



**HAL**  
open science

# Intérêt et optimisation de la sélection génomique chez la truite arc-en-ciel

Jonathan d'Ambrosio

## ► To cite this version:

Jonathan d'Ambrosio. Intérêt et optimisation de la sélection génomique chez la truite arc-en-ciel. Génétique animale. Université Paris-Saclay, 2020. Français. NNT : 2020UPASB030 . tel-03209220

**HAL Id: tel-03209220**

**<https://pastel.hal.science/tel-03209220>**

Submitted on 27 Apr 2021

**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.

# Intérêt et optimisation de la sélection génomique chez la truite arc-en-ciel

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°581 Agriculture, alimentation, biologie,  
environnement et santé (ABIES)  
Spécialité de doctorat : Génétique Animale  
Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France.  
Réfèrent : AgroParisTech

**Thèse présentée et soutenue à Paris-Saclay,  
le 17 décembre 2020, par  
Jonathan D'AMBROSIO**

## Composition du Jury

<b>Etienne VERRIER</b> Professeur, AgroParisTech	Président
<b>Pierre BOUDRY</b> Directeur de Recherche, Ifremer (Centre Bretagne)	Rapporteur & Examineur
<b>Hélène GILBERT</b> Directrice de Recherche, INRAE (Centre Occitanie-Toulouse)	Rapporteur & Examinatrice
<b>Laurent BOUFFIER</b> Chargé de Recherche, INRAE (Centre Nouvelle-Aquitaine Bordeaux)	Examineur
<b>Bertrand BED'HOM</b> Professeur, Muséum National d'Histoire Naturelle	Examineur
<b>Pierrick HAFFRAY</b> Ingénieur, SYSAAF	Examineur
<b>Florence PHOCAS</b> Directrice de recherche, INRAE (Centre Île-de-France - Jouy-en-Josas - Antony)	Directrice de thèse



Tout le monde est un génie. Mais si vous jugez un poisson à sa capacité de grimper à un arbre, il vivra toute sa vie en croyant qu'il est stupide. (Albert Einstein)

## Remerciements

Un peu de pression pour cette partie malgré l'absence de présentation de résultats car je pense que c'est la partie la plus lue dans une thèse. Elle permet de remercier les gens et de montrer que la recherche n'est pas une activité solitaire et que les interactions sociales sont importantes dans les sciences.

Je remercie le **SYSAAF** de m'avoir fait confiance pour cette thèse et donner l'occasion de travailler sur un sujet important pour la filière aquacole. Je remercie **Pierrick** de m'avoir encadré, d'avoir partagé ses connaissances en génétique et surtout sur l'aquaculture qui était un domaine inconnu pour moi. Je remercie également **Romain** et **Sophie** qui ont participé à mon encadrement, avec lesquels j'ai partagé des discussions passionnantes. J'ai une pensée pour les différentes personnes qui ont participé au projet SG-Truite, **Anastasia**, **Florian**, **Pierre**. Malgré le fait que je n'étais pas physiquement dans les locaux du **SYSAAF**, l'entreprise a réussi à garder un lien avec moi durant toute la thèse avec de nombreuses réunions, présentations, séminaires. Ils ont pu partager leur enthousiasme et me faire participer à la vie de l'entreprise. J'ai pu rencontrer beaucoup de personnes passionnantes. Je remercie **Daniel**, **Rosine**, **Marie-Christine** de m'avoir intégré dans l'équipe du **SYSAAF** malgré la distance et d'avoir facilité toutes les démarches administratives.

Je remercie les entreprises de sélection pour leur confiance et leur partage. J'ai même eu la chance de participer à du phénotypage chez Viviers de Sarrance et Murgat.

Je remercie les membres ou ex-membres de l'équipe GenAqua : **Mathilde**, **Edwige**, **Francine**, **Katy**, **Tristan**, **Florence**, **Clémence**, **Carole**, **Céline**, **Delphine**, **Marc**, **Thérèse**, **Nicolas**. Même si toutes ces personnes n'ont pas eu de lien direct avec mon travail de recherche, elles ont contribué à mon épanouissement dans l'équipe avec les discussions, des échanges très

intéressants, les pauses café, les pique-niques, les nombreux gâteaux. J'avoue qu'il y a des règles assez autoritaires comme recevoir un gâteau le jour de son anniversaire :).

Même sous la pression du covid19, l'équipe est restée soudée durant le confinement grâce à de nombreuses visios.

Je remercie mon partenaire de bureau **Nicolas**, pour sa bonne humeur, les échanges sur les langages de programmation et scripts et pour le badminton (1/2 finalistes en double). Mais notre amitié n'a pas tenu à cause du confinement, j'ai changé de bureau. **Céline** a même changé d'équipe pour ne plus me supporter. Elle est restée un soutien sans faille avec de nombreuses pauses café, des réunions Tupperware et pour ses gâteaux (notamment son gâteau aux noix).

Je remercie la cheffe **Mathilde** pour son encadrement de thèse, ses encouragements, ses conseils, ses connaissances en génétique quantitative et de la filière aquacole, son expérience dans la vulgarisation scientifique qui est tellement importante (surtout durant cette période). J'ai pu participer au Salon de l'Agriculture, à la Fête des Sciences, cela m'a permis de discuter et partager mes connaissances avec beaucoup de personnes.

Je remercie surtout **Florence** pour l'encadrement de ma thèse. Je ne peux que souhaiter aux futurs thésards d'avoir un encadrement comme le mien. Elle a pris le temps pour m'encadrer, partager ses connaissances en génétique quantitative, ses nombreuses relectures, sa patience, ses conseils, sa confiance en moi pour réaliser cette thèse. Elle a su m'apporter son aide quand j'en avais besoin et me laisser me débrouiller le reste du temps. Enfin, je ne regrette vraiment pas de l'avoir eu comme encadrante et j'ai eu beaucoup de chance d'avoir croisé son chemin.

J'ai eu de la chance d'avoir fait ma thèse au sein de GABI et au centre de l'INRAE de Jouy-en-Josas. Cela m'a permis de travailler dans un pôle de recherche dynamique, permettant la rencontre de nombreuses personnes et d'avoir fait de nombreux séminaires, même au niveau national avec les réseaux de SelGen R2D2 et GDivSelGen.

Je remercie les membres de mon comité de thèse, **Pascale, Pascal, Etienne, François**, d'avoir suivi les travaux de ma thèse, partagé leurs connaissances et su me poser de bonnes questions pour faire avancer la thèse.

Il y a, au centre, une association de doctorants (ce qui à mon avis devrait être obligatoire dans tous les centres de recherches) permettant la rencontre entre doctorants pour partager de nombreux moments autour d'un barbecue, de jeux sociétés, de repas... Cela permet un échange, du soutien, de l'écoute, des encouragements. Cela m'a permis de rencontrer de nombreuses personnes très intéressantes. J'ai eu la chance d'être président de cette association durant 1 an, j'ai été un président qui déléguait très facilement (peut être un peu trop), je remercie les membres du bureau de cette année-là : **Abarna, Audrey, Anna-Charlotte, Mathieu**.

Je remercie aussi **Mathilde** du podcast « Bien dans ma thèse » qui a créé un serveur Discord au début du confinement permettant de retrouver des doctorants francophones partout dans le monde (même des SHS). Celui-ci a été un réel soutien durant le confinement et post-confinement et durant la rédaction de ce manuscrit. Il y a eu tant de discussions, de débats, de jeux, de séminaires, de soutiens, d'encouragements, d'écoute, je dirais tout simplement que c'était formidable.

Je remercie mes amis les plus proches : **Antoine, Julien, Sandra, Aurélien, Élodie**, de m'avoir soutenu durant les études et la thèse. Une pensée particulière pour **Aurélien** et **Élodie** de m'avoir supporté, encouragé et motivé tout le long de mes études et de cette thèse.

