC110492758	<i>samd10</i>	sterile alpha motif domain-containing protein 10-like
	16	46540016	46543744	emilin3a	<i>emilin3a</i>	EMILIN-3
	16	46567913	46571385	LOC110492761	<i>opn5</i>	opsin-5-like
	16	46583240	46589315	LOC110492763	<i>rad21</i>	double-strand-break repair protein rad21 homolog
	16	46593042	46597455	LOC110492762	<i>snph</i>	syntaphilin
	16	46600097	46601328	LOC110492764	<i>fkbp1a</i>	peptidyl-prolyl cis-trans isomerase FKBP1A
	16	46607164	46608453	LOC110492766	<i>fam110a</i>	protein FAM110A
	16	46612536	46645311	LOC110491075	<i>prickle1</i>	protein prick1

Supplementary information S11 (continued)

16	46659721	46668308	LOC110492767	<i>plp2</i>	proteolipid protein 2
16	46708591	46722538	LOC110492768	<i>syp</i>	synaptophysin
16	46724994	46761095	LOC110491077	<i>cacna1d</i>	LOW QUALITY PROTEIN: voltage-dependent L-type calcium channel subunit alpha-1D
16	46764819	46765740	LOC110492770	<i>praf2</i>	PRA1 family protein 2
16	46766926	46770437	LOC110492773	<i>timm17b</i>	mitochondrial import inner membrane translocase subunit Tim17-B
16	46771068	46774894		<i>pqbp1</i>	polyglutamine-binding protein 1
16	46777971	46795794	LOC110492774	<i>gpkow</i>	G-patch domain and KOW motifs-containing protein
16	46798792	46853552	LOC110491078	<i>magi2</i>	membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2
16	46906638	46914324	LOC110492775	<i>iqsec1</i>	IQ motif and SEC7 domain-containing protein 1
16	46924374	46942426	LOC110492776	<i>kdm5c</i>	lysine-specific demethylase 5C
16	46942434	46944083	LOC118939461	<i>epr1</i>	extensin-1-like
16	46946225	46971540	LOC110492777	<i>pdzd4</i>	PDZ domain-containing protein 4
16	46996622	46997503	LOC110492778	<i>znf85</i>	zinc finger protein 85-like
20	19124574	19197977	LOC110499153	<i>psd3</i>	PH and SEC7 domain-containing protein 1
20	19207902	19241166	LOC110499151	<i>chuk</i>	inhibitor of nuclear factor kappa-B kinase subunit alpha
20	19263898	19277012	LOC110499148	<i>adam8</i>	disintegrin and metalloproteinase domain-containing protein 8
20	19263898	19276983	LOC110499148	<i>adam15</i>	zinc metalloproteinase-disintegrin-like batroxstatin-2
20	19277618	19294251	LOC110499147	<i>tubg2</i>	gamma-tubulin complex component 2
20	19295431	19299977	LOC110499149	<i>kctd5</i>	BTB/POZ domain-containing protein KCTD5
20	19326901	19349351	LOC110499145	<i>slc16a6</i>	monocarboxylate transporter 6
20	19368982	19375969	LOC110499144	<i>chad</i>	chondroadherin
chr20_a	19401951	19416245	LOC110499142	<i>ndella</i>	nuclear distribution protein nudE-like 1-A
20	19441424	19555711	LOC110498769	<i>arhgap44</i>	rho GTPase-activating protein 44
20	19569198	19582643	LOC110499141	<i>stx4</i>	syntaxin-4
20	19586425	19611153	LOC110499140	<i>srcap</i>	helicase SRCAP
20	19616220	19638879	LOC110499139	<i>auts2</i>	autism susceptibility gene 2 protein homolog
20	19654834	19660537	LOC110499137	<i>chmp2A</i>	charged multivesicular body protein 2a
20	19664929	19669043	LOC110499136	<i>mssl1</i>	male-specific lethal 1-like 1
20	19676040	19676731	LOC110499134	<i>gip</i>	gastric inhibitory polypeptide
20	19726823	19728056	LOC110499130	<i>hoxb6b</i>	homeobox protein Hox-B6b
20	19798136	19840524	LOC110498765	<i>skap1</i>	src kinase-associated phosphoprotein 1

Supplementary Informations S12. Listing of heterozygous regions (i.e., large region without ROH) in LB population.

region	Chromosome	Start (bp)	End (bp)	Size (kb)	number of SNPs	minimal number of individual with a ROH	maximum number of individual with a ROH
1	3	20932960	21787508	854.548	361	4	4
2	7	44945401	47769161	2823.76	1111	2	3
3	10	54005056	54580594	575.538	200	4	4
4	10	54643941	56366064	1722.123	591	4	4
5	12	31109809	31669452	559.643	191	4	4
6	12	81185993	82161784	975.791	244	1	4
7	13	46877469	48108041	1230.572	483	0	4
8	18	13731709	15067746	1336.037	275	2	4
9	19	10752774	11517322	764.548	217	2	4
10	22	25415420	27410424	1995.004	605	3	4
11	30	29137877	30108981	971.104	254	3	4
12	30	30859968	32126922	1266.954	383	2	4
13	30	35446017	36471348	1025.331	210	3	4
14	30	37203420	38189828	986.408	280	3	4

Supplementary Informations S13. Listing of heterozygous regions (i.e., large region without ROH) in LC population.

region	Chromosome	Start (bp)	End (bp)	Size (kb)	number of SNPs	minimal number of individual with a ROH	maximum number of individual with a ROH
1	9	29604620	30599568	994.948	299	3	3
2	9	36782568	37296123	513.555	194	3	3
3	10	56060307	56694250	633.943	209	3	3
4	10	64252947	66065580	1812.633	490	2	3
5	12	18755972	20390539	1634.567	571	2	3
6	12	81192894	82239170	1046.276	167	2	3
7	13	45409805	46334815	925.01	372	0	3
8	13	46955921	48360507	1404.586	525	0	3
9	13	48412138	50792890	2380.752	620	0	3
10	17	21658471	23182159	1523.688	373	1	3
11	18	14820820	15350497	529.677	124	1	3
12	18	15403295	17066735	1663.44	458	1	2
13	19	9963655	10822968	859.313	220	0	3
14	19	11354167	13374472	2020.305	649	1	3
15	19	13539209	14130588	591.379	142	3	3
16	19	48235900	49566091	1330.191	401	1	2
17	19	50103924	50830925	727.001	208	1	3
18	19	56046866	56558017	511.151	91	3	3
19	22	31593680	32450470	856.79	254	3	3
20	25	26381475	27123926	742.451	225	3	3
21	26	35933773	37051208	1117.435	385	2	3
22	26	38214654	39029655	815.001	219	2	3
23	28	36269231	36959816	690.585	105	2	3
24	28	37021001	37765217	744.216	88	2	3

Supplementary Informations S14. Listing of heterozygous regions (i.e., large region without ROH) in SY population.

region	Chromosome	Start (bp)	End (bp)	Size (kb)	number of SNPs	minimal number of individual with a ROH	maximum number of individual with a ROH
1	1	9995928	12700815	2704.887	766	0	0
2	1	12756348	13615891	859.543	260	0	0
3	1	13694438	16432502	2738.064	889	0	0
4	1	16769649	18862977	2093.328	813	0	0
5	1	19453041	20281029	827.988	325	0	0
6	1	20821696	21467957	646.261	226	0	0
7	1	50039426	51397664	1358.238	535	0	0
8	1	64203069	66688670	2485.601	824	0	0
9	1	71205959	73274631	2068.672	635	0	0
10	1	73325555	74508457	1182.902	417	0	0
11	1	79252384	82212217	2959.833	843	0	0

Supplementary information S14 (continued)

12	2	33519755	34936942	1417.187	661	0	0
13	3	6079582	7167560	1087.978	284	0	0
14	3	8089476	8823630	734.154	106	0	0
15	3	9396986	10468007	1071.021	294	0	0
16	3	23724843	24401991	677.148	390	0	0
17	3	28838436	33904578	5066.142	1945	0	0
18	3	34449542	35361491	911.949	256	0	0
19	3	36958088	37690227	732.139	239	0	0
20	3	39224017	40838733	1614.716	394	0	0
21	3	53828832	55275507	1446.675	458	0	0
22	3	55844016	57160888	1316.872	550	0	0
23	4	10647450	12451468	1804.018	696	0	0
24	4	24185879	25204131	1018.252	357	0	0
25	4	31986044	32644695	658.651	189	0	0
26	5	77838955	79008048	1169.093	349	0	0
27	5	82557158	83309957	752.799	185	0	0
28	5	84008896	84564957	556.061	231	0	0
29	5	90024309	90559355	535.046	141	0	0
30	6	12583054	14204443	1621.389	373	0	0
31	6	14263670	16551878	2288.208	680	0	0
32	6	22897905	23897833	999.928	356	0	0
33	7	46831570	48158847	1327.277	522	0	0
34	7	64023103	66823907	2800.804	1015	0	0
35	8	8202034	8947603	745.569	100	0	0
36	8	29102526	29909364	806.838	160	0	0
37	8	72496299	73341421	845.122	294	0	0
38	8	78608813	80190455	1581.642	360	0	0
39	8	81482943	82108400	625.457	100	0	0
40	9	28227651	29045520	817.869	308	0	0
41	9	36937152	37806587	869.435	347	0	0
42	9	44554227	48401243	3847.016	1232	0	0
43	9	49929590	50629204	699.614	252	0	0
44	9	51189621	51988651	799.03	293	0	0
45	9	52768782	53366238	597.456	254	0	0
46	9	53905186	54692097	786.911	275	0	0
47	9	61998436	63413746	1415.31	493	0	0
48	9	65860483	66384695	524.212	121	0	0
49	9	69249208	71080887	1831.679	303	0	0
50	10	26046570	27383092	1336.522	562	0	0
51	10	28124792	29515369	1390.577	475	0	0
52	10	31996493	33760979	1764.486	595	0	0
53	10	34392267	36408052	2015.785	809	0	0
54	10	36502029	38717701	2215.672	895	0	0
55	10	55367240	56366064	998.824	287	0	0
56	10	59787272	60351460	564.188	185	0	0
57	11	6499591	7079057	579.466	61	0	0
58	11	12855529	16098606	3243.077	1037	0	0
59	11	16172062	16842079	670.017	224	0	0
60	11	21660989	23864483	2203.494	783	0	0
61	11	24826348	25560350	734.002	320	0	0
62	11	30584895	32936491	2351.596	701	0	0
63	12	34982850	36382856	1400.006	480	0	0
64	12	45240807	50249926	5009.119	1730	0	0
65	12	50313458	53696292	3382.834	1269	0	0
66	12	54962035	57286529	2324.494	1066	0	0
67	12	59116763	60419336	1302.573	498	0	0
68	12	61333142	62359206	1026.064	387	0	0
69	12	64507422	65420039	912.617	365	0	0
70	12	66032718	66790820	758.102	224	0	0
71	13	18992032	19594843	602.811	141	0	0
72	13	20719697	22243683	1523.986	533	0	0
73	13	24084111	25052249	968.138	363	0	0
74	13	46756485	48071904	1315.419	511	0	0
75	13	48775052	50066030	1290.978	357	0	0
76	13	50626217	51169600	543.383	145	0	0
77	14	20619055	22766886	2147.831	828	0	0
78	14	24219701	25344261	1124.56	368	0	0
79	14	25876683	27021819	1145.136	392	0	0
80	14	29930169	30539210	609.041	59	0	0
81	14	30869401	31717098	847.697	146	0	0
82	14	35353338	37045410	1692.072	261	0	0
83	15	27409352	28085439	676.087	336	0	0
84	15	42383493	43300206	916.713	195	0	0
85	15	44523712	45463834	940.122	216	0	0
86	15	47469953	48956899	1486.946	424	0	0

Supplementary information S14 (continued)

87	16	22723968	23383314	659.346	215	0	0
88	16	38077984	41535018	3457.034	1392	0	0
89	16	42286150	47086319	4800.169	1971	0	0
90	16	52531106	55054556	2523.45	1052	0	0
91	16	59043386	61475372	2431.986	717	0	0
92	16	66985087	67547738	562.651	139	0	0
93	16	68078277	68793189	714.912	121	0	0
94	17	34910216	35918985	1008.769	309	0	0
95	17	39332230	42501914	3169.684	1080	0	0
96	17	61533393	62224049	690.656	245	0	0
97	17	62810946	63800370	989.424	409	0	0
98	17	66305795	68786941	2481.146	999	0	0
99	17	69296839	70161179	864.34	291	0	0
100	17	74491527	75689805	1198.278	441	0	0
101	17	90326033	91166068	840.035	63	0	0
102	18	13731709	15288872	1557.163	359	0	0
103	18	16175426	17066735	891.309	249	0	0
104	18	17168455	17704307	535.852	116	0	0
105	18	19233887	20270247	1036.36	325	0	0
106	18	22509956	23982294	1472.338	576	0	0
107	18	25809497	29181388	3371.891	1268	0	0
108	18	31158498	34111280	2952.782	996	0	0
109	18	41783538	42890910	1107.372	365	0	0
110	18	42986383	43556021	569.638	153	0	0
111	18	45397782	47433023	2035.241	674	0	0
112	18	47490558	48941445	1450.887	445	0	0
113	18	65304315	65977920	673.605	93	0	0
114	19	9963655	15643066	5679.411	1525	0	0
115	19	25663324	28022507	2359.183	816	0	0
116	19	29330569	30248983	918.414	291	0	0
117	19	30766731	31801808	1035.077	406	0	0
118	19	32970140	37766982	4796.842	1726	0	0
119	19	47660567	48176963	516.396	182	0	0
120	19	51244989	56094488	4849.499	1331	0	0
121	20	11280951	15420845	4139.894	1895	0	0
122	20	20294061	20890379	596.318	367	0	0
123	20	20996564	23444955	2448.391	1144	0	0
124	20	25271894	26167797	895.903	320	0	0
125	20	38715953	39349819	633.866	75	0	0
126	20	39410181	40242439	832.258	121	0	0
127	21	22775862	24232900	1457.038	697	0	0
128	21	43982326	46649829	2667.503	835	0	0
129	21	54876416	55523624	647.208	137	0	0
130	22	198279	858743	660.464	152	0	0
131	22	4683066	7471413	2788.347	652	0	0
132	22	10644372	13548835	2904.463	1053	0	0
133	22	14126878	19370628	5243.75	2025	0	0
134	22	19877998	21476613	1598.615	487	0	0
135	23	23234992	24899842	1664.85	426	0	0
136	23	33396585	34813889	1417.304	487	0	0
137	23	51259465	51984180	724.715	158	0	0
138	24	26890149	27490141	599.992	181	0	0
139	24	29575248	30114758	539.51	112	0	0
140	24	30391771	31124998	733.227	164	0	0
141	25	39879772	40519251	639.479	68	0	0
142	26	15800693	17398925	1598.232	554	0	0
143	26	20093536	23964494	3870.958	1343	0	0
144	26	25259480	26680562	1421.082	561	0	0
145	26	26739010	27941285	1202.275	550	0	0
146	26	29374165	30312829	938.664	277	0	0
147	26	34445265	35484131	1038.866	458	0	0
148	26	35534222	37495791	1961.569	738	0	0
149	26	37551099	38155907	604.808	167	0	0
150	27	40516992	41552222	1035.23	165	0	0
151	28	18909044	21387594	2478.55	937	0	0
152	28	27714323	30493191	2778.868	924	0	0
153	28	33597464	34255823	658.359	115	0	0
154	31	15554695	16232561	677.866	123	0	0
155	32	19112340	20576747	1464.407	432	0	0
156	32	33977829	34907084	929.255	314	0	0
157	32	35867706	36469478	601.772	125	0	0
158	32	36548045	37193446	645.401	65	0	0

Supplementary Informations S15. Listing of heterozygous regions (i.e., large region without ROH) in HA population.

region	Chromosome	Start (bp)	End (bp)	Size (kb)	number of SNPs	minimal number of individual with a ROH	maximum number of individual with a ROH
1	1	6552905	7093199	540.294	94	0	0
2	1	14182285	14919737	737.452	216	0	0
3	1	71493819	73274631	1780.812	542	0	0
4	1	73325555	75158337	1832.782	620	0	0
5	1	78571424	79120029	548.605	157	0	0
6	1	80802453	82212217	1409.764	337	0	0
7	1	82292185	83131310	839.125	192	0	0
8	1	83185373	84368100	1182.727	312	0	0
9	2	15593303	16329415	736.112	120	0	0
10	2	25240773	30181562	4940.789	1612	0	0
11	2	31304892	34848818	3543.926	1404	0	0
12	2	57441101	58559775	1118.674	353	0	0
13	2	59336589	61212536	1875.947	577	0	0
14	2	62058830	64840544	2781.714	929	0	0
15	2	65470591	67092904	1622.313	528	0	0
16	2	68517042	69416672	899.63	267	0	0
17	2	70063894	72076148	2012.254	804	0	0
18	2	73196639	75969389	2772.75	808	0	0
19	2	80002034	81762118	1760.084	492	0	0
20	2	82471541	84537874	2066.333	708	0	0
21	2	85181564	93020279	7838.715	2555	0	0
22	2	93074048	95759804	2685.756	561	0	0
23	3	9294197	10100470	806.273	203	0	0
24	3	12637565	14770578	2133.013	623	0	0
25	3	15608693	19188357	3579.664	839	0	0
26	3	19242365	20885903	1643.538	447	0	0
27	3	21435169	27425742	5990.573	2561	0	0
28	3	29009254	30164729	1155.475	468	0	0
29	3	30673080	33962864	3289.784	1243	0	0
30	3	42335393	43008651	673.258	256	0	0
31	3	43591501	48169605	4578.104	1614	0	0
32	3	49625770	51750003	2124.233	1023	0	0
33	3	59804061	62021736	2217.675	945	0	0
34	3	63232936	63924361	691.425	186	0	0
35	3	67836885	68611901	775.016	202	0	0
36	3	68747748	70120120	1372.372	359	0	0
37	3	71638415	72272914	634.499	177	0	0
38	3	75110357	77245303	2134.946	526	0	0
39	3	77852919	78596941	744.022	96	0	0
40	4	11361036	14134115	2773.079	1068	0	0
41	4	14188883	15852181	1663.298	722	0	0
42	5	8865413	11118443	2253.03	643	0	0
43	5	11935182	13152680	1217.498	412	0	0
44	5	13834919	14834799	999.88	383	0	0
45	5	16287482	16832332	544.85	152	0	0
46	5	16888538	17649113	760.575	204	0	0
47	5	22800586	24471982	1671.396	542	0	0
48	5	30629677	31172511	542.834	197	0	0
49	5	39085967	41259616	2173.649	1110	0	0
50	5	44455042	45191673	736.631	347	0	0
51	6	35113463	36323550	1210.087	490	0	0
52	6	36415093	38889129	2474.036	870	0	0
53	6	50056318	50660044	603.726	199	0	0
54	6	50744796	52108288	1363.492	373	0	0
55	6	52649148	54021483	1372.335	389	0	0
56	6	55307612	61057694	5750.082	2143	0	0
57	6	65406493	66074424	667.931	231	0	0
58	6	77461531	78996692	1535.161	357	0	0
59	7	58199682	59700368	1500.686	502	0	0
60	7	60672762	63046233	2373.471	678	0	0
61	7	63105828	63759474	653.646	206	0	0
62	7	71107461	72830716	1723.255	506	0	0
63	7	74898571	75818371	919.8	240	0	0
64	7	75870135	76737277	867.142	248	0	0
65	7	77313239	80200335	2887.096	791	0	0
66	7	80255995	80931254	675.259	101	0	0
67	8	22953729	24821667	1867.938	672	0	0
68	8	25597752	29033805	3436.053	1138	0	0
69	8	30078901	33995476	3916.575	1314	0	0
70	8	35263577	35905058	641.481	247	0	0