Enfin, je remercie toute ma famille proche de m'avoir encouragé dans la voie des études. Je suis le seul à être parti à l'Université, cela était une inconnue pour tous. Ils ne m'ont jamais mis de bâtons dans les roues, au contraire ils ont été un vrai soutien pour me permettre de poursuivre dans la voie que j'ai choisi. Je remercie tout particulièrement **mes parents, mon frère, Mélanie, mes deux petites nièces, Christelle, mes grands-parents** pour leur soutien sans faille, leurs encouragements et un clin d'œil supplémentaire à ma mère.

## Sommaire

<i>Remerciements</i>	<i>ii</i>
<i>Sommaire</i>	<i>1</i>
<i>Abréviations</i>	<i>5</i>
<i>Liste des tableaux</i>	<i>6</i>
<i>Liste des figures</i>	<i>9</i>
<b>Chapitre 1: Importance de l'aquaculture et de la sélection génomique chez les salmonidés dans le monde</b>	<b>13</b>
<b>1. Importance de l'aquaculture et de la trutticulture dans le monde</b>	<b>13</b>
1.1. Quelques chiffres sur la filière française de salmonidés	15
1.2. La France, un pays leader de la sélection de la truite arc-en-ciel	15
<b>2. Les principes de la sélection génomique</b>	<b>18</b>
2.1. Les modèles de prédiction génomique	18
2.1.1. Modèle animal et évaluation BLUP sur pedigree et/ou génomique	18
2.1.2. Modèle BayesC $\pi$ et sélection de variables	20
2.2. L'intérêt de la sélection génomique	22
2.3. Facteurs impactant la précision de l'évaluation génomique	23
<b>3. La sélection des salmonidés</b>	<b>25</b>
3.1. Schémas de sélection historiques en salmonidés	25
3.1.1. Sélection massale	25
3.1.2. Schéma de sélection de type « norvégien » en salmonidés	26
3.1.3. Schéma de sélection truticole français	26
3.2. Éléments sur les intérêts potentiels de la sélection génomique chez les salmonidés	30
<b>4. Les objectifs de la thèse</b>	<b>32</b>
<b>5. Références</b>	<b>33</b>
<b>Chapitre 2: Diversité génétique, effectifs efficaces et consanguinité dans les populations françaises de truite arc-en-ciel</b>	<b>39</b>
<b>1. Introduction</b>	<b>39</b>
<b>2. Analyse de l'outil génomique utilisé dans les travaux de thèse et validation des SNPs</b>	<b>41</b>
<b>3. Article sur la diversité génétique des lignées françaises de truite arc-en-ciel</b>	<b>45</b>
<b>4. Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite</b>	<b>66</b>
<b>5. Résumé</b>	<b>73</b>

<b>6. Annexes</b>	<b>74</b>
<b>7. Références</b>	<b>74</b>
<b>Chapitre 3: Architecture génétique et sélection des caractères de reproduction femelle.</b>	<b>77</b>
<b>1. Introduction</b>	<b>77</b>
<b>2. Matériels et méthodes</b>	<b>79</b>
2.1. Caractéristiques de la population étudiée	79
2.2. Phénotypes bruts collectés	80
2.3. Phénotypes élaborés	80
2.4. Génotypages	81
2.4.1. Puce moyenne densité 30 K	82
2.4.2. Puce basse densité (LD)	83
2.5. Modèles et méthodes d'analyse génétique	85
2.5.1. Modèle animal et évaluation BLUP sur pedigree	85
2.5.2. Modèle animal et évaluation GBLUP basée sur la parenté génomique	86
2.5.3. Modèle BayesC $\pi$ et sélection de variables	87
2.6. Comparaison de l'efficacité de la sélection selon différents scénarios	89
2.6.1. Les scénarios	89
2.7. Validation croisée	90
<b>3. Résultats &amp; discussion</b>	<b>91</b>
3.1. Paramètres génétiques	91
3.1.1. Héritabilité des caractères	91
3.1.2. Corrélations génétiques et phénotypiques	93
3.2. Identification de QTLs et gènes candidats	96
3.3. Efficacité de la sélection	102
3.3.1. Impacts liés aux caractéristiques de la population d'apprentissage	102
3.4. Impact de la densité en marqueurs de la puce à ADN	110
3.5. Progrès génétique et coût d'une évaluation génétique ou génomique de la production d'oeufs	114
3.5.1. Progrès génétique	114
3.5.2. Coût et gain économiques de l'amélioration génétique du nombre d'œufs de la ponte	115
<b>4. Résumé</b>	<b>118</b>
<b>5. Annexes</b>	<b>120</b>
<b>6. Références</b>	<b>136</b>
<b>Chapitre 4: Architecture génétique et sélection des caractères de production</b>	<b>141</b>
<b>1. Introduction</b>	<b>141</b>

<b>2. Matériels et méthodes</b>	<b>142</b>
2.1. Caractéristiques de la population étudiée	142
2.1.1. Phénotypes bruts collectés	143
2.1.2. Phénotypes élaborés	143
2.2. Géotypages	145
2.2.1. Puce moyenne densité (MD)	146
2.2.2. Puce basse densité (LD)	147
2.2.3. Répartition des marqueurs le long du génome	149
2.3. Modèles et méthodes d'analyse génétique	153
2.3.1. Modèle animal et évaluation BLUP sur pedigree	153
2.3.2. Modèle animal et évaluation GBLUP basée sur la parenté génomique	154
2.4. Comparaison de l'efficacité de la sélection selon différents scénarios	154
2.4.1. Les scénarios	154
2.4.2. Efficacité de la sélection	156
<b>3. Résultats &amp; discussion</b>	<b>157</b>
3.1. Paramètres génétiques	157
3.1.1. Héritabilité des caractères	157
3.1.2. Corrélations génétiques et phénotypiques entre les caractères	159
3.2. Efficacités comparées de la sélection sur pedigree ou génomique intra-génération	160
3.2.1. Evaluation avec la puce MD	160
3.2.2. Evaluation avec les puces LD	163
3.3. Efficacité de la sélection entre générations	169
3.3.1. Evaluation avec la puce MD	169
3.3.2. Evaluation avec la puce LD	175
3.4. Progrès génétiques attendus sur les caractères de découpe selon la stratégie d'évaluation:	177
<b>4. Résumé</b>	<b>178</b>
<b>5. Annexes</b>	<b>180</b>
<b>6. References</b>	<b>185</b>
<b>Chapitre 5: Architecture génétique et sélection pour la résistance à la nécrose pancréatique infectieuse.</b>	<b>189</b>
<b>1. Introduction</b>	<b>189</b>
<b>2. Matériels et méthodes</b>	<b>191</b>
2.1. Caractéristiques de la population étudiée et du challenge expérimental	191
2.2. Géotypages	192
2.3. Modèle animal et évaluations BLUP et GBLUP	193

2.4.	Modèle BayesC $\pi$ et sélection de variables	194
<b>3.</b>	<b>Résultats &amp; discussion</b>	<b>195</b>
3.1.	Challenge expérimental	195
3.2.	Paramètres génétiques	197
3.3.	QTL	199
3.4.	Efficacité de la sélection sur pedigree et génomique	202
<b>4.</b>	<b>Résumé</b>	<b>203</b>
<b>5.</b>	<b>Annexes</b>	<b>205</b>
<b>6.</b>	<b>Références</b>	<b>206</b>
	<i>Chapitre 6: Discussion générale : Optimisation des schémas de sélection génomique des truites arc-en-ciel</i>	<b>209</b>
<b>1.</b>	<b>Le choix d'outils de génotypage pertinents</b>	<b>210</b>
1.1.	Utilisation de puces LD	210
1.2.	Intérêt du séquençage ou du génotypage à haute densité	213
1.3.	Utilisation de la spectroscopie vibrationnelle pour construire la matrice de parenté	213
<b>2.</b>	<b>Mise en oeuvre de la sélection génomique dans le schéma actuel</b>	<b>215</b>
<b>3.</b>	<b>La conception d'un nouveau schéma de sélection optimisé pour l'utilisation de l'information génomique</b>	<b>217</b>
3.1.	De l'intérêt d'une mutualisation des populations de référence entre sélectionneurs	217
3.2.	De l'intérêt d'une revisite du plan d'accouplement	218
3.3.	De l'intérêt d'augmenter les effectifs de candidats à la sélection et/ou de collatéraux en population de référence	219
3.4.	De l'intégration de nouveaux caractères dans les schémas de sélection	220
<b>4.</b>	<b>Références</b>	<b>222</b>

## Abréviations

ADN : Acide désoxyribonucléique  
 BLUP : Best Linear Unbiased Prediction  
 DL : Déséquilibre de liaison  
 (G)EBV : (Genomic) Estimated Breeding Value  
 FAO : Food and Agriculture Organization of United Nations  
 F<sub>ST</sub> : Indice de fixation  
 GBLUP : Genomic BLUP  
 GWAS : Genome Wide Association Study  
 h<sup>2</sup> : Héritabilité  
 HD: High Density  
 IBD : Identity by descendance  
 IBS : Identity by state  
 kb : Kilobase  
 LD : Low Density  
 MAF : Minor allele frequency  
 Mb : Mégabase  
 MD : Medium Density  
 Ne : Effectif efficace  
 OCS : Optimal contribution selection  
 O<sub>ROH</sub> : Occurrence d'un SNP dans un ROH  
 PA : Population d'apprentissage  
 PR : Population de référence  
 PV : Population de validation  
 QTL : Quantitative Trait Locus  
 R<sub>g</sub> : Corrélacion génétique  
 ROH : Run of homozygosity  
 SG : Sélection génomique  
 ssGBLUP : Single step GBLUP  
 SNP : Single Nucleotide polymorphism

### Caractères étudiés :

SD : Semaine de ponte  
 FW : Poids de la femelle avant la ponte  
 PW : Poids de la femelle après la ponte  
 SW : Poids de la ponte  
 EN : Nombre d'oeufs dans la ponte  
 EW : Poids moyen d'un oeuf  
 ED : Diamètre moyen d'un oeuf  
 BW : Poids à l'abattage  
 GCW : Poids à l'abattage éviscéré  
 HGCW : Poids à l'abattage éviscéré et étêté  
 ATA : Surface de tissus adipeux sur la darne  
 FCA : paramètre a\* de couleur de la chair  
 FCL : paramètre L\* de couleur de la chair  
 TD : Durée de vie post-infection  
 D6 : Survie à 6 jours post-infection

## Liste des tableaux

### Chapitre 1 :

**Tableau 1 :** Liste des différentes études sur l'efficacité de la sélection génomique par rapport à une sélection sur pedigree en salmonidés. .... 31

### Chapitre 2 :

**Tableau 2 :** Corrélations de Pearson entre la proportion d'individus pour chaque SNP dans un ROH entre les différentes cohortes étudiées..... 71

**Annexe - tableau 1 :** Effectif efficace calculé à partir des différentes lignées et différentes générations sans le chromosome 5 et 20.....74

**Annexe - tableau 2 :** Effectif efficace calculé à partir des différentes lignées et différentes générations avec l'ensemble des chromosomes.....74

### Chapitre 3 :

**Tableau 3 :** Statistiques descriptives des phénotypes élaborés mesurés dans les cohortes G<sub>14</sub> et F<sub>15</sub>. .... 81

**Tableau 4 :** Nombre d'individus (génotypés et phénotypés passant les contrôles de qualité) dans le jeu de données final et structuration familiale des cohortes G<sub>14</sub> et F<sub>15</sub>. .... 82

**Tableau 5 :** Nombre de SNPs (N.SNP) et statistiques des distances (en kb) entre SNPs successifs pour chaque puce ainsi que les paramètres utilisés pour le choix des marqueurs (para. Plink). .... 84

**Tableau 6 :** Récapitulatif du choix et du nombre d'individus dans la population d'apprentissage et validation pour chaque scénario. .... 90

**Tableau 7 :** héritabilité des caractères sur la diagonale (en gras), corrélations génétiques entre les caractères au-dessus de la diagonale, corrélations phénotypiques en dessous de la diagonale..... 91

**Tableau 8 :** héritabilités estimées pour les caractères de reproduction femelle de 3 lignées de truite arc-en-ciel issues de la même population de base (Su et al. 1997) et d'une population de saumon coho (Gall and Neira 2004). .... 92

**Tableau 9 :** Corrélations génétiques estimées pour les caractères de reproduction femelle de 3 lignées de truite arc-en-ciel issues de la même population de base (Su et al. 1997) et d'une population de saumon coho (Gall and Neira 2004). .... 94

**Tableau 10 :** Statistiques descriptives des QTL détectés pour les caractères de reproduction femelle en GWAS par approches GBLUP et BayesC $\pi$ . .... 97

**Tableau 11 :** Précision moyenne (r) et coefficient moyen de dispersion (b) des 40 simulations d'évaluation BLUP et GBLUP sur les caractères de reproduction avec une population d'apprentissage de 1077 individus (PA+) ou seulement de 672 individus (PA-). Les écarts-types d'échantillonnage sont indiqués entre parenthèses. .... 103

**Tableau 12 :** Précision (r) et coefficient de dispersion (b) pour les évaluations BLUP et GBLUP selon les scénarios PA<sub>g</sub>, PA<sub>g</sub> et PA<sub>-</sub> (moyennes et écarts-types (entre parenthèses) calculés sur 40 simulations)..... 107

<b>Tableau 13</b> : Précision moyenne ( $r$ ) et coefficient de dispersion moyen ( $b$ ) d'une évaluation GBLUP pour les caractères de reproduction calculés sur 40 simulations du scénario PA+ et différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	112
<b>Tableau 14</b> : Perte de précision moyenne sur les six caractères de reproduction et écart-type d'échantillonnage (entre parenthèses) dans les scénarios PA- et PA+ des évaluations GBLUP sur puce LD par rapport à la précision des évaluations GBLUP sur puce MD. .....	113
<b>Tableau 15</b> : Progrès génétique attendu ( $\Delta G$ ) par génération pour un taux de sélection de 10% et une sélection mono-caractère et expression en % par rapport à la moyenne du caractère ( $\Delta G/y$ ).....	115
<b>Annexe - tableau 3</b> : Précision moyenne ( $r$ ) et coefficient de dispersion moyen ( $b$ ) d'une évaluation GBLUP pour les caractères de reproduction calculés sur 40 simulations du scénario PA- et différentes puces LD.....	121
<b>Chapitre 4 :</b>	
<b>Tableau 16</b> : Statistiques descriptives des phénotypes élaborés mesurés dans les deux générations des animaux passant les contrôles de qualité (génotypage et phénotypage). .....	144
<b>Tableau 17</b> : Structurations familiale des 2 739 individus (1 315 génération 9 et 1424 génération 10) passant les contrôles de qualité (génotypage et phénotypage). .....	146
<b>Tableau 18</b> : Nombre de SNPs (N. SNP), statistiques de distances (en kb) entre SNPs successifs et paramètres utilisés pour le choix des marqueurs (para. Plink) de chaque puce construite à partir de la population AQ. ....	148
<b>Tableau 19</b> : Déséquilibre de liaison ( $r^2$ ) entre les marqueurs successifs des différentes puces LD à partir des génotypes de la population AQ et pourcentage de couverture des puces par rapport à la puce MD et nombre moyenne de SNP par zone.....	150
<b>Tableau 20</b> : Récapitulatif du nombre d'individus considérés dans les populations d'apprentissage (issus des générations 9 et 10) et de validation (issus de la génération 10 sauf pour le scénario SW1) et du nombre de simulations réalisées pour chaque scénario .....	156
<b>Tableau 21</b> : hérabilités des caractères sur la diagonale (en gras), corrélations génétiques au-dessus de la diagonale, corrélations phénotypiques en dessous de la diagonale. ....	157
<b>Tableau 22</b> : hérabilités ( $h^2$ ) estimées pour le poids chez les espèces de salmonidée. ....	158
<b>Tableau 23</b> : Précision moyenne ( $r$ ) et coefficient moyen de dispersion ( $b$ ) des 40 simulations d'évaluations BLUP et GBLUP pour les scénarios SW1 et SW2. ....	161
<b>Tableau 24</b> : Perte ou gain de précision (%) des scénarios SB1 à SB5 par rapport au scénario intra-génération SW2. ....	171
<b>Tableau 25</b> : Progrès génétique théorique par génération pour un taux de sélection de 10% exprimé en unité physique ( $\Delta G$ ) et en pourcentage de la performance moyenne ( $\Delta G/y$ ) pour le poids éviscéré et étêté (HG CW) et le rendement de découpe (HG CW/BW)...	177