Supplementary information S15 (continued)

71	8	49146433	49892228	745.795	266	0	0
72	8	50583421	52480916	1897.495	735	0	0
73	8	52744409	54887958	2143.549	881	0	0
74	8	61631866	63976651	2344.785	1010	0	0
75	8	64554878	66732042	2177.164	974	0	0
76	8	68156291	69288533	1132.242	441	0	0
77	8	70029742	71147330	1117.588	422	0	0
78	8	71217118	73603443	2386.325	852	0	0
79	8	74803457	75871447	1067.99	329	0	0
80	8	75931361	77378125	1446.764	281	0	0
81	8	78185995	80190455	2004.46	475	0	0
82	9	13218724	14906289	1687.565	356	0	0
83	9	15594228	16133523	539.295	122	0	0
84	9	16398944	17073824	674.88	154	0	0
85	9	17507844	19119585	1611.741	392	0	0
86	9	19564644	23218063	3653.419	1284	0	0
87	9	23826948	25270270	1443.322	571	0	0
88	9	25961818	28125052	2163.234	866	0	0
89	9	43609604	44208459	598.855	270	0	0
90	9	44734498	47093107	2358.609	775	0	0
91	9	47620754	53170806	5550.052	1995	0	0
92	9	53775888	56794059	3018.171	1118	0	0
93	10	8236427	10151929	1915.502	353	0	0
94	10	12947668	14061420	1113.752	355	0	0
95	10	17159164	18169068	1009.904	275	0	0
96	10	19097028	22587936	3490.908	1119	0	0
97	10	23128459	23683915	555.456	197	0	0
98	10	25254815	29433437	4178.622	1568	0	0
99	10	31881719	33397309	1515.59	524	0	0
100	10	56314062	56956427	645.232	169	0	0
101	10	59210021	59733141	523.12	141	0	0
102	10	61534151	63759768	2225.617	511	0	0
103	10	64835120	66228108	1392.988	368	0	0
104	10	68070942	68877741	806.799	186	0	0
105	11	10188768	12146524	1957.756	520	0	0
106	11	14297456	16098606	1801.15	596	0	0
107	11	16172062	17426956	1254.894	483	0	0
108	11	19394852	21489030	2094.178	778	0	0
109	11	23150104	25961802	2811.698	1044	0	0
110	11	27060937	28939858	1878.921	566	0	0
111	11	32292792	34711564	2418.772	766	0	0
112	11	34780145	35626913	846.768	219	0	0
113	11	43019084	43581200	562.116	166	0	0
114	11	45053522	46281673	1228.151	386	0	0
115	11	58168153	58737091	568.938	265	0	0
116	11	58798610	59872527	1073.917	448	0	0
117	11	60408385	61007517	599.132	213	0	0
118	11	61063063	62410762	1347.699	450	0	0
119	11	63009737	64390653	1380.916	391	0	0
120	11	66161065	66892599	731.534	270	0	0
121	11	67514110	70041255	2527.145	744	0	0
122	11	70847151	71390575	543.424	93	0	0
123	12	7912689	8473625	560.936	105	0	0
124	12	11081860	12037381	955.521	228	0	0
125	12	13428683	14099510	670.827	198	0	0
126	12	16600007	17948918	1348.911	379	0	0
127	12	45440159	46315421	875.262	270	0	0
128	12	47609696	49995491	2385.795	909	0	0
129	12	50804086	53600704	2796.618	1056	0	0
130	12	54211546	54853854	642.308	269	0	0
131	12	55463344	56271859	808.515	392	0	0
132	12	61498677	62137394	638.717	271	0	0
133	12	62714439	63566242	851.803	297	0	0
134	12	65978161	74844774	8866.613	3323	0	0
135	12	76431588	78215528	1783.94	722	0	0
136	12	79014420	81115945	2101.525	639	0	0
137	12	83050923	83932712	881.789	174	0	0
138	13	15370611	17227934	1857.323	301	0	0
139	13	22286711	22811866	525.155	201	0	0
140	13	23318787	24331102	1012.315	428	0	0
141	13	42608820	43205861	597.041	276	0	0
142	13	45148833	45742342	593.509	268	0	0
143	13	46959537	48199678	1240.141	480	0	0
144	14	3921454	5999055	2077.601	714	0	0
145	14	6552219	10099915	3547.696	1134	0	0

Supplementary information S15 (continued)

146	14	11222214	18813011	7590.797	2686	0	0
147	14	35353338	36642768	1289.43	205	0	0
148	15	8227290	9170115	942.825	111	0	0
149	15	9416836	10062990	646.154	80	0	0
150	15	14563121	15322042	758.921	224	0	0
151	15	16598439	17212115	613.676	213	0	0
152	15	18243879	19731495	1487.616	590	0	0
153	15	28160422	30607544	2447.122	800	0	0
154	15	43832360	48956899	5124.539	1337	0	0
155	15	49043502	51036178	1992.676	725	0	0
156	15	56256591	60639433	4382.842	1557	0	0
157	16	44351305	45584171	1232.866	412	0	0
158	17	21359365	23244499	1885.134	493	0	0
159	17	23317779	24856514	1538.735	488	0	0
160	17	25779383	26388685	609.302	185	0	0
161	17	27653117	32665761	5012.644	1784	0	0
162	17	33782953	35536469	1753.516	580	0	0
163	17	40950947	41771889	820.942	237	0	0
164	17	48736646	49509033	772.387	274	0	0
165	17	55172164	59504036	4331.872	1801	0	0
166	17	59626461	62064670	2438.209	851	0	0
167	17	73890492	75992699	2102.207	735	0	0
168	18	15697646	17065357	1367.711	384	0	0
169	18	17927574	18679323	751.749	144	0	0
170	18	19575860	22972998	3397.138	1127	0	0
171	18	25419826	26293462	873.636	308	0	0
172	18	27071787	28333048	1261.261	491	0	0
173	18	32039677	35254426	3214.749	1038	0	0
174	18	35909314	40129393	4220.079	1394	0	0
175	18	40277544	41584395	1306.851	568	0	0
176	18	41643394	42170596	527.202	180	0	0
177	18	64648185	66368574	1720.389	283	0	0
178	19	10209615	19365302	9155.687	2338	0	0
179	19	19448588	21605967	2157.379	661	0	0
180	19	45793006	47608852	1815.846	810	0	0
181	19	47660567	48176963	516.396	182	0	0
182	19	48235900	49482481	1246.581	367	0	0
183	19	54193274	55928746	1735.472	622	0	0
184	19	58389489	59374453	984.964	236	0	0
185	20	373366	1218393	845.027	187	0	0
186	20	1331764	3237918	1906.154	292	0	0
187	20	6093106	6908663	815.557	312	0	0
188	20	22095357	23784167	1688.81	746	0	0
189	20	24299458	25301089	1001.631	222	0	0
190	20	26176102	28688352	2512.25	1027	0	0
191	20	33548049	36578562	3030.513	854	0	0
192	20	38086839	38656901	570.062	104	0	0
193	20	38715953	39349819	633.866	75	0	0
194	21	50144477	53120948	2976.471	1008	0	0
195	22	198279	1110770	912.491	193	0	0
196	23	14349333	16464599	2115.266	787	0	0
197	23	17015896	18674677	1658.781	568	0	0
198	23	21299369	21851397	552.028	147	0	0
199	23	37769421	38479484	710.063	209	0	0
200	23	38995905	43439910	4444.005	1267	0	0
201	23	44228632	45383236	1154.604	284	0	0
202	23	47432580	48188646	756.066	174	0	0
203	23	48252224	48769186	516.962	140	0	0
204	23	51416882	51984180	567.298	127	0	0
205	23	52042543	52586447	543.904	88	0	0
206	23	53537923	54935459	1397.536	300	0	0
207	23	56575062	57525412	950.35	146	0	0
208	24	1512015	4458288	2946.273	1288	0	0
209	24	4561302	11852218	7290.916	2622	0	0
210	24	12033368	12885427	852.059	342	0	0
211	24	13935958	15096589	1160.631	324	0	0
212	24	16156796	18511203	2354.407	870	0	0
213	24	20326336	24911570	4585.234	1449	0	0
214	24	25566149	26163785	597.636	148	0	0
215	24	26890149	27548969	658.82	185	0	0
216	24	27620214	28328024	707.81	104	0	0
217	24	28418572	28945934	527.362	110	0	0
218	24	29588649	30114758	526.109	109	0	0
219	24	30391771	31124998	733.227	164	0	0
220	25	13449914	19519550	6069.636	1986	0	0

Supplementary information S15 (continued)

221	25	23373628	25998862	2625.234	830	0	0
222	26	12045346	12705447	660.101	255	0	0
223	26	14632337	21355854	6723.517	2491	0	0
224	26	28610557	30312829	1702.272	624	0	0
225	26	30367194	31077047	709.853	192	0	0
226	26	31656514	32972225	1315.711	483	0	0
227	26	33146667	34209627	1062.96	455	0	0
228	27	8425031	10451035	2026.004	851	0	0
229	27	14666566	19595960	4929.394	1659	0	0
230	27	37731330	38346183	614.853	88	0	0
231	28	1796569	4011700	2215.131	752	0	0
232	28	6840452	8522421	1681.969	579	0	0
233	28	8584013	9946910	1362.897	422	0	0
234	28	11806899	12630116	823.217	339	0	0
235	28	15340817	19427768	4086.951	1367	0	0
236	28	21043258	22819475	1776.217	626	0	0
237	28	27761168	29524127	1762.959	586	0	0
238	28	30415774	35120554	4704.78	1381	0	0
239	28	35371576	36218106	846.53	186	0	0
240	29	56625	649552	592.927	161	0	0
241	29	7051533	9565087	2513.554	832	0	0
242	29	10225665	10898695	673.03	232	0	0
243	29	11590910	12550658	959.748	389	0	0
244	29	13412254	16929331	3517.077	1222	0	0
245	29	29098222	31312506	2214.284	902	0	0
246	29	31385549	33097877	1712.328	691	0	0
247	29	33168403	34645288	1476.885	458	0	0
248	29	36344967	36979026	634.059	199	0	0
249	30	231533	1284990	1053.457	81	0	0
250	30	6361042	7098173	737.131	117	0	0
251	30	7535490	8818653	1283.163	212	0	0
252	30	13759867	14268443	508.576	225	0	0
253	30	32768582	34358699	1590.117	382	0	0
254	31	26786964	27428768	641.804	141	0	0
255	31	27583087	28103070	519.983	161	0	0
256	31	28660091	30649939	1989.848	707	0	0
257	31	31011378	32713851	1702.473	536	0	0
258	31	32805594	34523365	1717.771	775	0	0
259	32	11157219	12924605	1767.386	290	0	0
260	32	13669546	14363814	694.268	134	0	0
261	32	16487584	17149054	661.47	166	0	0
262	32	19303112	22390203	3087.091	1013	0	0
263	32	26426712	28687441	2260.729	799	0	0
264	32	28785720	29576825	791.105	251	0	0
265	32	31297745	35193754	3896.009	1438	0	0

Supplementary Information S16. Listing of the 29 genes which composed the four heterozygous genomic regions.

Region	Chromosome	Start (bp)	End (bp)	GeneID_NCBI	Gene name	Protein_name	
chr10_a	10	56251640	56645651	ctna2	<i>ctna2</i>	catenin alpha 2	
	13	46941309	47041211	LOC110485195	<i>rbfox3</i>	RNA binding protein fox-1 homolog 3-like	
	13	47395433	47400215	LOC110486591	<i>enpp7</i>	ectonucleotide pyrophosphatase/phosphodiesterase family member 7	
	13	47403950	47407911	LOC110486592	<i>cbx2</i>	chromobox protein homolog 2	
	13	47432794	47433462	LOC110486593	<i>nog3</i>	noggin-3-like	
	13	47461417	47466583	LOC110486594	<i>pctp</i>	phosphatidylcholine transfer protein	
	13	47469699	47470130	tmem100a	<i>tmem100a</i>	transmembrane protein 100	
	13	47482712	47504517	LOC110486595	<i>mmd</i>	monocyte to macrophage differentiation factor	
	13	47511898	47522430	LOC110486596	<i>hlf</i>	hepatic leukemia factor	
	13	47566915	47597083	LOC110486597	<i>foxk2</i>	forkhead box protein K2	
	13	47612912	47613316	LOC110487550	<i>pirt</i>	phosphoinositide-interacting protein	
	13	47639788	47706368	LOC110486600	<i>shisa6</i>	protein shisa-6	
	13	47806788	47814889	LOC110486601	<i>elac2</i>	zinc phosphodiesterase ELAC protein 2	
	chr13_a	13	47818946	47830360	LOC110486602	<i>map2k4</i>	dual specificity mitogen-activated protein kinase kinase 4
		13	47831328	47832595	LOC110486605	<i>tmem220</i>	transmembrane protein 220
		13	47833460	47835282	LOC110486603	<i>sco1</i>	protein SCO1 homolog, mitochondrial
		13	47875547	47911501	LOC110486607	<i>gas7</i>	growth arrest-specific protein 7
13		47920503	47921621	LOC110486610	<i>rcvrn</i>	recoverin-like	
13		47927253	47932236	LOC110486609	<i>gsg1l</i>	germ cell-specific gene 1-like protein	
13		47933613	47934227	LOC110486608	<i>hs3st3b1</i>	heparan sulfate glucosamine 3-O-sulfotransferase 3B1-like	
13		47948696	47950307	LOC110486613	<i>dhrrs7c</i>	dehydrogenase/reductase SDR family member 7C-B	
13		47953208	47958915	LOC110486612	<i>u2af2</i>	splicing factor U2AF 65 kDa subunit	
13		47962136	47983861	LOC110486615	<i>mprip</i>	myosin phosphatase Rho-interacting protein	
13		47986112	47988937	LOC110486616	<i>tcf20</i>	transcription factor 20	
13		48002720	48010337	LOC110486618	<i>sbk1</i>	serine/threonine-protein kinase SBK1	
13	48042698	48071525	LOC110486619	<i>radil</i>	ras-associating and dilute domain-containing protein		
chr19_a	19	10749245	10758666	LOC110498644	<i>smarca5</i>	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	
	19	10760826	10811242	LOC110498645	<i>frem2</i>	FRAS1-related extracellular matrix protein 2	
chr19_b	19	11399493	11400947	LOC110497370	<i>pou4f2</i>	POU domain, class 4, transcription factor 2-like	

Supplementary information S17. Percentage of protein homology between rainbow trout and nine other vertebrate species for all genes annotated in heterozygous regions chr10_a, chr13_a, chr19_a and chr19_b.

	gene_ID	Human	Mouse	Goat	Cattle	Pig	Chicken	Zebrafish	Medaka	Atlantic salmon	
chr10_a	<i>ctma2</i>	95.47	95.58	95.43	91.53	91.53	95.36	98.34	98.01	99.89	
	<i>rbfox</i>	76.36	74.54	75.25	75.59	76.00	74.14	80.05	95.91	97.55	
	<i>enpp7</i>	62.06	58.66	58.42	58.57	57.67	61.48	70.46	72.59	98.53	
	<i>cbx2</i>	33.12	35.48	40.23	39.65	39.42	44.29	54.98	51.23	94.78	
	<i>nog3</i>	56.89	56.89	57.33	57.33	56.89	61.16	69.2	72.77	97.75	
	<i>pctp</i>	66.47	69.94	65.32	64.16	66.67	72.16	76.37	79.44	98.91	
	<i>tmem100a</i>	65.22	63.04	62.77	62.32	63.77	68.12	78.17	73.19	97.90	
	<i>mmd</i>	86.84	86.4	85.96	85.53	86.4	82.48	88.09	87.61	99.59	
	<i>hlf</i>	70.23	71.06	71.10	69.58	69.9	73.15	74.42	77.74	95.02	
	<i>foxx2</i>	75.20	75.57	75.45	75.61	74.14	76.46	74.71	79.60	96.78	
	<i>pirt</i>	-	-	-	-	-	-	-	-	64.71	100
	<i>shisa6</i>	49.47	50.85	48.75	49.64	51.13	52.99	78.63	65.55	98.97	
	<i>elac2</i>	61.17	60.37	60.91	61.43	60.39	57.68	68.95	70.95	96.91	
	chr13_a	<i>map2k4</i>	89.84	90.3	88.53	89.84	89.84	97.97	91.67	90.62	99.76
<i>tmem220</i>		45.96	42.50	37.97	43.75	45.00	43.75	61.18	64.60	94.29	
<i>scol</i>		69.27	73.15	72.43	72.43	71.96	66.36	79.44	81.19	98.36	
<i>gas7</i>		69.91	71.09	70.84	70.38	71.5	71.02	84.10	71.43	97.87	
<i>rcvrn</i>		47.09	51.28	42.4	48.15	41.27	48.68	84.90	88.54	98.96	
<i>gsg11</i>		43.92	48.51	45.25	45.25	44.44	50.72	70.73	68.05	97.08	
<i>hs3st3b1</i>		70.70	70.10	69.12	68.63	68.63	71.08	78.92	76.47	99.02	
<i>dhrs7c</i>		55.56	54.07	54.98	54.81	53.33	56.09	63.33	65.17	96.96	
<i>u2af2</i>		79.20	79.00	78.95	79.20	78.66	47.47	80.31	78.25	98.5	
<i>mprip</i>		44.23	44.73	45.38	65.24	59.56	65.55	68.30	65.85	93.73	
<i>tcf20</i>		35.38	35.59	35.78	35.29	34.9	37.26	44.81	68.87	93.94	
<i>sbk1</i>		64.56	64.56	64.56	64.56	64.91	55.2	77.12	80.36	97.88	
<i>radil</i>		37.89	36.94	37.19	37.07	40.11	38.64	65.88	68.07	95.95	
chr19_a		<i>smarca5</i>	88.86	88.65	88.75	88.75	88.75	89.67	91.37	92.15	99.61
	<i>frem2</i>	67.63	65.73	66.27	67.63	66.65	67.74	84.22	83.71	98.51	
chr19_b	<i>pou4f2</i>	89.23	89.23	91.88	89.23	89.54	85.02	92.97	91.21	100.00	



Development of a High-Density 665 K SNP Array for Rainbow Trout Genome-Wide Genotyping

Maria Bernard^{1,2}, Audrey Dehaullon¹, Guangtu Gao³, Katy Paul¹, Henri Lagarde¹, Mathieu Charles^{1,2}, Martin Prchal⁴, Jeanne Danon⁵, Lydia Jaffrelo⁵, Charles Poncet⁵, Pierre Patrice⁶, Pierrick Haffray⁶, Edwige Quillet¹, Mathilde Dupont-Nivet¹, Yniv Palti³, Delphine Lallias¹ and Florence Phocas^{1*}

¹INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas, France, ²INRAE, SIGENAE, Jouy-en-Josas, France, ³USDA, REE, ARS, NEA, NCCCWA, Kearneysville, WV, United States, ⁴South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Faculty of Fisheries and Protection of Waters, University of South Bohemia, Vodňany, Czechia, ⁵INRAE-UCA, Plateforme Gentyane, UMR GDEC, Clermont-Ferrand, France, ⁶SYSAAF, Campus de Beaulieu, Rennes, France

OPEN ACCESS

Edited by:

Nguyen Hong Nguyen,
University of the Sunshine Coast,
Australia

Reviewed by:

James W. Kijas,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia
Tereza Manousaki,
Hellenic Centre for Marine Research
(HCMR), Greece

*Correspondence:

Florence Phocas
florence.phocas@inrae.fr

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 11 May 2022

Accepted: 24 June 2022

Published: 18 July 2022

Citation:

Bernard M, Dehaullon A, Gao G,
Paul K, Lagarde H, Charles M,
Prchal M, Danon J, Jaffrelo L,
Poncet C, Patrice P, Haffray P,
Quillet E, Dupont-Nivet M, Palti Y,
Lallias D and Phocas F (2022)
Development of a High-Density
665 K SNP Array for Rainbow Trout
Genome-Wide Genotyping.
Front. Genet. 13:941340.
doi: 10.3389/fgene.2022.941340

Single nucleotide polymorphism (SNP) arrays, also named « SNP chips », enable very large numbers of individuals to be genotyped at a targeted set of thousands of genome-wide identified markers. We used preexisting variant datasets from USDA, a French commercial line and 30X-coverage whole genome sequencing of INRAE isogenic lines to develop an Affymetrix 665 K SNP array (HD chip) for rainbow trout. In total, we identified 32,372,492 SNPs that were polymorphic in the USDA or INRAE databases. A subset of identified SNPs were selected for inclusion on the chip, prioritizing SNPs whose flanking sequence uniquely aligned to the Swanson reference genome, with homogenous repartition over the genome and the highest Minimum Allele Frequency in both USDA and French databases. Of the 664,531 SNPs which passed the Affymetrix quality filters and were manufactured on the HD chip, 65.3% and 60.9% passed filtering metrics and were polymorphic in two other distinct French commercial populations in which, respectively, 288 and 175 sampled fish were genotyped. Only 576,118 SNPs mapped uniquely on both Swanson and Arlee reference genomes, and 12,071 SNPs did not map at all on the Arlee reference genome. Among those 576,118 SNPs, 38,948 SNPs were kept from the commercially available medium-density 57 K SNP chip. We demonstrate the utility of the HD chip by describing the high rates of linkage disequilibrium at 2–10 kb in the rainbow trout genome in comparison to the linkage disequilibrium observed at 50–100 kb which are usual distances between markers of the medium-density chip.