<b>Annexe - tableau 4 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SW1 et SW2. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	180
<b>Annexe - tableau 5 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SW1 et SW2 et les différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	181
<b>Annexe - tableau 6 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB1 à SB5. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	182
<b>Annexe - tableau 7 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB2 et SB5. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	183
<b>Annexe - tableau 8 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB2 et SB5 et les différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.....	184

## Chapitre 5 :

<b>Tableau 26 :</b> Structuration familiale des 679 familles identifiées de la cohorte challengée et génotypée. ....	192
<b>Tableau 27 :</b> Composantes de la variance pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) selon différents modèles génétiques (ecart-type d'erreur entre parenthèses) .....	198
<b>Tableau 28 :</b> Statistiques descriptives des QTL détectés pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) par approche GBLUP et BayesC $\pi$ . ....	200
<b>Tableau 29 :</b> Précision moyenne (r) et coefficient moyen de dispersion (b) d'évaluations BLUP et GBLUP pour la résistance à l'IPN : .....	202

## Chapitre 6 :

<b>Tableau 30 :</b> Gain de précision par rapport au BLUP d'une sélection intra-cohorte basée sur un GBLUP et des génotypes LD* et MD pour les caractères de reproduction et production évalués dans une population d'apprentissage d'environ 1 000 individus. .	209
--	-----

## Liste des figures

### Chapitre 1 :

- Figure 1 :** Évolution de la production de produits aquatiques (pêche et aquaculture) de 1950 à 2018 dans le monde (FAO 2020)..... 13
- Figure 2 :** principales productions aquacoles dans l'Union européenne (2017)..... 14
- Figure 3 :** Précision attendue d'après (Goddard 2009) de l'évaluation génomique en fonction de la taille de la population de référence, de l'héritabilité ( $h^2$ ) du caractère et de l'effectif efficace de la population ( $N_e$ )..... 24
- Figure 4 :** Description du schéma de sélection actuellement appliqué par les entreprises de sélection trutticole françaises (d'après Haffray, 2018)..... 28
- Figure 5 :** Illustration de la méthode de sélection massale intra-groupe PROSPER (Chevassus et al. 2004) ..... 29

### Chapitre 2 :

- Figure 6 :** Explication d'un ROH selon McQuillan et al. (2008)..... 41
- Figure 7 :** Densité des 50820 marqueurs SNPs de la puce MD avec une position unique sur le génome pour chaque 1 Mb des 29 chromosomes de l'assemblage américain. .... 43
- Figure 8 :** Visualisation du problème de polymorphisme de sonde ou duplication à partir du logiciel Axiom Analysis Suite 2.0. .... 44
- Figure 9 :** Répartition des 38 350 marqueurs fiables par Mb pour chacun des 29 chromosomes de l'assemblage américain..... 45
- Figure 10 :** Chronologie de création et de sélection des différentes cohortes des lignées commerciales étudiées durant la thèse..... 66
- Figure 11 :** Effectif efficace ( $N_e$ ) sur les 10 générations précédant la cohorte génotypée, calculé à partir du déséquilibre de liaison moyenné par chromosome (sans les chromosomes 5 et 20) pour les différentes lignées étudiées dans le projet 57K-Truite (ligne pointillée) et dans le projet SG-Truite (ligne pleine). .... 68
- Figure 12 :** Boîtes à moustaches de la consanguinité totale ( $F_{ROH}$ , boîtes à moustaches pleines) et de la consanguinité récente ( $F_{ROH>10Mb}$ , boîtes à moustaches transparentes) pour les cohortes étudiées dans le projet 57K-Truite (boîtes à moustaches hachurées) et SG-Truite (boîtes à moustaches non hachurées). .... 70
- Figure 13 :** Proportion d'individus où les SNPs sont conjointement dans des ROHs de  $SA_B$  (axe x) et SC (axe y). .... 71

### Chapitre 3 :

- Figure 14 :** Schéma de parenté entre les cohortes..... 79
- Figure 15 :** Manhattan plot des QTLs détectés en analyse BayesC $\pi$  pour les caractères de reproduction femelle..... 98

- Figure 16 :** Manhattan plot représentant l'association entre les SNP par une approche BayesC $\pi$  et les caractères SW (violet), EW (jaune), ED (rouge) avec le SNP peak pour chaque QTL évident sur le chromosome 1 entre 60 et 80 Mn (A) et le chromosome 2 entre 5Mb et 20Mb (B). ..... 101
- Figure 17 :** Boîtes à moustaches du gain de précision (%) pour le BLUP (A) et le GBLUP (B) pour chaque répétition et chaque caractère entre une population d'apprentissage de 1077 individus (PA+) et une population d'apprentissage de 672 individus (PA-). ..... 104
- Figure 18 :** Boîtes à moustaches du gain de précision (%) entre GBLUP et BLUP pour chaque simulation et chaque caractère avec une population d'apprentissage de (A) 1077 individus (PA+) ou (B) 672 individus (PA-). ..... 105
- Figure 19 :** Heatmap du gain de précision pour chaque simulation et chaque caractère entre les scénarios (A) GBLUP PA- versus BLUP PA-, (B) GBLUP PA+ versus BLUP PA+, (C) GBLUP PA+ versus GBLUP PA- et (D) GBLUP PA- versus BLUP PA+. ..... 106
- Figure 20 :** Histogramme de la précision de l'évaluation génétique pour 6 caractères de reproduction femelle en BLUP (couleur claire) et GBLUP (couleur foncée) avec la puce 30K pour 3 scénarios de population d'apprentissage (barre hachurée diagonalement : scénario PAg, barre hachurée horizontalement : scénario PAF et barre pleine : scénario PA-). ..... 108
- Figure 21 :** Histogramme de la précision des évaluations GBLUP et ssGBLUP pour les scénarios PAg, PAF et PA- pour le caractère SW (non ajusté pour FW). ..... 110
- Figure 22 :** Schéma sur 3 générations de parents sélectionnés produisant chacune 2 cohortes de 10 000 femelles reproductrices en première ponte. .... 116
- Figure 23 :** Gain économique marginal cumulé sur 90 mois d'une sélection sur pedigree (BLUP : bleu) ou génomique (GBLUP : vert) du nombre d'œufs de la ponte, selon le coût du génotypage. .... 117
- Annexe - figure 1:** Boîtes à moustaches de la perte de précision (%) entre le BLUP, les puces LD (GBLUP) versus un GBLUP avec la puce MD pour chaque caractère avec une population d'apprentissage de 1077 individus (PA+). ..... 120

## Chapitre 4 :

- Figure 24 :** Exemple de traitement numérique à partir d'une image de la darne pour définir les caractères étudiés. © Jerome Bugeon. .... 145
- Figure 25 :** Couverture du génome par les puces LD de densité 6K (absence ou présence d'au moins 1 marqueurs par zone de 1Mb couverte par la puce MD). .... 152
- Figure 26 :** Boîtes à moustache des pertes de précision (en %) des évaluations BLUP (boîtes à moustaches noires) et GBLUP en LD (boîtes à moustaches colorées) par rapport aux évaluations GBLUP en MD des caractères BW, GCW et HGCW pour chaque scénario intra-génération : A - SW1 et B - SW2. .... 165
- Figure 27 :** Boîtes à moustaches des pertes de précision (en %) des évaluations BLUP (boîtes à moustaches noires) et GBLUP en LD (boîtes à moustaches colorées) par rapport aux évaluations GBLUP en MD des caractères ATA, FCL et FCA pour chaque scénario intra-génération : A - SW1 et B - SW2. .... 168
- Figure 28 :** Boîtes à moustache de la précision et du coefficient de dispersion d'une évaluation GBLUP avec 30K marqueurs (boîtes à moustaches pleines) et une évaluation BLUP (boîtes à

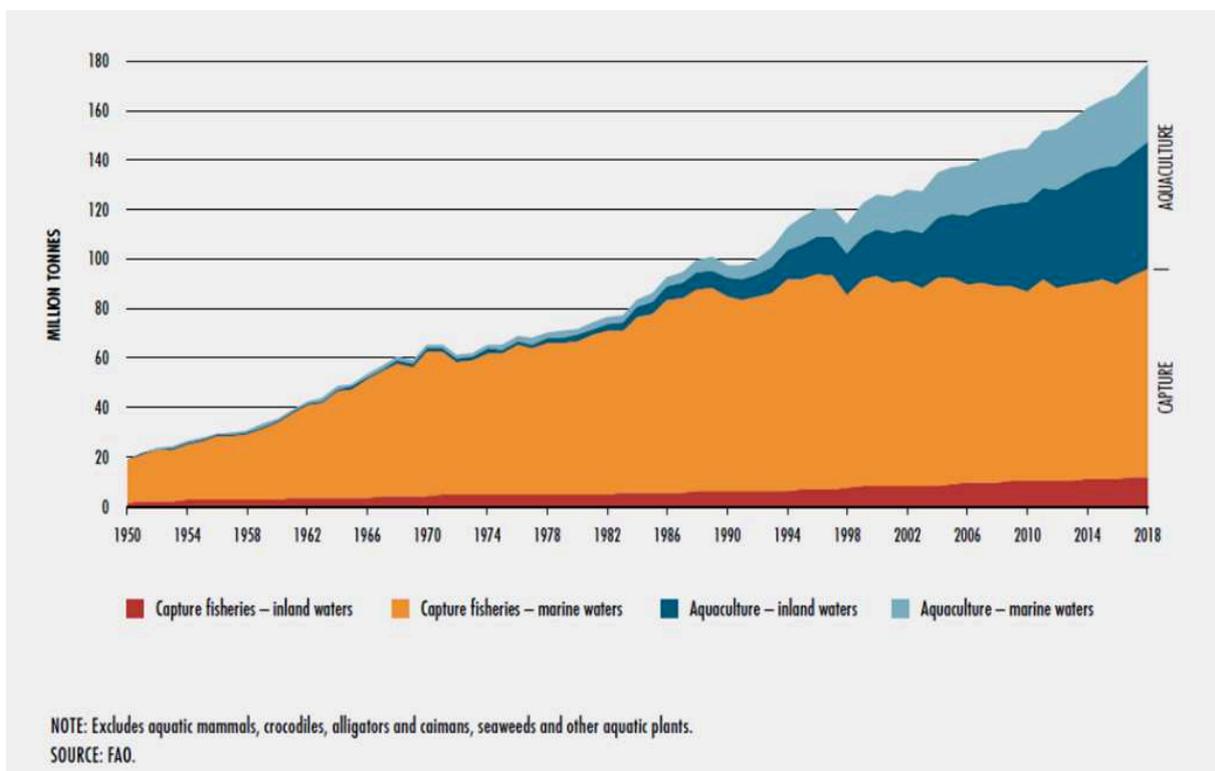
moustaches transparentes) des différents échantillonnages des 7 scénarios testés pour les caractères de poids et découpe.....	170
<b>Figure 29 :</b> Boîtes à moustache de la précision et du coefficient de dispersion d'une évaluation GBLUP avec 30K marqueurs (boîtes à moustaches pleines) et une évaluation BLUP (boîtes à moustaches transparentes) des différents échantillonnages des 7 scénarios testés pour les caractères sur la darne..	172
<b>Figure 30:</b> Précision d'une évaluation génomique pour les caractères de production pour trois scénarios : SW2 (couleur rose), SB2 (couleur verte) et SB5 (couleur violette) en fonction du nombre de marqueur (1K à 30K).	176
<b>Chapitre 5 :</b>	
<b>Figure 31 :</b> Mortalité cumulée en fonction des jours post-infection pour les individus génotypés (courbe bleue) et non génotypés (courbe verte).	196
<b>Figure 32 :</b> Boîte à moustache de la durée de vie post-infection (en jours) pour les 9 familles ayant au moins 5 descendants challengés et génotypés	197
<b>Figure 33 :</b> Manhattan plot des QTLs détectés en analyse BayesC $\pi$ (incluant un effet polygénique) pour la résistance à l'IPN.....	201
<b>Figure 34 :</b> Boîtes à moustaches du gain de précision du GBLUP versus BLUP pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) pour les 40 simulations avec une population d'apprentissage de 1 000 individus et une population de validation de 310 individus.	203
<b>Annexe - figure 2 :</b> Manhattan plot des QTLs détectés en analyse GBLUP pour la résistance à l'IPN.....	205
<b>Annexe - figure 3 :</b> Boîtes à moustaches de la précision de l'évaluation génétique BLUP (boîtes à moustache transparentes) et GBLUP (boîtes à moustache pleines) pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) avec une population d'apprentissage de 1 000 individus.....	206
<b>Chapitre 6 :</b>	
<b>Figure 35 :</b> Description du schéma de sélection actuellement appliqué par les entreprises de sélection trutticole françaises (d'après Haffray et al. 2018).....	215



## Chapitre 1: Importance de l'aquaculture et de la sélection génomique chez les salmonidés dans le monde

### 1. Importance de l'aquaculture et de la trutticulture dans le monde

La production, issue de la pêche et de l'aquaculture, est en constante augmentation (Figure 1), passant d'environ 20 millions de tonnes en 1950 à 178 millions de tonnes en 2018 (FAO 2020). Depuis les années 1990, c'est avant tout l'aquaculture qui permet cette augmentation. Elle représentait 15% des ces produits dans les années 1990 et 46% en 2018 (soit 73 millions de tonnes) (FAO 2020). La FAO estime que la production aquatique devrait atteindre 204 millions de tonnes en 2030, dont 66% provenant de l'aquaculture. L'augmentation de l'aquaculture d'ici à 2030 concernera tous les continents, même si elle sera plus importante en Afrique (48%), Océanie (41%), Asie (32%), Amérique du Sud (30%) ou du Nord (27%) qu'en Europe (18%) (FAO 2020).