Keywords: SNP, single nucleotide polymorphism, sequence, high-density chip, linkage disequilibrium, rainbow trout, doubled haploid lines, isogenic lines

INTRODUCTION

Next-generation sequencing (NGS) has transformed the fields of quantitative, ecological and evolutionary genetics by enabling the discovery and cost-effective genotyping of thousands to millions of variants across the genome, allowing for genome-wide association studies (GWAS) of complex traits, genomic selection (GS) through accurate inference of relationships among individuals (Meuwissen and Goddard, 2010), inbreeding (Kardos et al., 2015), population

structure, and genetic diversity studies. Large numbers of densely genotyped individuals are required to get accurate results thanks to a high SNP density along the genome that constructs strong linkage disequilibrium between SNP and causative mutations (de Roos et al., 2008). However, regardless of the animal or plant species, it remains very challenging to cost-effectively genotype large numbers of individuals at polymorphic sites in all the genomes. An appealing strategy is to use a cheaper and reduced-density SNP chip with markers being chosen for optimizing the imputation accuracy to higher density genotypes. Genotype imputation describes the process of predicting genotypes that are not directly assayed in a sample of individuals (Marchini and Howie, 2010). Imputation has become a standard practice in research to increase genome coverage and improve GS accuracy and GWAS resolution, as a large number of samples can be genotyped at lower density (and lower cost) then imputed up to denser marker panels or to sequence level, using information from a limited reference population (Phocas, 2022).

Two main methods are employed for large-scale and genome-wide SNP genotyping. Array-based methods use flanking probe sequences to interrogate pre-identified SNPs (often named “SNP chips”). The alternative genotyping-by-sequencing (GBS) methods call SNPs directly from the genome (Davey et al., 2011). In GBS methods, either restriction enzymes are used to target sequencing resources on a limited number of cut sites (Baird et al., 2008) or low-coverage whole genome resequencing is performed. Low-coverage GBS followed by imputation has been proposed as a cost-effective genotyping approach for human genetics studies (Pasaniuc et al., 2012), as well as farmed species (Gorjanc et al., 2017) that cannot afford a high development of genomic tools. Nevertheless, compared to GBS methods, SNP chips offer a robust and easily replicable way of genotyping samples at a consistent set of SNPs, with very low rates of missing data.

Medium (~thousands to tens of thousands of loci) and high (~hundreds of thousands of loci) density SNP chips have been routinely developed for commercial species to perform genomic selection (Meuwissen et al., 2001) and to identify genes playing significant roles in livestock and crop performances (Goddard et al., 2016). SNP chips developed for model organisms or farmed species have also been utilised to address evolutionary and conservation questions, in particular in animal populations. For example, they have been used to identify signatures of adaptation in cattle (Gautier et al., 2010) or genes under selection in grey wolves (Schweizer et al., 2016), characterize the genetic diversity and inbreeding levels in pig (Silió et al., 2013), sheep (Mastrangelo et al., 2014), cattle (Rodríguez-Ramilo et al., 2015) or fish (D’Ambrosio et al., 2019), and infer the genomic basis of recombination rate variation in cattle (Sandor et al., 2012) or sheep (Johnston et al., 2016; Petit et al., 2017).

While there is now over ten fish and shellfish species for which commercial SNP arrays had been developed (Boudry et al., 2021), most of those contain only about 50 to 60 K SNPs. Such medium-density chips are sufficient for genomic selection purposes but are clearly too low-density tools for fine QTL detection and help in identification of causal variants. As

rainbow trout (*Oncorhynchus mykiss*) is a major academic model for a wide range of investigations in disciplines such as cancer research, toxicology, immunology, physiology, nutrition, developmental or evolutionary biology in addition to quantitative genetics and breeding (Thorgaard et al., 2002), it is important to get access to very high-density genomic tools for this salmonid species.

For rainbow trout, SNP discovery has been firstly done through sequencing of restriction-site associated DNA (RAD) libraries (Palti et al., 2014), reduced representation libraries (RRL) (Sánchez et al., 2009), and RNA sequencing (Sánchez et al., 2011). A first commercial medium-density Axiom® Trout Genotyping array (hereafter termed 57 K chip) has then been developed (Palti et al., 2015) and produced by Affymetrix (ThermoFisher). Since then it has been largely used in population genetics studies (Larson et al., 2018; D’Ambrosio et al., 2019; Paul et al., 2021), GWAS and GS accuracy works for various traits in farmed populations (Gonzalez-Pena et al., 2016; Vallejo et al., 2017; Vallejo et al., 2019; Reis Neto et al., 2019; Rodríguez et al., 2019; Yoshida et al., 2019; Frasin et al., 2019; Frasin et al., 2020; Karami et al., 2020; D’Ambrosio et al., 2020; Blay et al., 2021a; Blay et al., 2021b). However, out of the 57,501 SNPs included in this chip, nearly 20,000 were found to be unusable because they were either duplicated due to the ancestral genome duplication or showing primer polymorphism in five French commercial or experimental lines (D’Ambrosio et al., 2019). Of the 57,501 markers from the original chip, 50,820 are uniquely localized on the Swanson reference genome (Pearse et al., 2019), and in the remaining number, only 38,332 markers pass the control quality filters [no primer polymorphism, call rate > 97%, Minor Allele Frequency (MAF) > 0.001 over 3,000 fish from five French lines].

To overcome these limitations as well as to get access to a more powerful tool for GWAS and population genetics studies in rainbow trout, the aim of our study was to develop a high-density SNP array. To develop this resource for rainbow trout, we made use of a large set of resequencing data from 31 doubled haploid (DH) lines from Washington State University (WSU) and Institut National de la Recherche pour l’Agriculture, l’Alimentation et l’Environnement (INRAE). In the United States, 12 WSU DH lines have been created by androgenesis (Young et al., 1996) while in France 19 DH INRAE lines (called isogenic lines) were produced by gynogenesis (Quillet et al., 2007). The 12 WSU DH lines as well as seven of the INRAE isogenic lines served as basic material for the variant search and SNP selection for the 57 K chip (Palti et al., 2015).

In this study, we describe how we overcame the limitations of duplications in the rainbow trout genome, in order to identify and locate polymorphisms. We describe the subset of detected SNPs that was selected for inclusion on a custom high-density SNP chip. It was used to genotype 463 samples from two different French commercial populations. We test the genotyping success rates, that is, the proportion of SNPs included on the array that are polymorphic and successfully genotyped. We demonstrate the utility of this SNP chip to infer linkage disequilibrium in the genome of this species.

MATERIALS AND METHODS

Use of the USDA Database for Initial SNP Detection

Gao et al. (2018) constituted a first large SNP database (USDA1) by performing high coverage whole genome resequencing (WGS) with 61 unrelated samples, representing a wide range of rainbow trout and steelhead populations. Of the 61 samples, 11 were doubled-haploid lines from Washington State University (WSU), 12 were aquaculture samples from AquaGen (Norway), 38 were from wild and hatchery populations from a wide range of geographic distribution (California, Oregon, Washington and Idaho states in the United States; Canada; Kamtchatka Peninsula in Russia). Overall, 31,441,105 SNPs were identified with 30,302,087 SNPs located on one of the 29 chromosomes of the Swanson reference genome assembly (Omyk_1.0; GenBank, assembly accession GCA_002163495.1) (Pearse et al., 2019).

A second database (USDA2) with 17,889,078 SNPs coming from resequencing of 24 USDA samples was added to the initial USDA1 database. The samples were composed of 12 representatives from the USDA-NCCCWA odd-year class and 12 from the even-year class as previously described (NCBI BioProject PRJNA681179; Liu et al., 2021). The SNP discovery analysis followed the methods of (Gao et al., 2018).

By merging these two databases using BCFtools (Danecek et al., 2021), we constituted a single USDA database that contained 35,732,342 distinct SNPs, with 34,170,401 placed on the 29 chromosomes or mitochondrial chromosome of the Swanson reference genome. SNP filtering was performed to remove non bi-allelic variants and SNPs with MAF <1% using a `vcf_filter.py` homemade script (**Supplementary File S1**). The final USDA clean database contained 29,024,315 SNPs.

Whole Genome Resequencing of INRAE Isogenic Lines and Use of the INRAE Database for SNP Detection

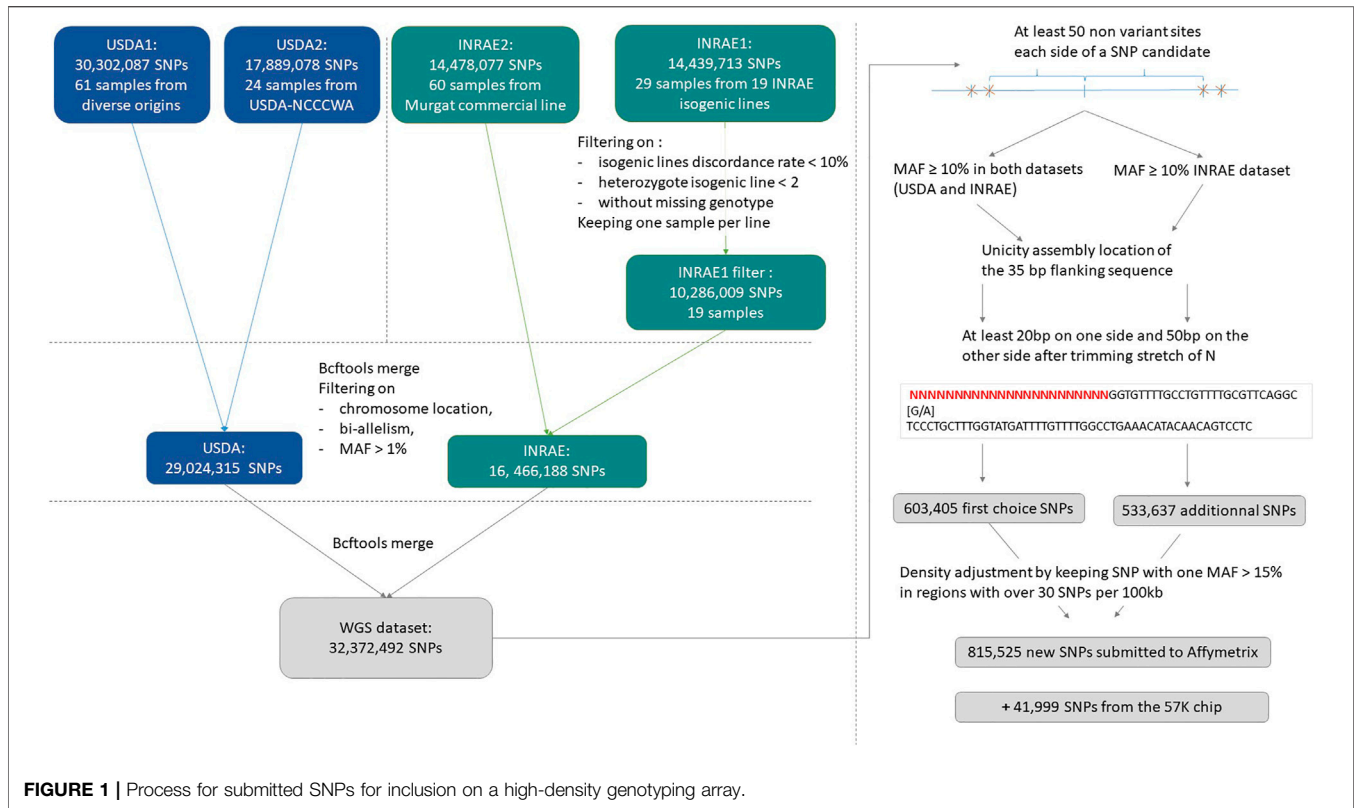
Genomic DNA was extracted from fin clips of 19 rainbow trout INRAE isogenic lines. Whole-genome paired-end sequencing libraries were prepared and sequenced using the Illumina HiSeq 2000, Hi Seq 3000 or HiSeq X-Ten platforms at a depth of genome coverage ranging from 10X to 32X per sample. The 19 isogenic lines were sequenced in two batches that were processed successively. The first batch contained sequencing data from 12 samples (doubled haploid individuals) coming from 11 isogenic lines. The second batch contained sequencing data from 17 samples (doubled haploid individuals) coming from 17 isogenic lines (9 lines already sequenced in batch 1; and 8 lines not previously sequenced). Overall, 10 out of the 19 isogenic lines were sequenced twice. This resulted in a total of 8,911,630,867 paired reads with a median of 321,575,464 per sample.

Sequence reads from each of the 12 samples from the first batch were mapped to the Swanson rainbow trout reference genome (GenBank assembly accession GCA_002163495.1;

Pearse et al., 2019) using BWA MEM v.0.7.12 (Li, 2013). We then ran Samtools sort [v1.3.1, (Danecek et al., 2021)] to sort the alignment data by chromosome and scaffold locations. Afterwards, PCR duplicates were marked using Picard Tools (v.2.1.1, Broad Institute, 2019) MarkDuplicates. Variant calling was then performed for each sample using GATK (v3.7; McKenna et al., 2010) HaplotypeCaller (options `-stand_call_conf 30 -mbq 10`), leading to 12 vcf files. A variant reference file containing 1,207,861 high quality SNPs was generated by keeping variants with $QUAL \geq 1,050$ from the vcf files. This file was then used for the recalibration step, using GATK BaseRecalibrator and PrintReads. The recalibrated BAM files were then used as input for the variant calling step using GATK HaplotypeCaller in ERC GVCF mode. The resulting 12 GVCF files were then merged into a single vcf file containing 24,944,575 variants using GATK GenotypeGVCFs. The vcf file was then filtered as follows using GATK VariantFiltration: $DP < 120$; $MQ < 30.0$; $QUAL < 600$; $AN < 12$. To filter out putative PSVs (Paralogous Sequence Variants), we filtered out variants with heterozygous genotypes in at least two of the 12 doubled haploid samples. The filtered vcf file from the first batch contained 11,113,836 variants.

The second samples sequence batch were analyzed following the same procedure as for the first batch with few updates. Prior to sequence alignment, sequences have been filtered using trimmomatic 0.36 (Bolger et al., 2014) to remove Illumina Truseq adapters, trim low quality bases, keep trimmed reads with a sufficient length and average quality. These parameters removed 3.8% of the reads, keeping 6,349,173,142 reads over the 17 samples. Alignment software was updated to use BWA MEM v.0.7.15. First calling to create a high quality variants set to recalibrate the BAM files was avoided by directly using the final vcf file from the first batch analysis. These recalibrated BAM have been submitted to GATK Haplotype caller as before to generate GVCF files. To increase confidence in the SNP calling, we also added 2 other SNP callers: Samtools mpileup and FreeBayes 1.1.0 (Garrison and Marth, 2012). GATK calling results were jointly genotyped using GATK GenotypeGVCFs on the 12 GVCF files from the first batch and the 17 newly generated GVCF files. This calling procedure resulted in 3 VCF files, one for each caller. Calling from GATK contained 29 samples (from the 19 isogenic lines, i.e., with 10 lines replicated) and 31,454,943 variants; Freebayes and Mpileup was used only on the second batch and contained 19 samples and 25,805,271 and 30,340,281 variants respectively.

The final step for variant calling was to intersect the 3 calling datasets using VCFtools_0.1.12a (Danecek et al., 2011), to keep only variants called by the 3 callers (genotypes kept were the GATK ones). SNP and INDEL were separated using GATK SelectVariants, and SNP were filtered with GATK VariantFiltration by following the GATK recommendations ($QD < 2.0$ || $MQ < 40.0$ || $FS > 60.0$ || $SOR > 3.0$ || $MQRankSum < -12.5$ || $ReadPosRankSum < -8.0$). This constitutes the INRAE1 variants dataset which includes 14,439,713 SNPs.



Using a `vcf_filter.py` homemade script to parse VCF file, INRAE1 dataset was filtered to keep only bi-allelic SNP localized on the 29 trout chromosomes or mitochondrial chromosome, fully genotyped for all 29 samples. As 10 isogenic lines were duplicated, we also checked genotype consistency and removed SNP with more than 1 isogenic line genotype discordance. Finally, we kept one sample per isogenic line (with the deepest sequencing) and filtered out SNP with more than 1 heterozygote genotype as they may represent duplicated genome regions. Among the 14,439,713 variants, we kept 10,286,009 SNPs (71.23%).

We merged them using BCFtools with a second dataset INRAE2, containing 14,478,077 SNPs called from 60 samples of a commercial line from “Les Fils de Charles Murgat” (Beaufort, France) and whose resequencing was described in Fraslin et al. (2020).