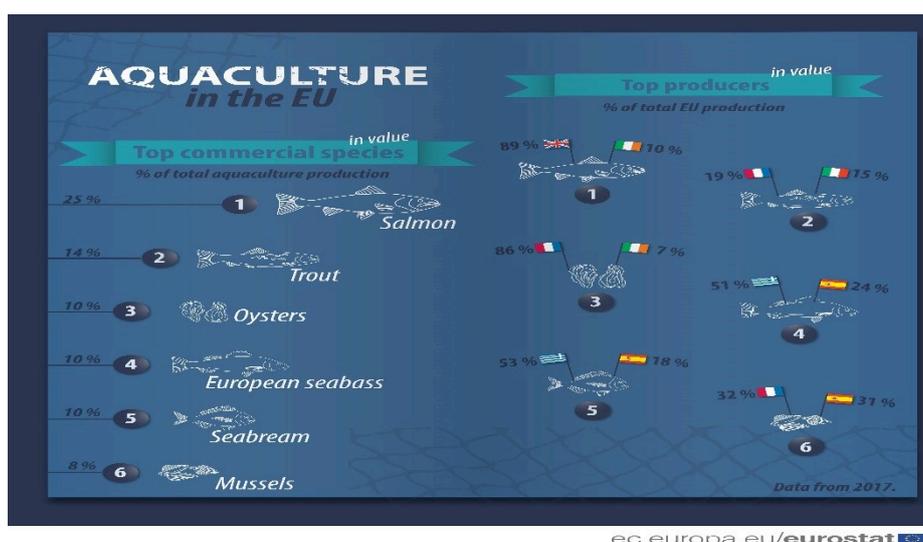


**Figure 1 :** Évolution de la production de produits aquatiques (pêche et aquaculture) de 1950 à 2018 dans le monde (FAO 2020).

L'aquaculture mondiale est dominée par l'Asie avec 89% de la production (58% par la Chine), suivie par l'Amérique du Sud avec 3,8% (1,5% par le Chili) et l'Europe avec 3,7% (1,6% par la Norvège).

Parmi les cinq espèces de poissons les plus élevées dans le monde, nous retrouvons 4 espèces de carpes représentant 32,8% de la production mondiale (10,5% carpe herbivore, *Ctenopharyngodon idellus* ; 8,8% carpe argentée, *Hypophthalmichthys molitrix* ; 7,7% carpe commune, *Cyprinus carpio* ; 5,8% carpe à grosse tête, *Hypophthalmichthys nobilis*). Le tilapia du Nil (*Oreochromis niloticus*) est la 3<sup>ème</sup> espèce la plus élevée dans le monde (8,3% de la production totale). Les principales espèces de salmonidés élevées sont le saumon Atlantique (*Salmo salar*) et la truite arc-en-ciel (*Oncorhynchus mykiss*) respectivement en 9<sup>ème</sup> position et en 15<sup>ème</sup> position dans le classement des espèces produites et représentant à elles deux 6,1% de la production mondiale respectivement (FAO 2020). Ainsi, 2 436 millions de tonnes de saumon Atlantique et 848 000 tonnes de truite arc-en-ciel ont été produites dans le monde en 2018.

Dans l'Union Européenne, les productions aquacoles principales sont en volume le saumon Atlantique (25%), puis la truite arc-en-ciel (14%), suivies par l'huître, le bar européen, la dorade et les moules soit environ 10% de la production aquacole Européenne représentait par chacune de ces espèces (Figure 2). En Europe, l'essentiel de la production de salmonidés est réalisée en Norvège avec 1,282 million de tonnes de saumon Atlantique (plus de la moitié de la production mondiale) et 68 000 tonnes de truite arc-en-ciel élevées en mer. La France est le 3<sup>ème</sup> pays en Europe en production de truite arc-en-ciel avec 33 119 de tonnes en 2018 dernière la Russie dont la production s'élève à 35 204 de tonnes en 2018 (AGRESTE 2020 et base de données FAO).



**Figure 2 :** principales productions aquacoles dans l'Union européenne (2017).

## 1.1. Quelques chiffres sur la filière française de salmonidés

La filière française aquacole est composée de 3 006 entreprises pour une production de 188 484 tonnes et un chiffre d'affaires annuel de 715 622 k€ (AGRESTE 2020). Les coquillages et crustacés représentent 77% de la production et 70% du chiffre d'affaires. Les poissons représentent 23% de la production et 30% du chiffre d'affaires. La principale espèce de poisson élevée en France est la truite arc-en-ciel.

La production de truite arc-en-ciel représente 80% de la production piscicole en France et 95% de celle des salmonidés (incluant aussi truite Fario, saumon de fontaine, ombre et omble chevalier) en 2018 (AGRESTE 2020). La production en France de truite arc-en-ciel se fait pour sa quasi-totalité en eau douce. L'élevage en mer est très minoritaire (125 tonnes par an) (AGRESTE 2020).

Outre la production de poissons, celle d'œufs de salmonidés pour la consommation est estimée à 111 tonnes par an, soit près du triple de la production française de caviar d'esturgeon (40 tonnes par an). La production d'œufs de truite représente 2% du chiffre d'affaires de la filière française de salmonidés (AGRESTE 2020).

Sur les 456 entreprises salmonicoles recensées pour l'année 2007, 25 entreprises étaient des écloséries, 126 faisaient uniquement du grossissement et 305 conduisaient les deux activités à la fois. En 2008, les 10 entreprises les plus importantes réalisaient 42% du chiffre d'affaires (AGRESTE 2011).

## 1.2. La France, un pays leader de la sélection de la truite arc-en-ciel

Les premiers essais de sélection expérimentale en aquaculture sont rapportés dès les années 1920 et les premiers programmes de sélection commerciale commencent dans les années 70 (e.g., Haffray 2019). Chavanne et al. (2016) dénombrent 10 programmes de sélection chez la truite arc-en-ciel et sept chez le saumon Atlantique en Europe.

La France a des entreprises leaders au niveau mondial en génétique de la truite arc-en-ciel avec une production totale, incluant les filiales implantées à l'étranger, avoisinant les 850 millions d'œufs embryonnés (communication personnelle SYSAAF). Les quatre entreprises françaises de sélection trutticole sont :

## Chapitre 1: Importance de l'aquaculture et de la trutticulture dans le monde

- Les Sources de l'Avance, filiale du Groupe Aqualande (<https://www.groupeaqualande.com>) créée en 1981, qui opère dans 41 piscicultures avec une production de l'ordre 15 000 tonnes de truites par an en France et en Espagne. Le groupe possède une centre de sélection et quatre écloseries de multiplication avec 650 millions d'œufs embryonnés commercialisés chaque année dans le monde à partir des sites de multiplication localisés en France et en Espagne.
- Milin Nevez, filiale de Bretagne Truite (<http://www.bretagne-truite.fr/>) commence son activité en 1996. En 2016, l'entreprise compte 60 sites d'élevage et une production de 7 000 tonnes de salmonidés. Elle vend des œufs embryonnés aux piscicultures adhérentes, mais aussi à des clients extérieurs en France et à l'international. Le programme de sélection a commencé en 2000.
- Viviers de Sarrance (<http://www.oeufsdetruite.fr/>) a été créée en 1975 et a mis en place un programme de sélection en 1995. Aujourd'hui, l'entreprise vend plus de 50 millions d'œufs embryonnés partout dans le monde.
- Charles Murgat (<https://www.charlesmurgat.com>) est la plus ancienne pisciculture familiale de France créée en 1898. Le programme de sélection chez la truite arc-en-ciel a commencé en 1994. Ils produisent quatre espèces de salmonidés (truite fario, omble chevalier, saumon de fontaine et truite arc-en-ciel).