This merged dataset was filtered like the merged USDA dataset, to keep bi-allelic SNP localized on the 29 trout chromosomes of the Swanson reference genome or mitochondrial chromosome, with a $MAF > 1\%$. The final INRAE cleaned database contained 16,466,188 SNPs.

Merging the USDA and INRAE SNP Databases and SNP Preselection

A total of 32,372,492 distinct SNPs were selected for consideration for the HD chip, from a combination (BCFtools merge) of the USDA and INRAE databases (<https://doi.org/10.5281/zenodo.6657091>).

An overview of the process to detect and select SNPs for inclusion on the array is provided in **Figure 1**.

SNPs were further filtered to be at least 50 base pairs from the closest identified SNP which resulted in a subset of 3,679,547 SNPs.

During `vcf` files merging, additional alleles may be added on shared variant positions and some variants previously removed from the INRAE dataset (on replicate discordance or high isogenic heterozygosity rate) may be reincluded. Thus, for a first SNP preselection, in addition to filtering SNPs with $MAF \geq 10\%$ in both the USDA and INRAE databases, we applied filters on bi-allelism variant and on a maximum number of 4 heterozygote INRAE isogenic lines.

Assessment included a check for duplicate flanking information suggesting repetitive elements, and an assessment of the complexity of the flanking sequence:

- 1) Unicity of at least one side 35 bp-sequence for each SNP. This was done by blasting (default `blastn` parameters, Camacho et al., 2009) the 35 bp on the reference assembly genome and by checking that the best match was unique and located on the expected chromosome.
- 2) Trimming of each side 50 bp sequence if it contained more than 3 successive N. Variants were kept if at least the shortest trimmed sequence contained 20 bp and the other 50 bp (`trimN.py` homemade script, **Supplementary File S2**).

This first high-quality selection represented 633,405 SNPs. Trimmed flanking sequence each side of the SNP was extracted

for all SNPs and formatted for Affymetrix (Thermo Fisher Scientific, United States) according to their specifications.

From this first submission to Affymetrix quality control, only 457,086 SNPs were qualified as recommended to be designable for the HD array and among them, only 351,755 were not ambiguous, meaning they were not of the type [G/C] or [T/A] that would require 4 probes instead of only two to distinguish the alleles.

To get sufficient recommended variants and to avoid the selection of markers that will use twice the space used by the others on the HD array, we decided to resubmit a large second set of variants to Affymetrix quality check. The same procedure was applied to produce a second more relaxed preselected set of SNPs by keeping SNPs with a MAF $\geq 10\%$ in the INRAE dataset only. This second preselection contained 533,637 additional SNPs. Among that additional set, 134,086 SNPs were specific to the INRAE dataset while the others were also present in the USDA dataset but with MAF below 10%.

We merged the first recommended set of 457,086 SNPs with this additional set of 533,637 SNPs. Then we removed all ambiguous SNPs of type [G/C], [C/G], [T/A] or [A/T]. Finally, densities were adjusted such that in regions with more than 30 SNPs retained per 100 kb by the previous filters, we only kept SNPs with MAF $\geq 15\%$ in at least one of the two INRAE or USDA databases.

This procedure resulted in a selection of 815,525 SNPs for the final submission in October 2020 to Affymetrix for assessment of the suitability of the SNPs for inclusion on a custom AXIOM 96HT SNP chip. Of the submitted SNPs, a total of 623,544 SNPs were deemed to be “designable” (recommended or neutral) in either the forward or reverse flanking sequence based on the Affymetrix pconvert score.

Keeping Informative Variants From the Medium-Density Axiom® Trout Genotyping Array

The INRAE and USDA research teams were willing to keep in the HD chip design the informative markers from the 57 K chip. Therefore 41,999 SNPs out of its 57,501 SNPs were designable in either forward and reverse directions and were kept for the HD chip design.

At the only exception of 8 specific SNPs, all the markers had a unique position on the Swanson reference genome and MAF $> 5\%$ in at least one French or North American population. Among them, 38,826 SNPs were also put on a 200 K chip that was built on 120 resequenced mostly “wild” genomes from over 40 locations from Russia, Alaska Canada down through Washington, Oregon and California (Ben Koop’s personal communication).

Selection of SNPs for the HD-Trout SNP Chip

In total 664,531 SNPs corresponding to 701,602 probesets (some SNPs were tiled in both directions as both their forward and reverse flanking sequence was assessed to be neutral) passed the Affymetrix final quality control to be designed on the custom HD

Axiom array. Only 40,987 of the 41,999 SNPs from the 57 K chip remained on the HD final design.

Among the selected SNPs, 664,503 were mapped on the 29 chromosomes of the Swanson reference genome (**Figure 2**), while 28 were positioned on the mitochondrial genome.

Based on the Swanson reference genome mapping, the average (median) SNP density on the chromosomes was 293 (324) SNPs per Mb (**Figure 2**), with SNP density varying from 2 to 774 SNPs per Mb. The average (median) inter-marker distance was 2.9 kb (1.3 kb). Maximum inter-marker distance was 243 kb and only 5.4% of inter-marker distance was over 10 kb (0.02% over 100 kb).

Single Nucleotide Polymorphism Mapping on the Arlee Reference Genome

Recently USDA/ARS (Gao et al., 2021) released a second reference genome assembly (USDA_OmykA_1.1; GenBank assembly accession GCA_013265735.3) for *Oncorhynchus mykiss* as long reads-based *de-novo* assembly for a second WSU DH line, named Arlee line, had been performed. Because Arlee lineage was closer from the INRAE isogenic lines (Palti et al., 2014) than Swanson lineage, it was decided to keep for further analysis only the SNPs that were mapped uniquely on one of the 32 chromosomes of this new reference genome.

Therefore, the 701,602 SNP probe flanking sequences of the HD chip were realigned to the new Arlee reference genome using BLAST (blastn with default task, i.e., megablast, and filtering options -evalue 1e-5 -perc_identity 95 -best_hit_overhang 0.1 -best_hit_score_edge 0.1).

SNP Validation

Given that the objective of our work was to get a powerful genomic tool with a large number of SNPs useful in diverse rainbow trout populations, we included fish from two French commercial lines that were genetically distinct from the lines used from the SNP discovery step.

We used fin samples collected from “Bretagne Truite” (Plouigneau, France) and “Viviers de Sarrance” (Sarrance, France) commercial lines, hereafter named LB and LC lines respectively, that were sampled for the FEAMP project Hypotemp (n° P FEA470019FA1000016). Pieces of caudal fin sampled from 463 fish (288 from LB line and 175 from LC line) were sent to Gentyane genotyping platform (INRAE, Clermont-Ferrand, France) for DNA extraction using the DNAdvance kit from Beckman Coulter following manufacturer instructions and genotyping using the newly constructed HD SNP array.

The first round of quality control was done by ThermoFisher software AxiomAnalysisSuite™ with threshold values of 97% for SNP call rate and 90% for sample call rate. All the 288 individuals of the LB line passed the preliminary control, while 174 out of the 175 individuals from LC line passed the control quality. Following array hybridization and imaging, genotypes were called using default settings in the Axiom Analysis Suite software and exported from the software in PLINK (Purcell et al., 2007) format.

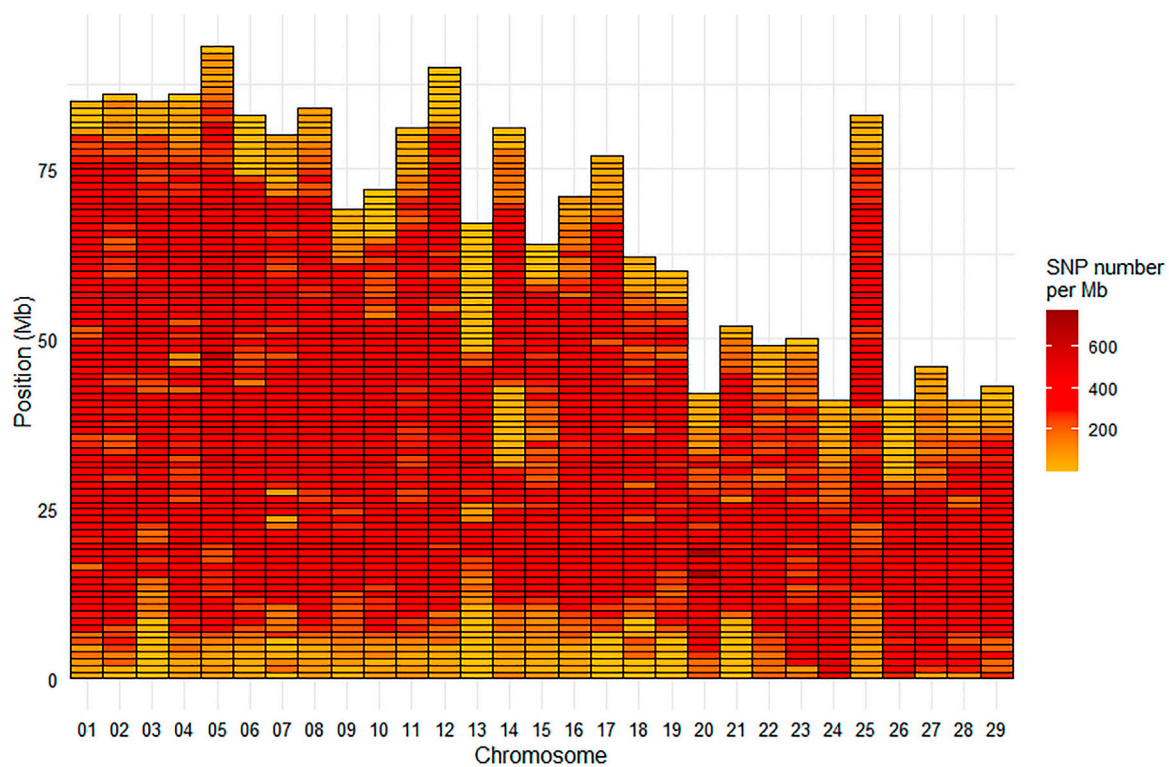


FIGURE 2 | Marker density per Mb for the HD Trout Affymetrix array with 664,503 SNPs positioned on the 29 chromosomes of the Swanson genome reference.

In addition, we used the WGS information of 20 samples sequenced in Gao et al. (2018)' study (with average genome coverage above $\times 20$) to extract their genotypes for SNPs included in the HD chip and positioned on the Arlee reference genome (Gao et al., 2021; GCA_013265735.3). Those samples came from hatchery (Dworhak, L. Quinault, Quinault, Shamania) and wild (Elwha) populations from the North-West of United States (Four samples for each of the five populations) and were proved to be genetically close to each other and very distant from the Norwegian Aquagen aquaculture population (Gao et al., 2018). The idea was to infer and compare the level of linkage disequilibrium across the HD markers from wild/hatchery American populations and farmed French selected lines.

Allele Frequencies and Linkage Disequilibrium Across Populations

We then used PLINK v1.9 (www.cog-genomics.org/plink2) to calculate allele frequencies, filter SNPs at low MAF or individuals with high identity by descent (IBD) values and derive linkage disequilibrium (LD) measured as the correlation coefficient r^2 , using the mapping of the SNP probe flanking sequences to the Arlee genome.

Allele frequencies were calculated per population for each SNP. SNPs were then filtered to only those with a MAF $\geq 5\%$, leaving 249,055 variants for American populations, 420,778 SNPs for the LB line, and 423,061 SNPs for the LC line. The set of individuals was also filtered using “*rel-cutoff*

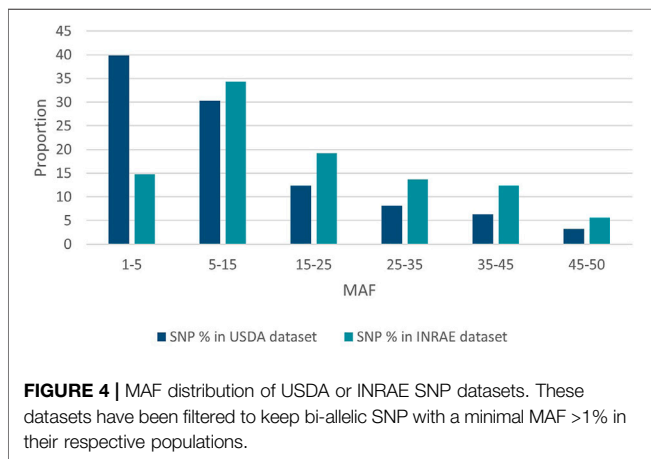
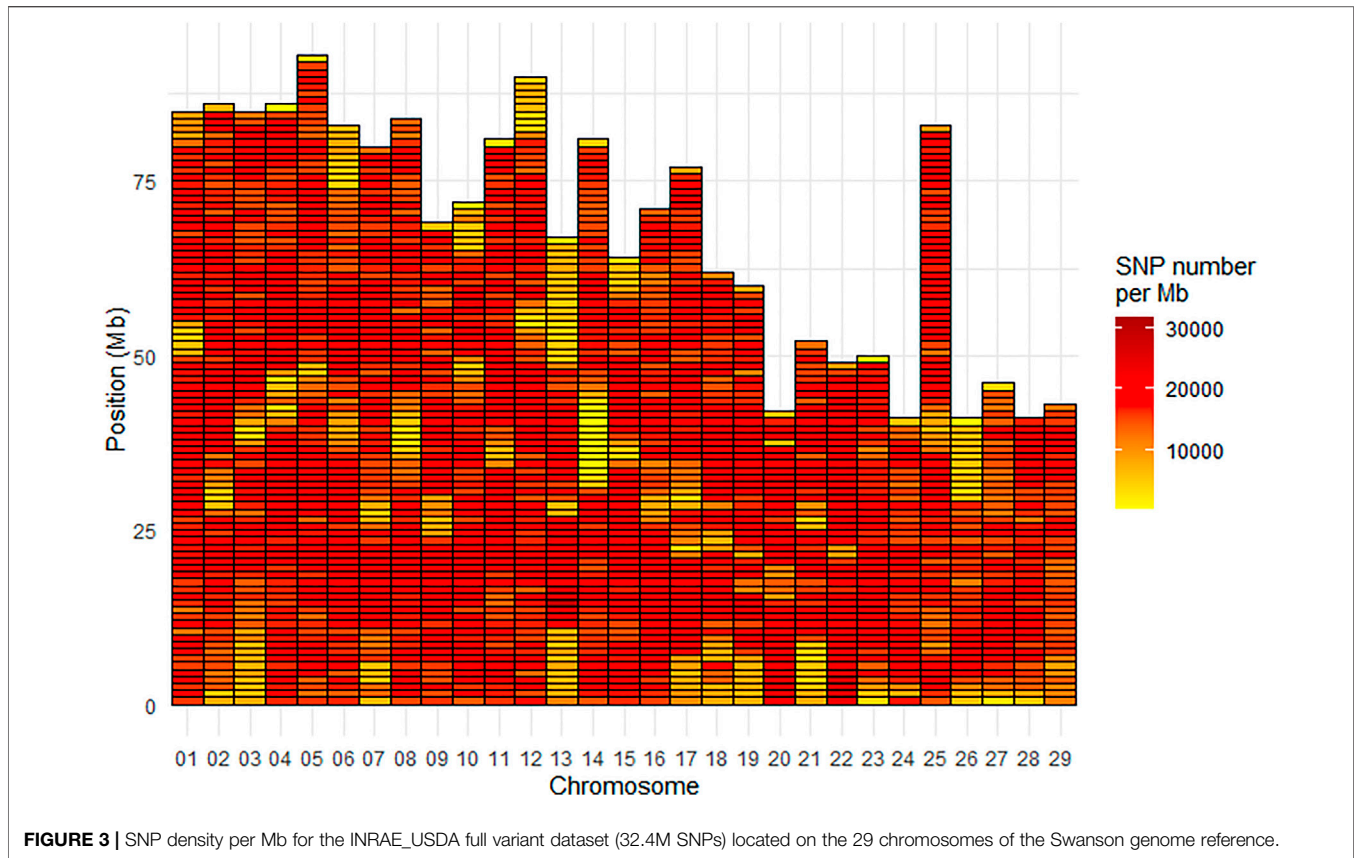
0.12” to exclude one member of each pair of samples with observed genomic relatedness above 0.12, keeping 120 samples across populations, corresponding to 20, 45, and 55 individuals for American populations, LB and LC lines, respectively. Linkage disequilibrium (r^2) between all pairs of SNPs on the same chromosome and at physical distances up to 1 Mb was then calculated using the PLINK options “*--r2 --ld-window 50000 --ld-window-kb 1001 --ld-window-r2 0.0*”. The r^2 values were binned into 2 kb units and per-bin averages calculated using R (R Core Team, 2019) for all chromosomes. The LD decay over physical distance up to 100 kb was then plotted in R.

RESULTS

SNP Identification and Characterization Based on the Swanson Reference Genome

Density of SNPs varied strongly from one chromosome to another with average SNP density per Mb ranging from 13,200 for Omy26 to 20,132 for Omy22. Across all chromosomes, the average SNP density per Mb was 16,483 SNPs (Figure 3). The Mb with the minimum density contained 451 SNPs while the Mb with the highest density contained 31,819 SNPs.

SNP identified in USDA or INRAE databases differed in terms of MAF distribution (Figure 4): 70% and 49% of SNPs had a MAF below 15% (40% and 15% had a MAF below 5%, respectively)



while only 9.5% and 18% of SNPs had a MAF above 35% in the USDA and INRAE datasets respectively.

HD Chip

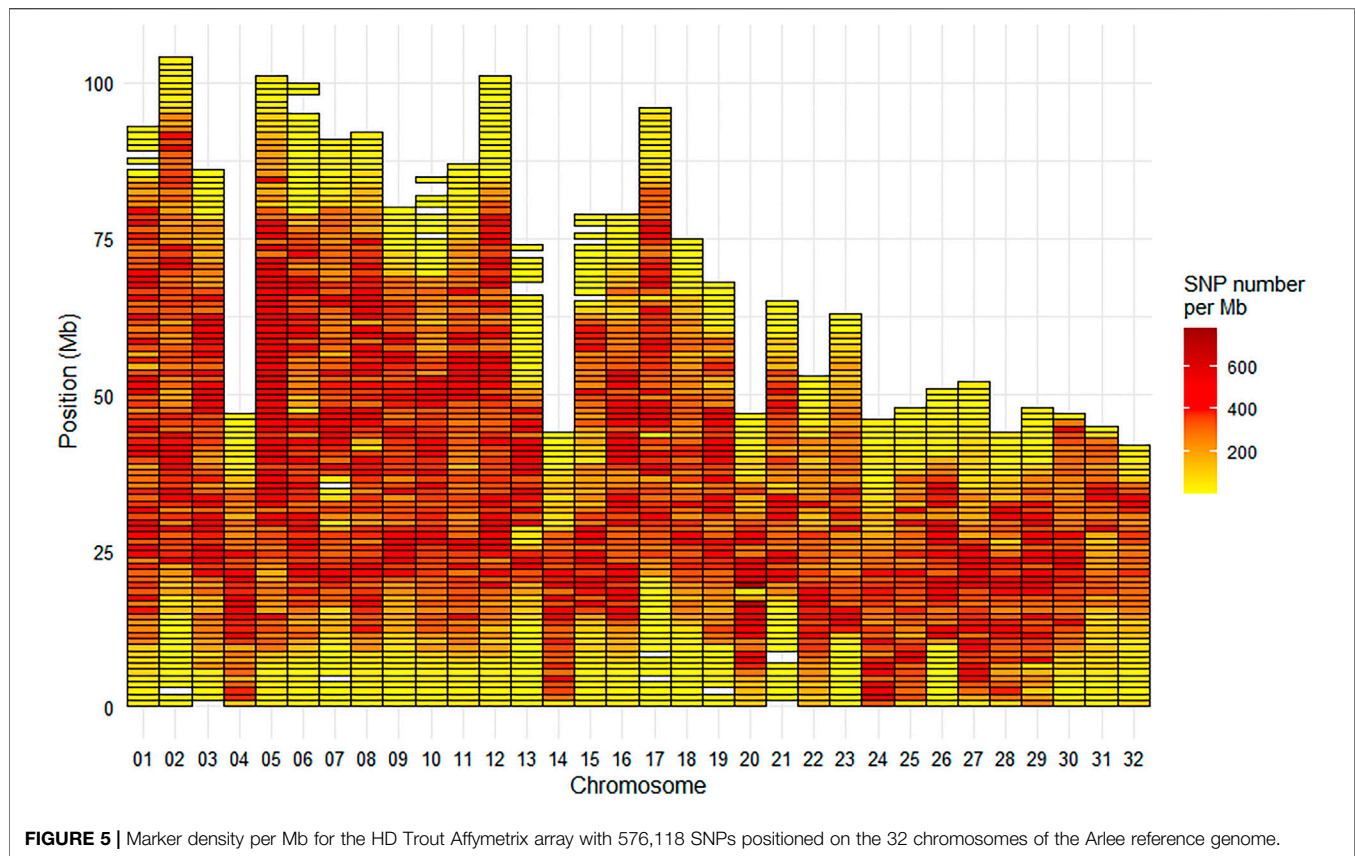
Based on genotyping the 288 LB samples, 65.34% of markers were polymorphic, had individuals with all three genotypes, and passed Affymetrix filtering metrics in the Axiom Analysis Suite software to be categorized as “PolyHigh Resolution” variants. Of those that “failed” to be in that category, 15.35% passed filtering metrics but were monomorphic, 10.71% passed filtering metrics but the

minor allele homozygote was missing, and the remainder 8.60% failed due to low call rates or other quality filters. The total number of best recommended markers was 91.81% corresponding to 610,115 SNPs out of the 664,531 genotyped variants.

Based on genotyping the 175 LC samples, 69.91% of markers were polymorphic, had individuals with all three genotypes, and passed Affymetrix filtering metrics in the Axiom Analysis Suite software to be categorized as “PolyHigh Resolution” variants. Of those that “failed” to be in that category, 5.63% passed filtering metrics but were monomorphic, 14.84% passed filtering metrics but the minor allele homozygote was missing, and the remainder 9.62% failed due to low call rates or other quality filters. The total number of best recommended markers was 90.86% corresponding to 603,768 SNPs out of the 664,531 genotyped variants.

Of the 664,531 SNPs which passed the Affymetrix quality filters and were included on the HD chip, 576,118 SNPs mapped uniquely on both reference genomes, and 12,071 SNPs did not map at all on the Arlee reference genome. **Supplementary Data S1** indicates both positions on the Swanson and Arlee reference genomes. Among those 576,118 SNPs, 38,948 SNPs were kept from the initial 57 K chip.

On the Arlee mapping (GCA_013265735.3), the average SNP density on the chromosomes was one SNP per 3.8 kb, or 266 SNPs per Mb. The median inter-marker distance was 1.5 kb with only 7% of the distances between successive



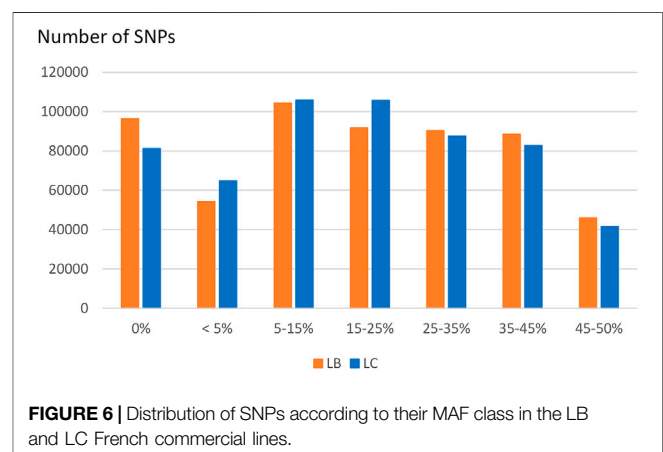
markers being above 10 kb. The largest gap was 4.16 Mb at the end of chromosome Omy6, the second largest gap was 2.94 Mb at the end of chromosome Omy10 and the third largest gap was 2.75 Mb at the end of chromosome Omy13 (Figure 5). Only five other gaps were above 2 Mb with values ranging from 2.3 to 2.5 Mb on chromosomes Omy7, Omy10, Omy15, and Omy21.

Finally, PLINK v1.9 software (www.cog-genomics.org/plink2) was used for a final SNP filtering based on keeping for further analysis SNPs with call rate above 95% and a deviation test from Hardy-Weinberg equilibrium (HWE) with a p -value $< 10e-7$ within each population. For LB line, 571,319 markers (474,937 being polymorphic) were kept after removing 1,136 miss genotyped SNPs and 3,663 ones with severe deviation from HWE. For LC line, 569,030 markers (487,940 being polymorphic) remained after removing 2,574 miss genotyped SNPs and 4,592 ones with severe deviation from HWE.

Regarding the American sequenced population, we extracted from the vcf files the genotypes for the 576,118 SNPs that were retained on the HD chip. Only 338,660 of those markers were polymorphic in the American population.

MAF Distribution in the Two French HD Genotyped Populations

Compared to variants called from sequence data, the MAF distribution of the HD selected SNPs was skewed to common alleles (Figure 6) with over 70% of SNPs with MAF above 5% in



each of the two populations, and over 20% of SNPs with MAF over 35% in both populations. Among polymorphic SNPs (MAF > 0.001), the average (median) MAF was 24.1% (23.6%) in the LB line and 23.0% (21.5%) in the LC line.

Linkage Disequilibrium Analysis

The median intermarker distance was 2 kb and the corresponding average r^2 between neighbouring markers was 0.47, 0.44, and 0.36 in LB, LC, and American population, respectively. As expected, average r^2 tended to decrease with increasing

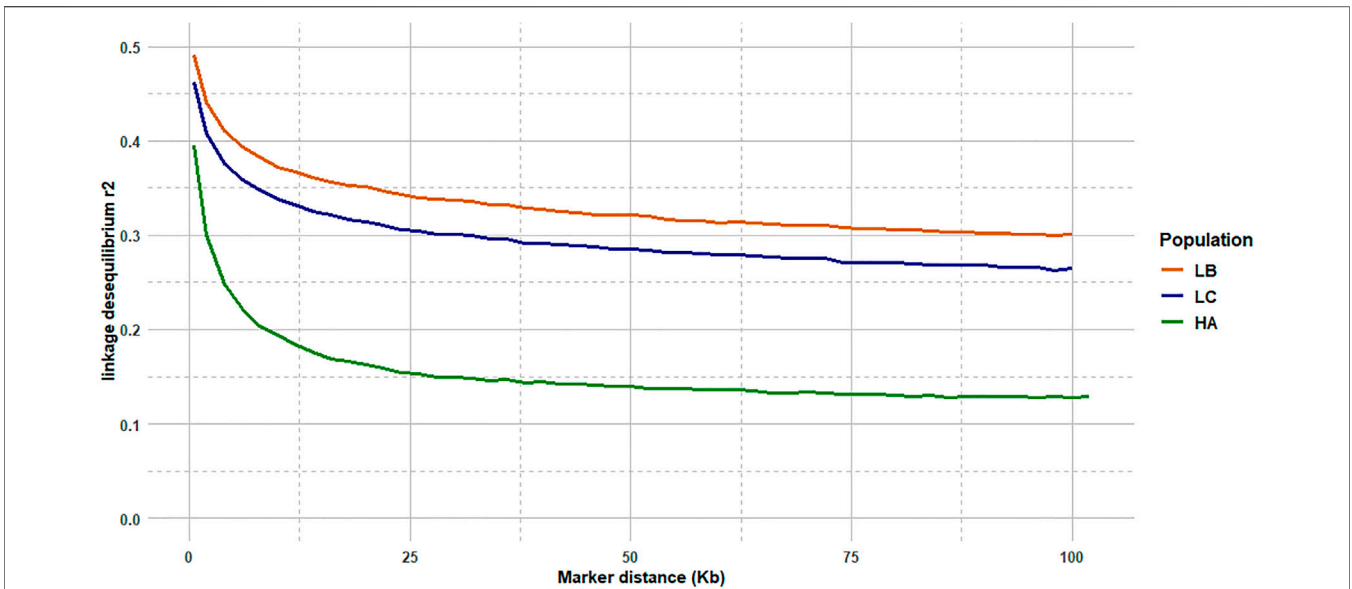


FIGURE 7 | LD decay from 2 to 100 kb intermarker distances (average over the 32 chromosomes) for the LB and LC French commercial lines and the HA American population.

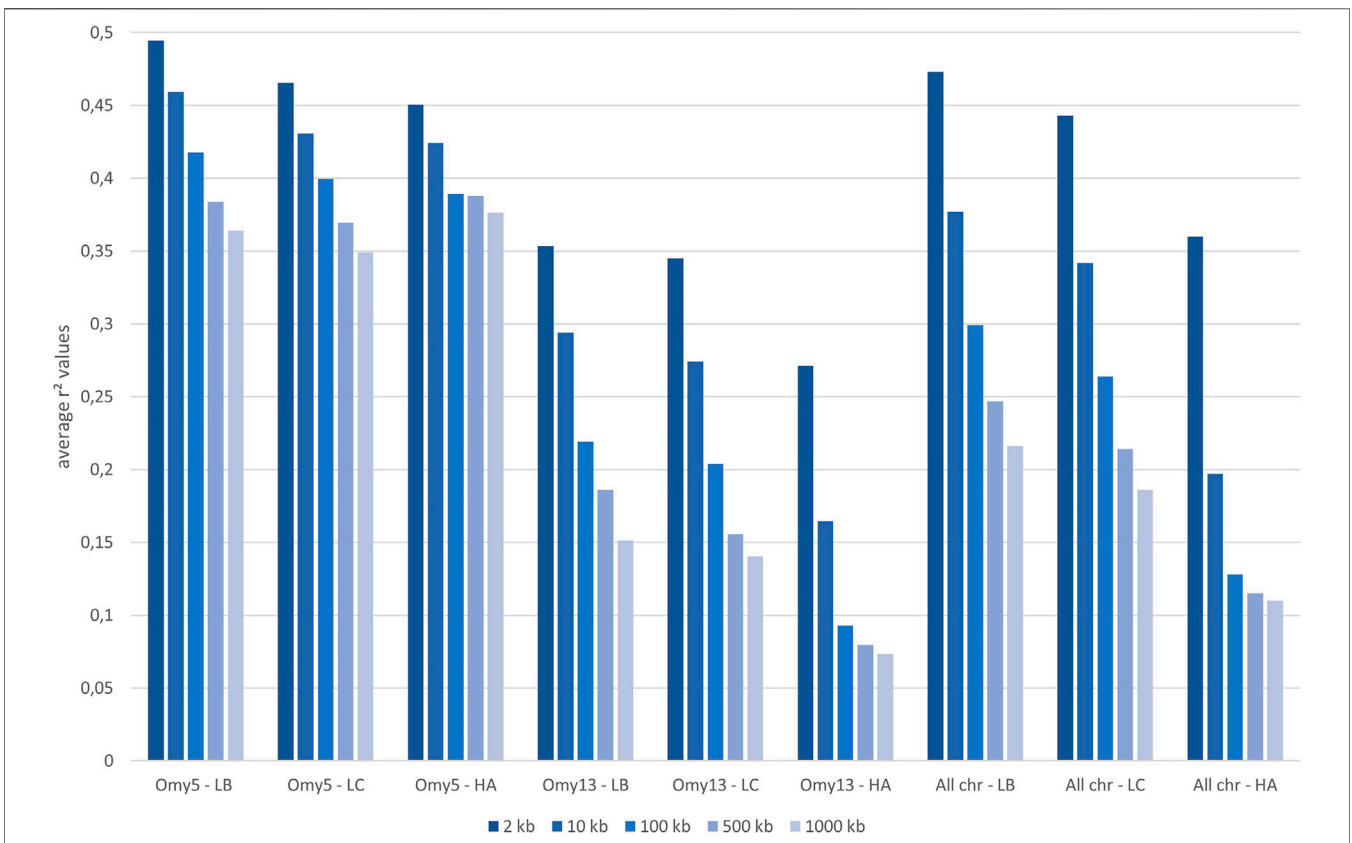


FIGURE 8 | Average linkage disequilibrium (r^2 values) from 2 to 1,000 kb derived for all chromosomes and only for Omy5 or Omy13 in populations LB, LC, and HA, respectively.

distance between pairs of markers in all populations studied, the most rapid decline being over the first 10 kb (Figure 7). Linkage disequilibrium was very high, with r^2 reaching 0.42, 0.39, and 0.27 at the average intermarker distance (4 kb) for LB, LC, and American population respectively; at 50 kb distance, r^2 average values were 0.32, 0.29, and 0.14 (Figure 7). At 500 kb, values were 0.25, 0.21, and 0.18 and values were still 0.22, 0.19, and 0.11 at 1 Mb, respectively for LB, LC, and the American population (Figure 8). However, those r^2 values may vary strongly from one chromosome to another as shown on Figure 8 for chromosomes Omy5 and Omy13 with respectively higher and lower linkage disequilibrium observed in comparison to the average values derived for all chromosomes.

DISCUSSION

In this study, based on the resequencing of tens of individuals from a diverse range of populations, we developed a high density (665K) SNP array that will be used for numerous applications, including genomic populations studies, GWAS or genomic selection. In fish, the first very high-density chip, named 930K XHD Ssal array, was developed for Atlantic Salmon using 29 fish from Aquagen lines and was a powerful tool to identify the key role of *VGLL3* gene on age at maturity (Barson et al., 2015) or the epithelial cadherin gene as the major determinant of the resistance of Atlantic salmon to IPNV (Moen et al., 2015). A similar approach was used in Atlantic salmon with whole genome re-sequencing of 20 fish from three diverse origins to generate a catalogue of 9.7M SNPs that were then filtered to design a 200K SNP chip (Yáñez et al., 2016). A similar number of 9.6M SNPs were identified for the development of a 700 K SNP chip in catfish (Zeng et al., 2017). Recently, a set of 82 fish were collected from six different locations of China and re-sequenced to identify 9.3M SNPs to design a 600K SNP chip for large yellow croaker (Zhou et al., 2020).

Based on the resequencing of 85 samples by USDA and 79 samples by INRAE, we identified 32,372,492 SNPs that were variants ($MAF \geq 1\%$) in either the USDA or the INRAE sets. More precisely, 29.0 and 16.4 million SNPs were identified in the USDA and INRAE datasets respectively for equivalent number of sequenced individuals. The higher number of SNPs detected in the USDA dataset probably resulted from the larger number of diverse populations included in the USDA dataset. The USDA database included 11 doubled haploid individuals and 50 individuals from 7 commercial, hatchery or wild populations, compared to the INRAE database that included 19 doubled haploid individuals derived from one experimental line and 60 individuals sampled from a single French commercial line. For comparison purposes, the influence of the numbers of sequenced individuals and populations or breeds on the number of identified SNPs can be exemplified in two large-scale projects, the 1000 human genomes project and the 1000 bull genomes project. In the human genome, a pilot phase identified ~15 million SNP based on the WGS of 179 individuals from four populations (Altshuler et al., 2010); increasing the number of sequences to 2,504 coming from

26 populations across the world increased considerably the number of identified SNPs to 84.7 million (Auton et al., 2015; Fairley et al., 2020). Similarly, the first phase of the 1000 bull genomes project identified 26.7 million SNPs based on the resequencing of 234 bulls from 3 breeds (Daetwyler et al., 2014); again, the number of SNPs increased to 84 million by sequencing 2,703 individuals from 121 breeds (Hayes and Daetwyler, 2019). Another study in chicken highlights the importance of sequencing a diverse set of individuals to identify a large catalogue of SNPs: WGS of 243 chickens from 24 chicken lines derived from diverse sources lead to the detection of about 139 million putative SNPs (Kranis et al., 2013).

In this study, the average distance between two successive variants was 60 bp, indicating important polymorphism level in the rainbow trout genome. This is consistent with the average SNP rate over all chromosomes of one SNP every 64 bp previously reported by Gao et al. (2018) in the Swanson rainbow trout reference genome. Such short average distance between successive variants was a strong limiting factor to preselect SNPs to design the HD chip. Indeed, an important technical issue in SNP array design is that very high SNP densities can potentially cause allele dropout when genotyping due to interferences between polymorphism at the marker position and at the probe designs that have to be monomorphic sequences flanking the marker candidates. When searching for markers with inter-marker distance over 50 bp that could be considered in the HD array design, we could only retain 3.68M SNPs.

Across all chromosomes, the average SNP density per Mb was 16,483 SNPs, i.e., slightly higher than the ~15,600 SNPs per Mb reported by Gao et al. (2018), although density of SNPs varied strongly from one chromosome to another (from 13,200 for Omy26 to 20,132 for Omy22). Interestingly, the lower SNP densities on Omy26 was also described in Gao et al. (2018) and associated with a higher proportion of SNPs being filtered out as putative paralogous sequence variants (PSV), as this chromosome shares high sequence homology with other chromosome arms in the genome as a result of delayed rediploidization. Stronger variation in average SNPs density among chromosomes has been reported previously in chickens (Kranis et al., 2013) and humans (Zhao et al., 2003), with average value of 78 and 83.3 SNPs per kb across the genome but with some chromosomes having only 3 (on chromosome Z) and 2 (on chromosome Y) SNPs reported per kb respectively.