Voici quelques exemples d'entreprises internationales avec des programmes de sélection bien avancés :

- Troutlodge (<https://www.troutlodge.com/en/>) est une entreprise des USA spécialisée en truite arc-en-ciel (rachetée par Hendrix Genetics en 2016 après le rachat de l'entreprise de sélection de saumon Atlantique Landcatch (GB) avec une production d'environ 500 millions d'œufs embryonnés de truite commercialisés dans plus de 70 pays. Dès 2015, Troutlodge a mis en place la sélection génomique pour améliorer la résistance à la flavobactérie, puis, pour la survie, la croissance, la qualité de la chair et l'efficacité alimentaire.
- AquaGen est une entreprise norvégienne spécialisée en sélection du saumon Atlantique et de la truite arc-en-ciel (<https://aquagen.no/en/>), filiale du leader mondial de la sélection de la poule pondeuse Wesjohann Group (Allemagne) aussi propriétaire des entreprises de sélection de tilapia aux Philippines (Genomar) et au Brésil (Aqualabel et AquaAmerica / AquaPorto). Son programme de sélection de la truite arc-en-ciel,

dérivant du programme conduit par Akvaforsk en 1972 a commencé en sélection familiale sur la croissance en mer et concerne maintenant une dizaine de caractères. Depuis 2015, ils ont mis en place une sélection assistée par marqueurs pour des QTLs de la résistance à la nécrose pancréatique infectieuse (IPN) (1 QTL) et/ou à la flavobactériose (2 QTLs). Depuis 2016, AquaGen a mis en place la sélection génomique pour l'une de ses lignées de saumon Atlantique et les caractères de croissance, résistance au pou de mer et résistance à une maladie des branchies (AGD : amoebic gill disease). Ils proposent des œufs embryonnés « estampillés » pour des aptitudes différentes à partir des QTLs identifiés (IPN, flavobactériose, pou de mer, maladie pancréatique du saumon causé par un alphavirus, syndrome de myopathie cardiaque, maturation sexuelle chez le saumon mâle, couleur du filet).

- Benchmark Genetics (<https://salmobreed.no/en/>) est un fond d'investissement anglais impliqué en aquaculture notamment en ayant racheté l'entreprise INVE spécialisée dans la production des aliments en éclosion mais aussi différentes entreprises de sélection aquacoles du saumon Atlantique Stofnfiskdur (Islande), SalmoBreed (Norvège), de tilapia (Honduras), de crevette blanche (Colombie) ou entrée récemment au capital du groupe grec Nireus pour la sélection du bar et de la daurade. Ils produisent environ 550 millions d'œufs embryonnés de saumon Atlantique par an. Pour le saumon, la sélection familiale a commencé en 1991 et la sélection génomique à partir de 2017. Trois lignées de saumon sont sélectionnées sur QTL pour IPN et en sélection génomique pour la résistance au pou de mer. Deux lignées sont sélectionnées pour la croissance en mer, l'une étant également sélectionnée pour la résistance à la maladie du pancréas du saumon et l'autre pour la résistance à l'anémie infectieuse du saumon (ISA). Cette entreprise propose aussi des œufs « estampillés » pour certaines aptitudes spécifiques.

On peut ainsi voir que les principales entreprises concurrentes ont déjà mis en œuvre la sélection génomique et/ou une sélection assistée par marqueurs pour des QTLs à effets forts entre 2015 et 2017. Elles proposent aux multiplicateurs et grossisseurs des œufs « estampillés » pour certains QTLs identifiés à une ou plusieurs résistances aux maladies.

## 2. Les principes de la sélection génomique

Le génotypage par puce à ADN contenant plusieurs milliers à centaines de milliers de marqueurs SNP (Single Nucleotide Polymorphism) permet, depuis la fin des années 2000, un accès et une lecture facile du patrimoine génétique des animaux, modifiant de ce fait considérablement les schémas de sélection dans toutes les espèces d'élevage (Le Roy et al. 2019). Les puces à ADN permettent la mise en œuvre de la SG qui consiste à prédire la valeur génétique d'un animal à partir de la lecture de son génotype. En effet, Meuwissen et al. (2001) démontrent la possibilité d'utiliser des milliers de marqueurs (SNPs) couvrant l'ensemble du génome pour estimer la valeur génétique des individus au travers des effets estimés pour chaque marqueur. Pour cela, les équations de prédiction des valeurs génétiques (EBV : estimate breeding value) sont établies à l'aide d'une population de référence constituée d'individus génotypés et phénotypés qui permettent d'estimer les effets des marqueurs. Dans un second temps, les EBV des candidats à la sélection génotypés mais sans performance connue, sont ensuite prédites grâce à la somme des effets estimés des marqueurs.

### 2.1. Les modèles de prédiction génomique

Les premières approches par régression ou stratégies bayésiennes de sélection de variables pour estimer les effets de SNP à des fins de prédiction génomique furent formalisées par Meuwissen et al. (2001). Une large variété de méthodes statistiques a depuis été proposée (Habier et al. 2011). Parmi celles-ci, j'ai utilisé deux modélisations différentes au cours de la thèse que je vous présente.

#### 2.1.1. Modèle animal et évaluation BLUP sur pedigree et/ou génomique

Depuis les années 1990-2000, les valeurs génétiques des animaux d'élevage sont estimées à partir d'un modèle de référence, appelé modèle animal qui est un modèle linéaire mixte décrivant les performances des individus par l'équation suivante :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

Où  $\mathbf{y}$  est le vecteur de performances,  $\mathbf{X}$  la matrice d'incidence pour le vecteur des effets fixes  $\boldsymbol{\beta}$ ,  $\mathbf{Z}$  la matrice d'incidence pour le vecteur  $\mathbf{u}$  des valeurs génétiques additives des animaux et  $\mathbf{e}$  le vecteur des résidus du modèle.

Les effets environnementaux identifiés (par exemple, les effets de la cohorte de naissance) jouant sur la performance sont inclus dans le modèle en effets fixes afin de discriminer au mieux les individus sur leurs valeurs génétiques. Les effets de milieu non identifiés sont englobés dans les résidus du modèle. Les EBV sont prédites par la méthode Best Linear Unbiased Prediction (BLUP) (Henderson 1975), permettant l'estimation conjointe des effets fixes et des effets aléatoires par la résolution du système d'équations suivant :

$$\begin{bmatrix} \mathbf{X}\mathbf{X}' & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_u^2} \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$\mathbf{A}^{-1}$  : l'inverse de la matrice de parenté des individus basé sur le pedigree ( $\mathbf{A}$ ),  $\sigma_e^2$  la variance résiduelle,  $\sigma_u^2$  la variance génétique.

Le GBLUP (ou BLUP génomique) proposé par Habier et al. (2007) est très similaire à un BLUP classique, mais la matrice de parenté sur pedigree ( $\mathbf{A}$ ) est remplacée par une matrice de parenté génomique ( $\mathbf{G}$ ) calculée à l'aide des marqueurs selon la formule suivante (VanRaden 2008) :

$$\mathbf{G} = \frac{\mathbf{z}\mathbf{z}'}{2 * \sum_{i=1}^m (p_i * (1 - p_i))},$$

où  $m$  est le nombre de SNP et  $p_i$  la fréquence de l'allèle majeur du SNP  $i$ .

Le GBLUP permet de ne prédire des EBVs que pour les individus génotypés. Cela a ensuite été étendu afin d'intégrer l'information des animaux génotypés et non-génotypés dans un modèle dit « Single step GBLUP » (Misztal et al. 2009) afin de prendre en compte en une seule étape l'ensemble des informations de phénotypes et génotypes disponibles, et de fournir simultanément une valeur génétique à tous les animaux qu'ils soient génotypés ou non.

Legarra et al. (2009) proposent pour cela de remplacer la matrice  $\mathbf{G}$  par une matrice  $\mathbf{H}$  construite ainsi :

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{G}_{12} + \mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix},$$

$\mathbf{A}$  correspondant à la matrice de parenté basée sur le pedigree,  $\mathbf{G}$  la matrice de parenté basée sur la génomique, les indices 1 et 2 représentent les individus non génotypés et génotypés, respectivement. L'inverse de la matrice  $\mathbf{H}$  pour les équations du modèle mixte est :

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

avec respectivement  $\mathbf{A}^{-1}$  et  $\mathbf{G}^{-1}$  l'inverse des matrices de parenté pedigree ( $\mathbf{A}$ ) et de parenté génomique ( $\mathbf{G}$ ), et  $\mathbf{A}_{22}^{-1}$  l'inverse de la matrice de parenté  $\mathbf{A}$  uniquement pour les individus génotypés.