There was also a heterogeneous distribution of SNPs along the chromosomes, with a minimum density per Mb as low as 451 SNPs, and a maximum of 31,819 SNPs. Areas with less SNP density generally located at the telomeric parts or the centromeric parts (for metacentric chromosomes) of some chromosomes (e.g., Omy13 and Omy14) (Figure 3). Such heterogeneous distribution of SNPs has been previously reported in Eukaryotes, with potential explanations including heterogeneous recombination across the genome. It has been reported in a meta-analysis in eukaryotes that “heterogeneity in the distribution of crossover across the genome is a key determinant of heterogeneity in the distribution of genetic variation within and between populations” (Haenel et al., 2018). One broad-scale and general pattern observed within chromosomes is a lower recombination rate around centromeres

(Stapley et al., 2017) and higher rates at the telomeric parts (Sakamoto et al., 2000; Anderson et al., 2012). Because higher recombination rates are observed in telomeric than centromeric regions of chromosomes, a higher number of variants may be expected in the telomeres. However, in general, the telomeres have very long patterns of repeats which generate problems in reads mapping. In the centromeric regions, it is unclear whether or not suppressed recombination is linked to highly repetitive regions (Talbert and Henikoff, 2010). Last but not least, the complexity of the rainbow trout genome with its recent whole genome duplication and partial rediploidization, and patterns of tetrasomic inheritance (Pearse et al., 2019), can potentially explain the difficulties to sequence and assemble some parts of its genome and hence detect SNPs. In a recent paper, Gui et al. (2022) have reported several phenomena (such as massive sequence divergences, extensive chromosome rearrangements, large-scale transposon bursts) occurring during the polyploidization and rediploidization that could explain the difficulties in assembling the complex genomes of Salmonids and other tetraploid fish species. Indeed, rainbow trout has a high content (57.1%) of repetitive sequences (Pearse et al., 2019), similar to the 59.9% reported for Atlantic salmon (Lien et al., 2016).

Taking advantage of the biological characteristics of fish (external fertilization and embryonic development, viability of uniparental progeny), isogenic lines have been generated in some fish species (reviewed in Franěk et al., 2020), by either gynogenesis (Quillet et al., 2007) or androgenesis (Young et al., 1996) in rainbow trout. Both USDA and INRAE datasets included the sequencing of 11 and 19 doubled-haploid individuals respectively from 30 different isogenic lines. This number, quite large and unique in fish, makes it possible to take advantage of both the within-line characteristics (homozygosity, isogenicity) and between-line variability. In particular, rainbow trout isogenic lines are being used for the development of genomic tools: the trout genome is the result of a whole genome duplication event that occurred about 96 Mya ago (Berthelot et al., 2014). Therefore, many genomic regions remain in a pseudo-tetraploid status, which complicates sequence assembly and development of genetic markers because of the difficulty to distinguish true allelic variants from PSVs. Therefore, homozygous individuals were used to produce the first genome sequence and reference transcriptome (Berthelot et al., 2014), subsequent improved genome assemblies (Pearse et al., 2019; Gao et al., 2021), and also to validate the large set of SNPs used in the first 57K SNP chip (Palti et al., 2014, 2015). In the present study, as in Gao et al. (2018), putative PSVs were filtered out by using information from the isogenic lines' genotypes, in order to generate a comprehensive catalogue of reliable SNPs in rainbow trout and then filter out SNPs to be included onto the HD SNP chip.

The 665 K SNP chip was designed based on the Swanson reference genome (Pearse et al., 2019). Only 576K SNPs were uniquely positioned on the Arlee reference genome, which led to a few gaps over 1 Mb based on this reference genome (Figure 5) while there was no gap over 250 kb on the Swanson reference genome (Figure 2). Genetic and genomic differences between the Swanson and Arlee lines have previously been studied (Palti et al., 2014). It is also known that the two lines differ in their chromosomes' numbers, the Swanson line having $2N = 58$ with 29 haploid chromosomes (Phillips and Ráb, 2001) and the Arlee line $2N = 64$ with 32 haploid

chromosomes (Ristow et al., 1998). This is not surprising as there are some variable chromosome numbers in rainbow trout populations, associated with Robertsonian centric fusions or fissions, as for instance fission splitting metacentric chromosome 25 observed in Swanson genome into two acrocentric chromosomes in French lines (Guyomard et al., 2012; D'Ambrosio et al., 2019). Depending on the rainbow trout populations, the number of haploid chromosomes (N) varies from 29 to 32 and evidence suggests that the redband trout with $2N = 58$ is the most ancestral type (Thorgaard et al., 1983). In the Arlee karyotype the haploid chromosome number is 32 because chromosomes Omy4, Omy14, and Omy25 are divided into six acrocentric chromosomes (Gao et al., 2021). Note that Arlee chromosomes Omy30, Omy31, and Omy32 correspond to the p-arms of, respectively, Omy4, Omy25, and Omy14 on the Swanson genome.

The 664,531 SNPs successfully genotyped on 463 individuals across two French commercial populations represent a valuable tool for ongoing genomic studies on the genomic architecture of traits, the population evolution history and genetic diversity as well as for the assessment of inbreeding and the genetic effects of management practices in farmed populations. The quality of our new HD chip compared very well with other HD chips developed for salmonid complex genomes. Indeed respectively 90.9% and 91.8% of SNPs were validated in LB and LC samples as high quality based on clustering properties, while corresponding values given for Coho salmon array (Barriá et al., 2019) and Atlantic salmon (Yáñez et al., 2016) were 82.5 % and 79.6% for 200 K Affymetrix Axiom myDesign Custom Arrays. The HD chip is a powerful genomic tool that allow not only to have on average all along the genome a very high density of markers in comparison to the 57 K chip, but also to significantly reduce the number of large gaps (>1 Mb) in the genome coverage. In particular, the extremely low coverage at the telomeric parts of most of the chromosomes or at the centromeric part of metacentric chromosomes have been drastically reduced and the 2 regions spanning over 10 Mb each without any markers on Omy13 (see **Supplementary Figure S1**) have been drastically reduced, leaving just a large gap of 2.75 Mb at the end of Omy13 on the Arlee reference genome. This remaining gap is likely due to the fact that the entire chromosome Omy13 shares high sequence homology with other chromosome arms due to delay in re-diploidization (Gao et al., 2018). The next step will be to develop a new medium-density SNP array for rainbow trout keeping the 39K SNPs present on both the HD chip and the initial 57 K chip, but adding about 25 K SNPs of the HD chip to fill the large gaps without any SNP of the 57 K chip. This second version of the medium-density chip will be a very useful tool both for genomic selection and for cost-effective GWAS thanks to imputation to HD genotypes.

In our study, we illustrate the interest of the HD chip based on LD study across three different rainbow trout populations. The analysis of LD plays a central role in GWAS and fine mapping of QTLs as well as in population genetics to build genetic maps, to estimate recombination rates or effective population sizes as the expected value of r^2 is a function of the parameter $4Nc$, where c is the recombination rate in Morgan between the markers and N_e is the effective population size (Sved, 1971). The decay and extent of LD at a pairwise distance can be used to determine the evolutionary history of populations (Hayes et al., 2003; Santiago et al., 2020). Lines LB and

LC had the highest LD values in comparison to the American hatchery population (HA), potentially indicating lower effective population sizes in the French selected lines. The lower LD values in the American population may be partly linked to stratification in the sampled population gathered from diverse rivers, but however it helps to quantify the lower bound LD values at short distance that we may expect in hatchery populations. The higher than average LD observed on *Omy5* is likely caused by a large chromosomal double-inversion of 55 Mb (Pearse et al., 2019) which prevents recombination in fish.

In rainbow trout farmed populations, the level of strong LD ($r^2 > 0.20$) spans over 100 kb (D'Ambrosio et al., 2019) to 1 Mb (Vallejo et al., 2018; Vallejo et al., 2020). While a number of studies quantify in salmonids the presence of long-range LD from 50 kb to over 1 Mb either for commercial populations (Kijas et al., 2017; Vallejo et al., 2018; Barría et al., 2019; D'Ambrosio et al., 2019) or wild populations (Kijas et al., 2017), little is known on the LD at very short distances. Barría et al. (2019) indicated a maximum value of 0.21 in a Chilean Coho selected line for marker distance lower than 1 kb and a threshold value of $r^2 = 0.2$ reached at approximately 40 kb. In Atlantic salmon, $r^2 = 0.2$ was reached at approximately 200 kb in a Tasmanian farmed population coming from a single Canadian river without any further introgression (Kijas et al., 2017). In the Tasmanian salmon population, the average LD value for markers separated by 0–10 kb was 0.54 while the corresponding average LD value was only 0.04 in a Finish wild population (Kijas et al., 2017). Previous LD estimates at short distance in French rainbow trout lines ranged from 0.30 to 0.39 at an average distance of 10 kb based on a small set of markers distant from 0 to 20 kb (D'Ambrosio et al., 2019). No LD estimates at inter-marker distances below 10 kb could be reasonably derived in rainbow trout based on the 57 K chip.

In our study, regardless of the rainbow trout populations, the LD values at very short distances between markers (≤ 10 kb) were moderate (0.44–0.47 at 2 kb and 0.34–0.38 at 10 kb, respectively for LC and LB) compared to the ones observed at similar distances in cattle breeds (Hozé et al., 2013) where r^2 values were around 0.70 at 2 kb and in the range 0.50–0.55 at 10 kb whatever the breeds considered. This may be partly due to higher recombination rate in rainbow trout (1.67 cM/Mb; D'Ambrosio et al., 2019) than in cattle (1.25 cM/Mb; Arias et al., 2009), but it also indicates that the founder populations of rainbow trout farmed lines have presumably larger ancestral effective population sizes than cattle breeds. On the contrary, for marker distances over 100 kb, LD values decrease below 0.20 in cattle breeds, while average LD values are still 0.26–0.30 in LC and LB lines, respectively. This indicates stronger recent bottlenecks and selection rates in rainbow trout lines than in cattle breeds. Similar long-range LD was independently observed in two US commercial rainbow trout populations (Vallejo et al., 2018; Vallejo et al., 2020). The pattern of LD decay in rainbow trout commercial lines appears to be more similar to the one observed in conservation flocks of chicken from South Africa (Khanyile et al., 2015), with very similar values reported both at shorter distances than 10 kb, as well as at 500 kb distance where LD

values range from 0.15 to 0.24 depending on the conservation flocks and values of 0.21–0.25 were derived for LC and LB, respectively. A last factor that may contribute to this long-range LD in rainbow trout is the high crossing-over interference in males observed when plotting the linkage map distance between markers from the male vs. female linkage maps against the physical distance in base pairs. Sakamoto et al. (2000) have reported a 3.25:1 female to male linkage map distance ratio and Gonzalez-Pena et al. (2016) indicates that female/male recombination ratios were above 2.0 in all the 13 chromosomes known to have homologous pairing with at least one other chromosome arm, while in most of the non-duplicated chromosomes the ratio was generally lower. Because such high crossing-over interference in males were observed in families generated from sex-reversed XX males, we hypothesize that there must be a mechanism that is controlling meiosis in the sperm differently than in the eggs through a different regulation of gene expression not related to presence or absence of the *sY* gene.

We have demonstrated in this paper a substantial linkage disequilibrium between neighboring markers, suggesting the density of genotyped SNPs is well-designed to accurately tag most areas of the rainbow trout genome. We acknowledge that, by design, the minor allele frequency distribution of genotyped SNPs is skewed to common alleles, and variation has been predominantly sampled from common SNP shared by both French and North American farmed populations. While this may limit some analyses, we believe that the array will be an invaluable genomic resource for ongoing work investigating genetic diversity, genetic architecture of traits and adaptive potential in world-wide rainbow trout populations.

DATA AVAILABILITY STATEMENT

Raw sequence data that were generated for French isogenic lines are deposited in the ENA (Projects PRJEB52016 and PRJEB51847). The VCF file for the database of all the SNPs identified in this study including a file with allele frequency information for each SNP is deposited to the public repository ZENODO for downloading under the name "INRAE_USDA_MAF1.vcf.gz" (<https://doi.org/10.5281/zenodo.6657091>). The sequence and the genotypes of the three French commercial trout lines from "Les Fils de Charles Murgat" (Beaufort, France), "Bretagne Truite" (Plouigneau, France) and "Viviers de Sarrance" (Sarrance, France) breeding companies will be made available by request on the recommendation of Pierrick Haffray (SYSAAF, pierrick.haffray@inrae.fr).

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because this study used fin samples for breeding companies as part of their commercial breeding programs in compliance with Directive 98/58/CE on the protection of animals kept for farming purposes. As such, the project was not subjected to oversight by an institutional

ethic committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

MB, DL, and FP conceived and designed the study. MD-N, EQ, and PH were involved in the conceptualisation and funding acquisition for the project. YP and GG gave access to the USDA SNP databases. MB, AD, MC, and DL performed bioinformatics analyses on resequencing data. MB, AD, GG, and FP led the design of the 665K SNP array. PP coordinated the collection of samples to be genotyped with the 665K SNP array. JD, LJ, and CP performed the 665K SNP genotyping. KP, HL, MP, and FP analyzed the 665K genotyping data. MB, DL, and FP wrote the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This study was supported by INRAE, FranceAgrimer and the European Maritime and Fisheries Fund (Hypotemp project, n° P FEA470019FA1000016, and NeoBio project, n° R FEA470016FA1000008). The sequencing of the INRAE rainbow trout isogenic lines were partly funded by

REFERENCES

- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., et al. (2010). A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467, 1061–1073. doi:10.1038/nature09534
- Anderson, J. L., Rodríguez Mari, A., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., et al. (2012). Multiple Sex-Associated Regions and a Putative Sex Chromosome in Zebrafish Revealed by RAD Mapping and Population Genomics. *PLoS One* 7, e40701. doi:10.1371/journal.pone.0040701
- Arias, J. A., Keehan, M., Fisher, P., Coppiters, W., and Spelman, R. (2009). A High Density Linkage Map of the Bovine Genome. *BMC Genet.* 10, 18. doi:10.1186/1471-2156-10-18
- Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* 3, e3376. doi:10.1371/journal.pone.0003376
- Barria, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Leong, J. S., Rondeau, E. B., et al. (2019). Whole Genome Linkage Disequilibrium and Effective Population Size in a Coho Salmon (*Oncorhynchus Kisutch*) Breeding Population Using a High-Density SNP Array. *Front. Genet.* 10, 498. doi:10.3389/fgene.2019.00498
- Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., et al. (2015). Sex-dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon. *Nature* 528, 405–408. doi:10.1038/nature16062
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates. *Nat. Commun.* 5, 3657. doi:10.1038/ncomms4657
- Blay, C., Haffray, P., Bugeon, J., D'Ambrosio, J., Dechamp, N., Collewet, G., et al. (2021a). Genetic Parameters and Genome-wide Association Studies of Quality Traits Characterised Using Imaging Technologies in Rainbow Trout, *Oncorhynchus Mykiss*. *Front. Genet.* 12, 639223. doi:10.3389/fgene.2021.639223
- CRB-Anim (Biological Resource Centers for Domestic Animals).
- Blay, C., Haffray, P., D'Ambrosio, J., Prado, E., Dechamp, N., Nazabal, V., et al. (2021b). Genetic Architecture and Genomic Selection of Fatty Acid Composition Predicted by Raman Spectroscopy in Rainbow Trout. *BMC Genomics* 22, 788. doi:10.1186/s12864-021-08062-7
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current Status and Potential of Genomic Selection to Improve Selective Breeding in the Main Aquaculture Species of International Council for the Exploration of the Sea (ICES) Member Countries. *Aquac. Rep.* 20, 100700. doi:10.1016/j.aqrep.2021.100700
- Broad Institute (2019). Picard Toolkit: Broad Institute, GitHub Repository. Available at: <https://broadinstitute.github.io/picard/> (Accessed December 11, 2017).
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinform.* 10, 421. doi:10.1186/1471-2105-10-421
- D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., et al. (2019). Genome-wide Estimates of Genetic Diversity, Inbreeding and Effective Size of Experimental and Commercial Rainbow Trout Lines Undergoing Selective Breeding. *Genet. Sel. Evol.* 51, 26. doi:10.1186/s12711-019-0468-4
- D'Ambrosio, J., Morvezen, R., Brard-Fudulea, S., Bestin, A., Acin Perez, A., Guéméné, D., et al. (2020). Genetic Architecture and Genomic Selection of Female Reproduction Traits in Rainbow Trout. *BMC Genomics* 21, 558. doi:10.1186/s12864-020-06955-7
- Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., et al. (2014). Whole-genome Sequencing of 234 Bulls Facilitates Mapping of Monogenic and Complex Traits in Cattle. *Nat. Genet.* 46, 858–865. doi:10.1038/ng.3034
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve Years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi:10.1093/gigascience/giab008

ACKNOWLEDGMENTS

The SNP chip was developed in cooperation with Thermo Fisher and we particularly thank the following Thermo Fisher Scientific personnel for their direct contribution: Ruth Barral Arca, Marie-Laure Schneider, and Philippe Lavis. We are also grateful to the Genotoul bioinformatics platform (Toulouse Occitanie, doi:10.15454/1.5572369328961167E12) and the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi:10.15454/1.5572390655343293E12) for providing help, computing, and storage resources. We also thank the 3 French breeding companies “Les fils de Charles Murgat”, “Bretagne Truite,” and “Viviers de Sarrance” that provided samples for genome resequencing or genotyping on the 665K SNP array.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.941340/full#supplementary-material>

- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide Genetic Marker Discovery and Genotyping Using Next-Generation Sequencing. *Nat. Rev. Genet.* 12, 499–510. doi:10.1038/nrg3012
- de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage Disequilibrium and Persistence of Phase in Holstein-Friesian, Jersey and Angus Cattle. *Genetics* 179, 1503–1512. doi:10.1534/genetics.107.084301
- Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) Collection of Open Human Genomic Variation Resources. *Nucleic Acids Res.* 48, D941–D947. doi:10.1093/nar/gkz836
- Franěk, R., Baloch, A. R., Kašpar, V., Saito, T., Fujimoto, T., Arai, K., et al. (2020). Isogenic Lines in Fish—A Critical Review. *Rev. Aquacult* 12, 1412–1434. doi:10.1111/raq.12389
- Fraslin, C., Brard-Fudulea, S., D'Ambrosio, J., Bestin, A., Charles, M., Haffray, P., et al. (2019). Rainbow Trout Resistance to Bacterial Cold Water Disease: Two New Quantitative Trait Loci Identified after a Natural Disease Outbreak on a French Farm. *Anim. Genet.* 50, 293–297. doi:10.1111/age.12777
- Fraslin, C., Phocas, F., Bestin, A., Charles, M., Bernard, M., Krieg, F., et al. (2020). Genetic Determinism of Spontaneous Masculinisation in XX Female Rainbow Trout: New Insights Using Medium Throughput Genotyping and Whole-Genome Sequencing. *Sci. Rep.* 10, 17693. doi:10.1038/s41598-020-74757-8
- Gao, G., Nome, T., Pearse, D. E., Moen, T., Naish, K. A., Thorgaard, G. H., et al. (2018). A New Single Nucleotide Polymorphism Database for Rainbow Trout Generated through Whole Genome Resequencing. *Front. Genet.* 9, 147. doi:10.3389/fgene.2018.00147
- Gao, G., Magadan, S., Waldbierer, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., et al. (2021). A Long Reads-Based De-novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout. *G3 Genes Genomes Genetics* 11, jkab052. doi:10.1093/g3journal/jkab052
- Garrison, E., and Marth, G. (2012). Haplotype-based Variant Detection from Short-Read Sequencing. arXiv [q-bio.GN]. Available at: <https://arxiv.org/abs/1207.3907> (Accessed November 13, 2018).
- Gautier, M., Laloë, D., and Moazami-Goudarzi, K. (2010). Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds. *PLoS One* 5, e13038. doi:10.1371/journal.pone.0013038
- Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., and Hayes, B. J. (2016). Genetics of Complex Traits: Prediction of Phenotype, Identification of Causal Polymorphisms and Genetic Architecture. *Proc. R. Soc. B* 283, 20160569. doi:10.1098/rspb.2016.0569
- Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B. M., Kenney, P. B., et al. (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus Mykiss*). *Front. Genet.* 7, 203. doi:10.3389/fgene.2016.00203
- Gorjanc, G., Dumasy, J.-F., Gonen, S., Gaynor, R. C., Antolin, R., and Hickey, J. M. (2017). Potential of Low-Coverage Genotyping-By-Sequencing and Imputation for Cost-Effective Genomic Selection in Biparental Segregating Populations. *Crop Sci.* 57, 1404–1420. doi:10.2135/cropsci2016.08.0675
- Gui, J.-F., Zhou, L., and Li, X.-Y. (2022). Rethinking Fish Biology and Biotechnologies in the Challenge Era for Burgeoning Genome Resources and Strengthening Food Security. *Water Biol. Secur.* 1, 100002. doi:10.1016/j.watbs.2021.11.001
- Guyomard, R., Boussaha, M., Krieg, F., Hervet, C., and Quillet, E. (2012). A Synthetic Rainbow Trout Linkage Map Provides New Insights into the Salmonid Whole Genome Duplication and the Conservation of Synteny Among Teleosts. *BMC Genet.* 13, 15. doi:10.1186/1471-2156-13-15
- Haenel, Q., Laurentino, T. G., Roesti, M., and Berner, D. (2018). Meta-analysis of Chromosome-Scale Crossover Rate Variation in Eukaryotes and its Significance to Evolutionary Genomics. *Mol. Ecol.* 27, 2477–2497. doi:10.1111/mec.14699
- Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annu. Rev. Anim. Biosci.* 7, 89–102. doi:10.1146/annurev-animal-020518-115024
- Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Res.* 13, 635–643. doi:10.1101/gr.387103
- Hozé, C., Fouilloux, M.-N., Venot, E., Guillaume, F., Dassonneville, R., Fritz, S., et al. (2013). High-density Marker Imputation Accuracy in Sixteen French Cattle Breeds. *Genet. Sel. Evol.* 45, 33. doi:10.1186/1297-9686-45-33
- Johnston, S. E., Béréanos, C., Slate, J., and Pemberton, J. M. (2016). Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis Aries*). *Genetics* 203, 583–598. doi:10.1534/genetics.115.185553
- Karami, A. M., Ødegård, J., Marana, M. H., Zuo, S., Jaafar, R., Mathiessen, H., et al. (2020). A Major QTL for Resistance to *Vibrio Anguillarum* in Rainbow Trout. *Front. Genet.* 11, 607558. doi:10.3389/fgene.2020.607558
- Kardos, M., Luikart, G., and Allendorf, F. W. (2015). Measuring Individual Inbreeding in the Age of Genomics: Marker-Based Measures Are Better Than Pedigrees. *Heredity* 115, 63–72. doi:10.1038/hdy.2015.17
- Khanyile, K. S., Dzomba, E. F., and Muchadeyi, F. C. (2015). Population Genetic Structure, Linkage Disequilibrium and Effective Population Size of Conserved and Extensively Raised Village Chicken Populations of Southern Africa. *Front. Genet.* 6, 13. doi:10.3389/fgene.2015.00013
- Kijas, J., Elliot, N., Kube, P., Evans, B., Botwright, N., King, H., et al. (2017). Diversity and Linkage Disequilibrium in Farmed Tasmanian Atlantic Salmon. *Anim. Genet.* 48, 237–241. doi:10.1111/age.12513
- Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a High Density 600K SNP Genotyping Array for Chicken. *BMC Genomics* 14, 59. doi:10.1186/1471-2164-14-59
- Larson, W. A., Palti, Y., Gao, G., Warheit, K. I., and Seeb, J. E. (2018). Rapid Discovery of SNPs that Differentiate Hatchery Steelhead Trout from ESA-Listed Natural-Origin Steelhead Trout Using a 57K SNP Array. *Can. J. Fish. Aquat. Sci.* 75, 1160–1168. doi:10.1139/cjfas-2017-0116
- Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv [q-bio.GN]. Available at: <https://arxiv.org/abs/1303.3997> (Accessed December 11, 2017).
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic Salmon Genome Provides Insights into Rediploidization. *Nature* 533, 200–205. doi:10.1038/nature17164
- Liu, S., Gao, G., Layer, R. M., Thorgaard, G. H., Wiens, G. D., Leeds, T. D., et al. (2021). Identification of High-Confidence Structural Variants in Domesticated Rainbow Trout Using Whole-Genome Sequencing. *Front. Genet.* 12, 639355. doi:10.3389/fgene.2021.639355
- Marchini, J., and Howie, B. (2010). Genotype Imputation for Genome-wide Association Studies. *Nat. Rev. Genet.* 11, 499–511. doi:10.1038/nrg2796
- Mastrangelo, S., Di Gerlando, R., Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. T., et al. (2014). Genome Wide Linkage Disequilibrium and Genetic Structure in Sicilian Dairy Sheep Breeds. *BMC Genet.* 15, 108. doi:10.1186/s12863-014-0108-5
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- Meuwissen, T., and Goddard, M. (2010). Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics* 185, 623–631. doi:10.1534/genetics.110.116590
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819
- Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., et al. (2015). Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics* 200, 1313–1326. doi:10.1534/genetics.115.175406
- Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., et al. (2014). A Resource of Single-Nucleotide Polymorphisms for Rainbow Trout Generated by Restriction-Site Associated DNA Sequencing of Doubled Haploids. *Mol. Ecol. Resour.* 14, 588–596. doi:10.1111/1755-0998.12204
- Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2015). The Development and Characterization of a 57K Single Nucleotide Polymorphism Array for Rainbow Trout. *Mol. Ecol. Resour.* 15, 662–672. doi:10.1111/1755-0998.12337
- Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., et al. (2012). Extremely Low-Coverage Sequencing and Imputation Increases Power for Genome-wide Association Studies. *Nat. Genet.* 44, 631–635. doi:10.1038/ng.2283
- Paul, K., D'Ambrosio, J., and Phocas, F. (2021). Temporal and Region-specific Variations in Genome wide Inbreeding Effects on Female Size and

- Reproduction Traits of Rainbow Trout. *Evol. Appl.* 15, 645–662. doi:10.1111/eva.13308
- Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., et al. (2019). Sex-dependent Dominance Maintains Migration Supergene in Rainbow Trout. *Nat. Ecol. Evol.* 3, 1731–1742. doi:10.1038/s41559-019-1044-6
- Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., Moreno, C. R., et al. (2017). Variation in Recombination Rate and its Genetic Determinism in Sheep Populations. *Genetics* 207, 767–784. doi:10.1534/genetics.117.300123
- Phillips, R., and Ráb, P. (2001). Chromosome Evolution in the Salmonidae (Pisces): An Update. *Biol. Rev.* 76, 1–25. doi:10.1017/S1464793100005613
- Phocas, F. (2022). “Genotyping, the Usefulness of Imputation to Increase SNP Density, and Imputation Methods and Tools,” in *Chapter 4 in: Complex Trait Prediction*. Editors N. Ahmadi and J. Bartholomé (New York, NY: Springer Protocols), 113–138. doi:10.1007/978-1-0716-2205-6_4
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Quillet, E., Dorson, M., Leguillou, S., Benmansour, A., and Boudinot, P. (2007). Wide Range of Susceptibility to Rhabdoviruses in Homozygous Clones of Rainbow Trout. *Fish Shellfish Immunol.* 22, 510–519. doi:10.1016/j.fsi.2006.07.002
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available at: <https://www.R-project.org/> (Accessed January 27, 2022).
- Reis Neto, R. V., Yoshida, G. M., Lhorente, J. P., and Yáñez, J. M. (2019). Genome-wide Association Analysis for Body Weight Identifies Candidate Genes Related to Development and Metabolism in Rainbow Trout (*Oncorhynchus Mykiss*). *Mol. Genet. Genomics* 294, 563–571. doi:10.1007/s00438-018-1518-2
- Ristow, S. S., Grabowski, L. D., Ostberg, C., Robison, B., and Thorgaard, G. H. (1998). Development of Long-Term Cell Lines from Homozygous Clones of Rainbow Trout. *J. Aquatic Animal Health* 10, 75–82. doi:10.1577/1548-8667(1998)010<0075:dotcl>2.0.co;2
- Rodríguez, F. H., Flores-Mara, R., Yoshida, G. M., Barría, A., Jedlicki, A. M., Lhorente, J. P., et al. (2019). Genome-wide Association Analysis for Resistance to Infectious Pancreatic Necrosis Virus Identifies Candidate Genes Involved in Viral Replication and Immune Response in Rainbow Trout (*Oncorhynchus Mykiss*). *G3 Genes Genomes Genet.* 9, 2897–2904. doi:10.1534/g3.119.400463
- Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Hernández, D., and Villanueva, B. (2015). Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. *PLoS One* 10, e0124157. doi:10.1371/journal.pone.0124157
- Sánchez, C. C., Smith, T. P., Wiedmann, R. T., Vallejo, R. L., Salem, M., Yao, J., et al. (2009). Single Nucleotide Polymorphism Discovery in Rainbow Trout by Deep Sequencing of a Reduced Representation Library. *BMC Genomics* 10, 559. doi:10.1186/1471-2164-10-559
- Sánchez, C. C., Weber, G. M., Gao, G., Cleveland, B. M., Yao, J., and Rexroad, C. E. (2011). Generation of a Reference Transcriptome for Evaluating Rainbow Trout Responses to Various Stressors. *BMC Genomics* 12, 626. doi:10.1186/1471-2164-12-626
- Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., et al. (2000). A Microsatellite Linkage Map of Rainbow Trout (*Oncorhynchus Mykiss*) Characterized by Large Sex-specific Differences in Recombination Rates. *Genetics* 155, 1331–1345. doi:10.1093/genetics/155.3.1331
- Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle. *PLoS Genet.* 8, e1002854. doi:10.1371/journal.pgen.1002854
- Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., and Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol. Biol. Evol.* 37, 3642–3653. doi:10.1093/molbev/msaa169
- Schweizer, R. M., VonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., et al. (2016). Genetic Subdivision and Candidate Genes under Selection in North American Grey Wolves. *Mol. Ecol.* 25, 380–402. doi:10.1111/mec.13364
- Silió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., Óvilo, C., et al. (2013). Measuring Inbreeding and Inbreeding Depression on Pig Growth from Pedigree or SNP-Derived Metrics. *J. Anim. Breed. Genet.* 130, 349–360. doi:10.1111/jbg.12031
- Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., and Smadja, C. M. (2017). Variation in Recombination Frequency and Distribution across Eukaryotes: Patterns and Processes. *Phil. Trans. R. Soc. B* 372, 20160455. doi:10.1098/rstb.2016.0455
- Sved, J. A. (1971). Linkage Disequilibrium and Homozygosity of Chromosome Segments in Finite Populations. *Theor. Popul. Biol.* 2, 125–141. doi:10.1016/0040-5809(71)90011-6
- Talbert, P. B., and Henikoff, S. (2010). Centromeres Convert but Don't Cross. *PLoS Biol.* 8, e1000326. doi:10.1371/journal.pbio.1000326
- Thorgaard, G. H., Allendorf, F. W., and Knudsen, K. L. (1983). Gene-Centromere Mapping in Rainbow Trout: High Interference over Long Map Distances. *Genetics* 103, 771–783. doi:10.1093/genetics/103.4.771
- Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L., Ristow, S. S., et al. (2002). Status and Opportunities for Genomics Research with Rainbow Trout. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 133, 609–646. doi:10.1016/s1096-4959(02)00167-7
- Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic Selection Models Double the Accuracy of Predicted Breeding Values for Bacterial Cold Water Disease Resistance Compared to a Traditional Pedigree-Based Model in Rainbow Trout Aquaculture. *Genet. Sel. Evol.* 49, 17. doi:10.1186/s12711-017-0293-6
- Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., et al. (2018). Accurate Genomic Predictions for BCWD Resistance in Rainbow Trout are Achieved Using Low Density SNP Panels: Evidence that Long Range LD is a Major Contributing Factor. *J. Anim. Breed. Genet.* 135, 263–274. doi:10.1111/jbg.12335
- Vallejo, R. L., Cheng, H., Fragomeni, B. O., Shewbridge, K. L., Gao, G., MacMillan, J. R., et al. (2019). Genome-wide Association Analysis and Accuracy of Genome-Enabled Breeding Value Predictions for Resistance to Infectious Hematopoietic Necrosis Virus in a Commercial Rainbow Trout Breeding Population. *Genet. Sel. Evol.* 51, 47. doi:10.1186/s12711-019-0489-z
- Vallejo, R. L., Fragomeni, B. O., Cheng, H., Gao, G., Long, R. L., Shewbridge, K. L., et al. (2020). Assessing Accuracy of Genomic Predictions for Resistance to Infectious Hematopoietic Necrosis Virus with Progeny Testing of Selection Candidates in a Commercial Rainbow Trout Breeding Population. *Front. Vet. Sci.* 7, 590048. doi:10.3389/fvets.2020.590048
- Yáñez, J. M., Naswa, S., López, M. E., Bassini, L., Correa, K., Gilbey, J., et al. (2016). Genomewide Single Nucleotide Polymorphism Discovery in Atlantic Salmon (*Salmo Salar*): Validation in Wild and Farmed American and European Populations. *Mol. Ecol. Resour.* 16, 1002–1011. doi:10.1111/1755-0998.12503
- Yoshida, G. M., Carvalheiro, R., Rodríguez, F. H., Lhorente, J. P., and Yáñez, J. M. (2019). Single-step Genomic Evaluation Improves Accuracy of Breeding Value Predictions for Resistance to Infectious Pancreatic Necrosis Virus in Rainbow Trout. *Genomics* 111, 127–132. doi:10.1016/j.ygeno.2018.01.008
- Young, W. P., Wheeler, P. A., Fields, A. D., Thorgaard, G. H., and Vrijenhoek, R. C. (1996). DNA Fingerprinting Confirms Isogenicity of Androgenetically Derived Rainbow Trout Lines. *J. Hered.* 87, 77–81. doi:10.1093/oxfordjournals.jhered.a022960
- Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., et al. (2017). Development of a 690 K SNP Array in Catfish and its Application for Genetic Mapping and Validation of the Reference Genome Sequence. *Sci. Rep.* 7, 40347. doi:10.1038/srep40347
- Zhao, Z., Fu, Y.-X., Hewett-Emmett, D., and Boerwinkle, E. (2003). Investigating Single Nucleotide Polymorphism (SNP) Density in the Human Genome and its Implications for Molecular Evolution. *Gene* 312, 207–213. doi:10.1016/S0378-1119(03)00670-X

Zhou, T., Chen, B., Ke, Q., Zhao, J., Pu, F., Wu, Y., et al. (2020). Development and Evaluation of a High-Throughput Single-Nucleotide Polymorphism Array for Large Yellow Croaker (*Larimichthys Crocea*). *Front. Genet.* 11, 571751. doi:10.3389/fgene.2020.571751

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bernard, Dehaillon, Gao, Paul, Lagarde, Charles, Prchal, Danon, Jaffrelo, Poncet, Patrice, Haffray, Quillet, Dupont-Nivet, Palti, Lallias and Phocas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Résultats Annexe 3 : Effets de la consanguinité chromosomique sur les performances de production d'une lignée de truite arc-en-ciel française.

Dans le chapitre 3, nous avons étudié les effets de la consanguinité sur les performances de taille et de reproduction de truites arc-en-ciel françaises. Cette annexe a pour but de compléter l'étude en évaluant les caractères de production subissent également des phénomènes de dépression de consanguinité. Les caractères étudiés ont été le poids à l'abattage (BW, à ~17/18 mois), la taille (FL), le poids éviscéré (GCW), le poids étêté et éviscéré (HGCW) et la surface du tissu adipeux de la darne (ATA).

Matériel et Méthodes

a) Population

Pour cette étude nous avons utilisé une population de 2740 poissons, réparties en deux générations (1316 issus de la cohorte G9 et 1424 issus de la cohorte G10). Tous les individus ont été génotypés avec la puce moyenne densité Axiom 57k, dans le cadre du Projet SG-Truite no. RFEA47 0016 FA 1000016 et 57k-Truite (No. 2015-0638). Ici c'est l'assemblage de référence Swanson qui a été utilisé, c'est pourquoi nous étudions les effets de la consanguinité sur 30 chromosomes.

Après les tri qualités classique (snp_génotypés > 90% ; call_rate > 97%, MAF > 1% et écart non extrême aux fréquences de l'équilibre d'Hardy-Weinberg avec un seuil de p-value > 10⁻⁵), 32 013 SNPs sont conservés pour l'analyse.

b) Estimation de la consanguinité globale et locale

A l'aide de la fonction 'homozyg' du logiciel Plink v1-09 (avec les mêmes options que celles décrites au chapitre 3), le calcul des ROH a été effectué, ainsi que du coefficient de consanguinité global. Puis, le calcul des coefficients de consanguinité locaux a été effectué pour chaque chromosome de la manière suivante :

$$F_{Omyk,i} = \frac{\sum_k \text{Taille}(ROH_{k,i})}{LRegion_k}$$

où $\sum_k \text{Taille}(ROH_{k,i})$ est la somme des longueurs des ROHs du chromosome k porté par l'individu i . $LRegion_k$ est la taille totale de la region considérée.

c) Estimation des effets de la consanguinité globale et locale

Modèle globale à décrire tout d'abord

Pour estimer les effets de la consanguinité locale sur les performances y de production des individus, le modèle GBLUP suivant a été réalisé :

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \sum_{r=1}^{30} b_r \mathbf{F}_r + \mathbf{e}$$

Où $\boldsymbol{\beta}$ correspond au vecteur de l'effet de la generation; pour les caractères GCW et HGCW une covariables sur le poids (BW) a été prise en compte; pour le caractère ATA, la covariables 'Surface de la Darne' a été prise en compte.

Ensuite nous avons \mathbf{a} et \mathbf{e} les vecteurs des effets aléatoires des valeurs génétiques additives et des résidus. La matrice génomique \mathbf{G} a été construite à l'aide du programme *parallele90* (Vitezica et al., 2016).

Les coefficients de régression b_r de y sur les vecteurs \mathbf{F}_r rendent compte des effets de consanguinité le long du génome (pour cela on considère les 30 chromosomes).

NB : Dans ce modèle, lorsque nous ajoutons l'estimation des effets de dominance (avec les 30 vecteurs \mathbf{F}_r rendant compte des effets de la consanguinité le long du génome), notre modèle ne converge pas, il semble que dans ce cas, la structure des données ne permettent pas d'estimer les effets de dominance.

Résultats

Dans la population étudiée les niveaux de consanguinité globaux sont en moyenne de 19% (les valeurs minimale et maximale étant de 1% et 32%). Tout comme dans notre étude sur les effets locaux de la consanguinité chapitre 3, l'estimation des effets chromosomiques ensemble est plus précis et est meilleur modèle (selon les valeurs des AIC dans la **Table Annexe 2.1**).

Si l'on considère l'effet de la consanguinité globale, hormis pour ATA, une dépression de consanguinité entre -0,1% à 1,91% est estimée (**Table Annexe 2.2**). Les effets globaux de la consanguinité ne sont pas significatifs, tout comme la plupart des caractères de reproduction (cf Chapitre 3), c'est pourquoi il nous semble essentiel d'estimer les effets locaux de la consanguinité.

Pour chacun des caractères étudiés, au moins un chromosome parmi les 30, a un effet significatif de la consanguinité (**Figure Annexe 2.1**). Ces effets peuvent être positif ou négatif, ce résultat est très similaire à ce que nous avons pu observer dans le chapitre 3 sur les caractères de reproduction de la truite arc-en-ciel.

Table Annexe 2.1. Paramètres génétiques, variance phénotypique (V_p), et héritabilité (h^2) pour les caractères de production d'une population française de truite arc-en-ciel. Les écarts-types sont présentés entre parenthèse. Le modèle 1 est celui qui estime l'effet d'une seule valeur de consanguinité (=la consanguinité globale). Le modèle 2 estime les effets de la consanguinité présente dans les 30 chromosomes.

	Modèle1_Fglobal			Modèle2_Fomy		
	V_p	h^2	AIC	V_p	h^2	AIC
BW	298240 (9342,4)	0,26 (0,03)	42061,7	297970 (9432,4)	0,27 (0,04)	41731,0
FL	1352,5 (39,7)	0,15 (0,03)	27400,1	1359,1 (40,58)	0,16 (0,03)	27212,4
GCW	779,6 (30,9)	0,54 (0,03)	25158,0	781,2 (31,22)	0,54 (0,03)	25014,1
HGCW	1147,3 (13,5)	0,46 (0,03)	26404,5	1140,5 (43,35)	0,46 (0,03)	26222,4
ATA	36342 (13314)	0,42 (0,03)	35962,3	36481 (1348)	0,42 (0,03)	35710,4

Table Annexe 2.2. Valeurs des coefficients de régression b estimés pour les effets de la consanguinité globale sur les caractères de production (SE entre parenthèses). B^* correspond à l'effet de la consanguinité sur la performance moyenne pour une augmentation du niveau du coefficient de consanguinité globale F de 0,1 point, c'est-à-dire lorsque l'on passe d'un coefficient de consanguinité de 0,10 à 0,20.

Trait	b	B^* (%)
BW	-461,754 (382,155)	-1,91 (1,58)
FL	-45,232 (23,346)	-0,83 (0,43)
GCW	-22,247 (15,918)	-0,10 (0,07)
HGCW	-19,000 (20,004)	-0,10 (0,10)
ATA	178,417 (114,604)	1,22 (0,78)

Si l'on se place à l'échelle globale, on n'observe pas d'effet significatif de la consanguinité. Qu'en est-il localement ?

A l'échelle du chromosome, nous observons que la consanguinité présente dans au moins un chromosome a un effet significatif sur une des performances de production. Les effets de la consanguinité sur le poids à l'abattage (BW) sont plus importants (positifs et négatifs) par rapport aux autres caractères. La consanguinité de quatre chromosomes (omy7, omy12, omy13 et omy25) a un effet significatif sur le poids à l'abattage. D'autres performances comme GCW ou HGCW ont des effets de la consanguinité certes très faibles mais significatifs (omy3, omy12, omy22, omy27 et omy30).

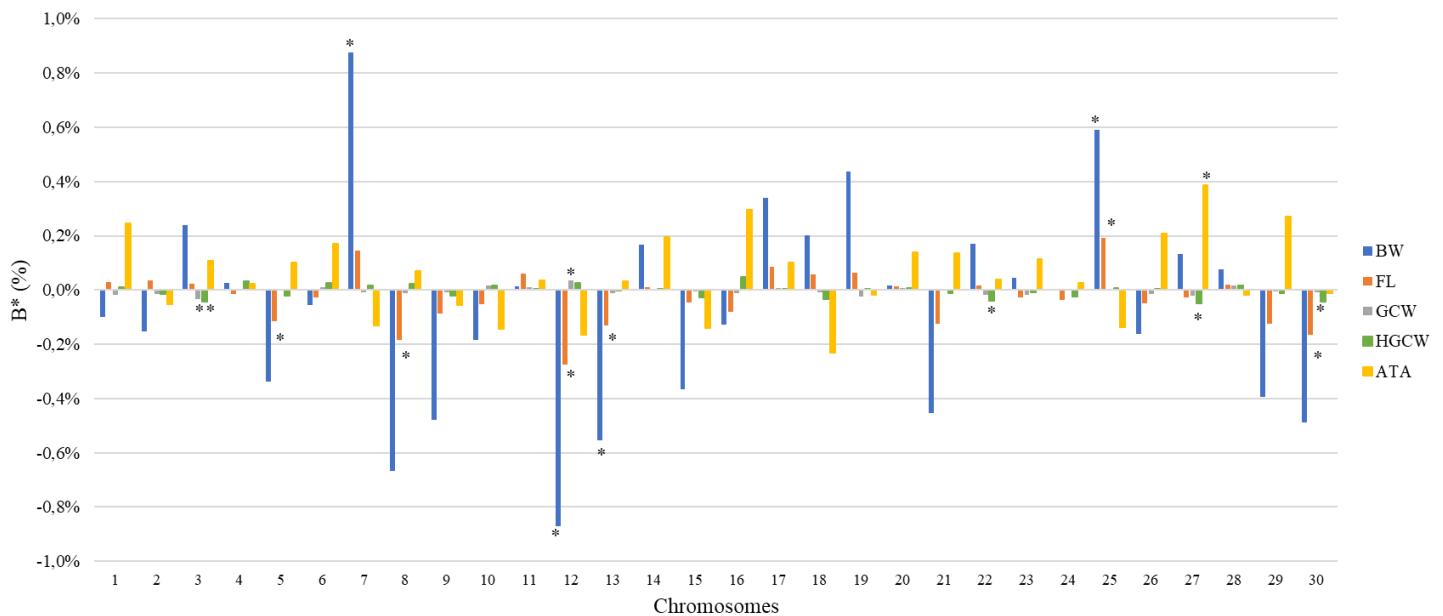


Figure Annexe 2.1. Estimation des effets B^* sur les performances de la consanguinité observée sur chaque chromosome (modèle 2). B^* correspond à l'effet de la consanguinité sur la performance moyenne pour une augmentation du niveau du coefficient de consanguinité globale F de 0,1 point, c'est-à-dire lorsque l'on passe d'un coefficient de consanguinité de 0,10 à 0,20.

Conclusion

Tout comme pour les caractères de reproduction, des effets significatifs de la consanguinité locale (à l'échelle du chromosome) ont pu être observés pour des caractères de production. En toute logique on pourra s'attendre à identifier d'autres effets significatifs de la consanguinité dans les régions au sein de chaque chromosome. La dépression de consanguinité n'est donc pas uniquement liée aux caractères de reproduction mais peut également toucher les caractères de production des truites.

Résultats Annexe 4 : Effets locaux de la consanguinité sur la survie du stade œillé jusqu'à trois semaines après la première alimentation dans la lignée synthétique INRAE.

Dans le chapitre 2 de cette thèse, nous nous sommes concentrés sur l'identification des verrous reproductifs qui pouvaient influencer la survie des descendants de chaque reproducteur. Nous avons ainsi pu estimer les paramètres génétiques du caractère de survie du stade œillé jusqu'à trois semaines après la première alimentation. La survie n'était affectée ni par le niveau de consanguinité globale des individus, ni par celui de leurs parents. Toutefois, la consanguinité locale peut-elle avoir des effets significatifs sur la survie des individus ? De plus, peut-on relier les régions ne contenant aucun ROHs (cf Chapitre 4) communes aux quatre populations de truites arc-en-ciel (adultes) avec des effets de la consanguinité locale sur la survie embryonnaire ?

Matériels et Méthodes

a) Les animaux et génotypage

La même population issue de la lignée synthétique INRAE que celle utilisée dans les travaux du chapitre 2 a été considérée, soit 860 embryons/alevins phénotypés et génotypés avec 32 725 SNPs conservés pour l'analyse après contrôle qualité. Ici c'est l'assemblage de référence Arlee qui a été utilisé, c'est pourquoi les analyses ont été effectuées sur la base de 32 chromosomes.

b) Estimation de la consanguinité locale

Sur la base des ROH identifiés dans le chapitre 2, le calcul des coefficients de consanguinité locaux, dans des régions r de 10Mb sur le chromosome k , a été effectué de la manière suivante:

$$F_{Omyk_r,i} = \frac{\sum_{k_r} Taille(ROH_i)}{LRegion_{k_r}}$$

où $\sum_{k_r} Taille(ROH_i)$ est la somme de la longueur des ROHs d'un individu i dans la région k_r . $LRegion_{k_r}$ est la longueur totale de la région considérée.

c) Modèle linéaire BLUP – effets de la consanguinité locale

Pour estimer les effets de la consanguinité locale sur la survie des individus du stade œillé jusqu'à trois semaines après la première alimentation, le modèle GBLUP suivant a été utilisé pour décrire la performance y (vivant (prend la valeur 0) ou mort (prend la valeur 100)) :

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{d} + \mathbf{Z}\mathbf{c} + \sum_r b_r \mathbf{F}_r + b_{sire} \mathbf{F}_{sire} + b_{dam} \mathbf{F}_{dam} + \mathbf{e}$$

Où $\boldsymbol{\beta}$, \mathbf{a} , \mathbf{d} , \mathbf{c} et \mathbf{e} sont respectivement les vecteurs des effets fixes environnementaux (incubateur), des effets aléatoires des valeurs génétiques additives, des effets de dominance, des effets communs dus à la combinaison mère x incubateur, et les résidus. \mathbf{X} , \mathbf{Z} et \mathbf{W} sont les matrices d'incidence pour $\boldsymbol{\beta}$, \mathbf{a} et \mathbf{d} . Les vecteurs \mathbf{a} et \mathbf{d} contiennent les valeurs génétiques additives et de dominances des 860 individus phénotypés et génotypés reliés par la matrice de relation additive génomique \mathbf{F} et la matrice de relation de dominance \mathbf{D} . Les matrices génomiques \mathbf{G} et \mathbf{D} ont été construites à l'aide du programme *parallele90* (Vitezica et al., 2016).

Les coefficients de régression b_r de y sur les vecteurs des niveaux de consanguinité locaux \mathbf{F}_r rendent compte des effets de consanguinité le long du génome. Pour cela on considère soit les 32 chromosomes, soit les n_k régions qui correspondent à 31 chromosomes entiers + 1 chromosome 'découpé' en régions de 10Mb. Par exemple, si on choisit de 'découper' le chromosome 13 en 5 régions égales, alors $n_k = 31 + 5 = 36$.

Résultats

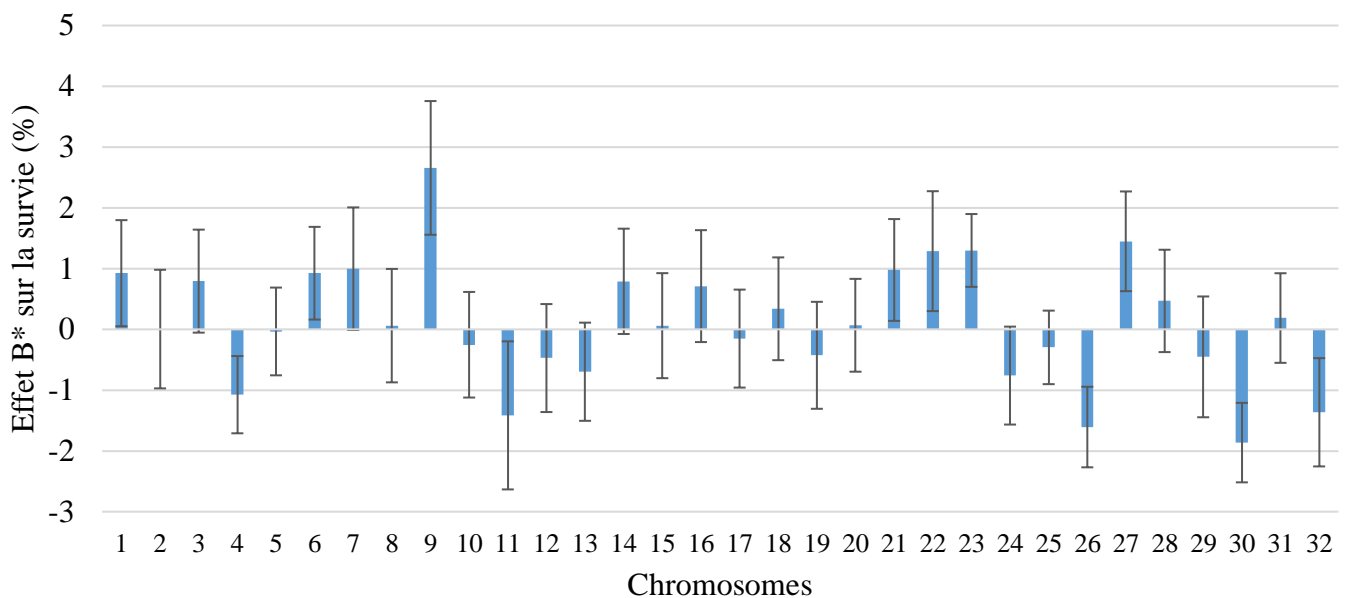


Figure annexe 3.1. Estimation des effets de la consanguinité de chaque chromosome sur la survie des individus du stade œillé jusqu'à trois semaines après la première alimentation. B^* : Pourcentage de variation de la performance pour un gain de 0,10 unité du coefficient de consanguinité. Autrement dit lorsque le coefficient de consanguinité d'un individu passe de 0,10 à 0,20 alors sa performance varie de B^*

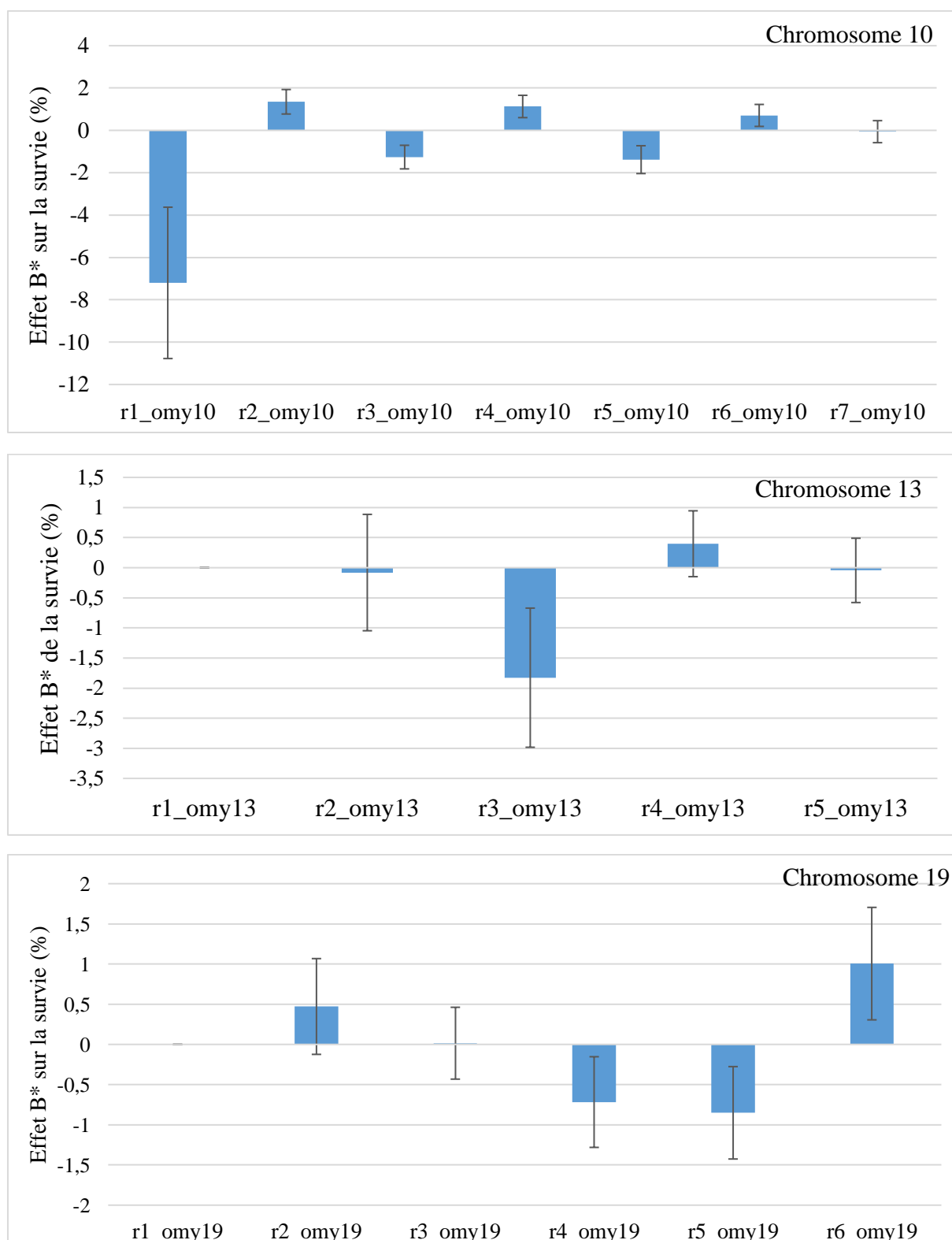


Figure annexe 3.2. Estimation des effets de la consanguinité locale des chromosomes 10, 13 et 19 sur la survie des individus du stade œillé jusqu'à trois semaines après la première alimentation. B* : Pourcentage de variation de la performance pour un gain de 0,10 unité du coefficient de consanguinité. Autrement dit lorsque le coefficient de consanguinité d'un individu passe de 0,10 à 0,20 alors sa performance varie de B*

Les effets de la consanguinité chromosomique sur la survie des individus lorsque le coefficient de consanguinité d'un individu augmente de 0,10 unité (exemple : $F = 10\% \rightarrow 20\%$), sont présentés en Figure annexe 3.1. Comme démontré dans le chapitre 3, malgré un effet de la consanguinité globale non significatif (cf Chapitre 2), la consanguinité locale peut avoir des effets significatifs, soit négatifs (chromosomes 4, 11, 26, 30, 32), soit positifs (chromosomes 6, 9, 21, 22, 23, 27).

En zoomant sur les larges régions génomiques non-consanguines (c'est-à-dire sans ROHs) partagées des chromosomes 10, 13 et 19 (cf Chapitre 4). On s'aperçoit que pour les régions concernées (r6_omy10 ; r5_omy13 ; r2_omy19) la consanguinité (identifiée au sein de région de 10Mb) ne semble avoir aucun effet significatif sur la survie des individus.

Conclusion

Nous avons pu voir que la consanguinité locale a des effets variables et significatifs le long du génome sur le caractère de survie précoce. De plus si l'on regarde plus précisément les effets de la consanguinité des régions ne contenant pas de ROHs (populations adultes) identifiées dans le Chapitre 4 de cette thèse, on peut voir que la consanguinité ne semble pas avoir d'effet sur la survie embryonnaire.

Toutefois, ici nous ne regardons que la survie précoce, alors que la détection des régions non-consanguines a été réalisée sur des populations de truites adultes. L'élimination des génotypes homozygotes défavorables (consanguins) de ces différentes régions pourrait survenir soit avant le stade embryonnaire, soit après l'installation du stade trophique (c'est-à-dire après 3 semaines après la première alimentation).