Les effets des marqueurs peuvent ensuite être prédits par la résolution de l'équation suivante (Wang et al. 2012):

$$\hat{\mathbf{a}} = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{\mathbf{u}}_g \quad (2)$$

où  $\mathbf{D}$  est la matrice diagonale des poids associés aux effets de SNPs ( $\mathbf{D}=\mathbf{I}$  en GBLUP),  $\mathbf{Z}$  la matrice d'incidence qui relie les génotypes aux SNPs des individus génotypés au vecteur  $\hat{\mathbf{a}}$  des effets estimés des SNPs et  $\hat{\mathbf{u}}_g$  le vecteur des GEBV des individus génotypés ( $\hat{\mathbf{u}}_g = \mathbf{Z}\hat{\mathbf{a}}$ ).

### 2.1.2. Modèle BayesC $\pi$ et sélection de variables

Le GBLUP fait l'hypothèse que tous les marqueurs ont un effet sur le caractère, les effets étant distribués selon une loi normale de très faible variance (variance génétique divisée par le nombre de marqueurs). Ce modèle est cohérent avec l'hypothèse du modèle génétique infinitésimal de Fisher (1918) supposant qu'un caractère quantitatif est expliqué par un très grand nombre de gènes ayant tous de très faibles effets additifs. Toutefois, nous ne connaissons pas *a priori* le degré de polygénie des caractères. Certains caractères vont avoir une structure oligogénique avec un ou quelques QTLs majeurs. Par exemple, pour la résistance à l'IPN dans une lignée japonaise de truite arc-en-ciel quelques QTL expliquent 27 à 30% de la variance phénotypique (Ozaki et al. 2001), tandis que d'autres caractères se révèlent très polygéniques comme les caractères de production chez la truite arc-en-ciel (Gonzalez-Pena et al. 2016; Neto et al. 2019; Ali et al. 2020). Le GBLUP va sous-estimer les effets des marqueurs réellement importants et sur-estimer les effets des marqueurs qui ne jouent pas sur l'expression des caractères. Des méthodes de sélection de variables, comme des méthodes bayésiennes, peuvent résoudre cela. De nombreuses méthodes bayésiennes de sélection de variables permettent de s'approcher plus de la réalité biologique et d'un nombre fini de gènes jouant sur les caractères (Meuwissen et al. 2001; Habier et al. 2011).

Pour mieux estimer les effets des SNPs ayant un réel effet sur le caractère, la méthode BayesC $\pi$  (Habier et al. 2011) a été considérée dans la thèse. Le modèle considéré est un modèle de

mélange avec une proportion  $\pi$  de SNPs ayant des effets non nuls (distribués selon une loi normale de moyenne nulle et variance suivant une loi de Chi2-inverse) et une proportion  $(1-\pi)$  de SNPs sans effet par méthode de Monte-Carlo par chaîne de Markov (avec algorithme de Metropolis-Hastings). Celui-ci permet de partir sans *a priori* sur les effets des SNPs. À chaque cycle de tirage, une proportion  $\pi$  de SNPs auront un effet et une proportion  $1-\pi$  n'aura aucun effet. Un grand nombre d'itérations (plusieurs centaines de milliers) est nécessaire et les cycles retenus pour les estimations doivent être assez indépendants les uns des autres, ce qui demande un temps de calcul très important (plusieurs heures à quelques jours de CPU pour un caractère et 1300 individus). Le modèle d'analyse des performances est le suivant :

$$y_i = \mu + X\beta_i + \sum_{j=1}^n \delta_j z_{ij} a_j + e_i \quad (4),$$

avec  $y_i$  la performance de l'animal  $i$ ,  $\mu$  la moyenne du caractère,  $X$  la matrice d'incidence pour les effets fixes  $\beta$  d'une performance,  $n$  le nombre de SNP,  $z_{ij}$  le génotype (codé 0 et 2 pour les homozygotes et 1 pour l'hétérozygote) au locus  $j$  pour animal  $i$ ,  $a_j$  l'effet du SNP  $j$ ,  $e_i$  le résidu,  $\delta_j$  est une indicatrice : à l'itération  $k$ ,  $\delta_j=1$  si le SNP  $j$  est inclus dans le modèle (probabilité  $\pi$ ) et  $\delta_j=0$  s'il n'est pas inclus (probabilité  $1-\pi$ ).

Cette approche a été utilisée dans la thèse pour les études d'associations pangénomiques (GWAS) car elle permet non seulement de mieux estimer les effets des principaux QTLs, mais aussi de préciser leur localisation le long du génome par rapport à une analyse en GBLUP.

La SG a commencé par révolutionner les schémas de sélection des races bovines laitières dès 2008 grâce à l'élaboration de la première puce de génotypage 54K utilisée pour les espèces d'élevage. La SG a été mise en place dans les principales races bovines laitières à partir de 2009 (Guillaume et al. 2011). Il était attendu que le progrès génétique annuel puisse être doublé par la SG en raison de la forte réduction de l'intervalle de génération liée à l'abandon du testage sur descendance des taureaux (Boichard et al. 2016). L'augmentation du progrès génétique annuel réalisé en France grâce à la SG n'a été *in fine* que de 33, 50 et 71% pour les index de synthèse, respectivement pour les races bovines Holstein, Montbéliarde et Normande. Ces écarts entre progrès attendus et observés peuvent être expliqués par une utilisation de reproducteurs peu en cohérence avec les objectifs de sélection définis par les races et, pour la race Holstein, une nette augmentation de la consanguinité engendrant une perte de variabilité génétique (Doublet et al. 2019).

## 2.2. L'intérêt de la sélection génomique

La SG permet potentiellement une augmentation du progrès génétique annuel ( $\Delta G$ ) en intervenant sur plusieurs composantes de l'équation du sélectionneur décrite ci-dessous.

$$\Delta G = \frac{ir\sigma_g}{T},$$

- $i$  : intensité de sélection ; c'est la différence de performances moyennes exprimée en unité d'écart-type entre les parents sélectionnés et les candidats à la sélection pour le caractère soumis à la sélection.
- $r$  : la précision de l'évaluation génétique ou génomique.
- $\sigma_g$  : l'écart type génétique du caractère.
- $T$  : intervalle de génération ; c'est l'âge moyen des parents à la naissance de leurs descendants (qui seront eux-mêmes potentiellement reproducteurs), celui-ci est fortement dépendant de la biologie de l'espèce et des pratiques zootechniques.

La SG permet une évaluation précoce des candidats à la sélection, ce qui peut permettre de réduire l'intervalle de génération quand le phénotypage est tardif comme pour les caractères laitiers mesurés chez les filles d'un taureau. Cet aspect n'a que peu d'intérêt *a priori* en sélection piscicole où les caractères sont presque tous évalués avant la maturité sexuelle. Seuls les caractères de reproduction pourraient bénéficier de cet avantage.

En revanche, la SG peut sensiblement améliorer la précision de la sélection, en particulier pour les caractères non mesurables sur les candidats à la sélection comme les caractères létaux intéressants en aquaculture (rendement de découpe, résistance aux maladies) qui sont mesurés sur des collatéraux. En évaluation sur pedigree, sans information d'apparentement génomique, tous les individus d'une même famille partagent alors la même valeur génétique estimée. Or, l'aléa de méiose qui distingue deux pleins-frères ne peut être capturé que par la prise en compte de leurs ressemblances génomiques. La SG permet seule d'estimer une valeur génétique différente pour chaque individu au sein d'une même famille, améliorant ainsi la précision de l'évaluation et l'efficacité de la sélection à court terme, ainsi qu'à plus long terme en évitant de surreprésenter des familles parmi les reproducteurs des nouvelles générations et donc en limitant l'augmentation de la consanguinité (Daetwyler et al. 2007; Meuwissen et al. 2020).

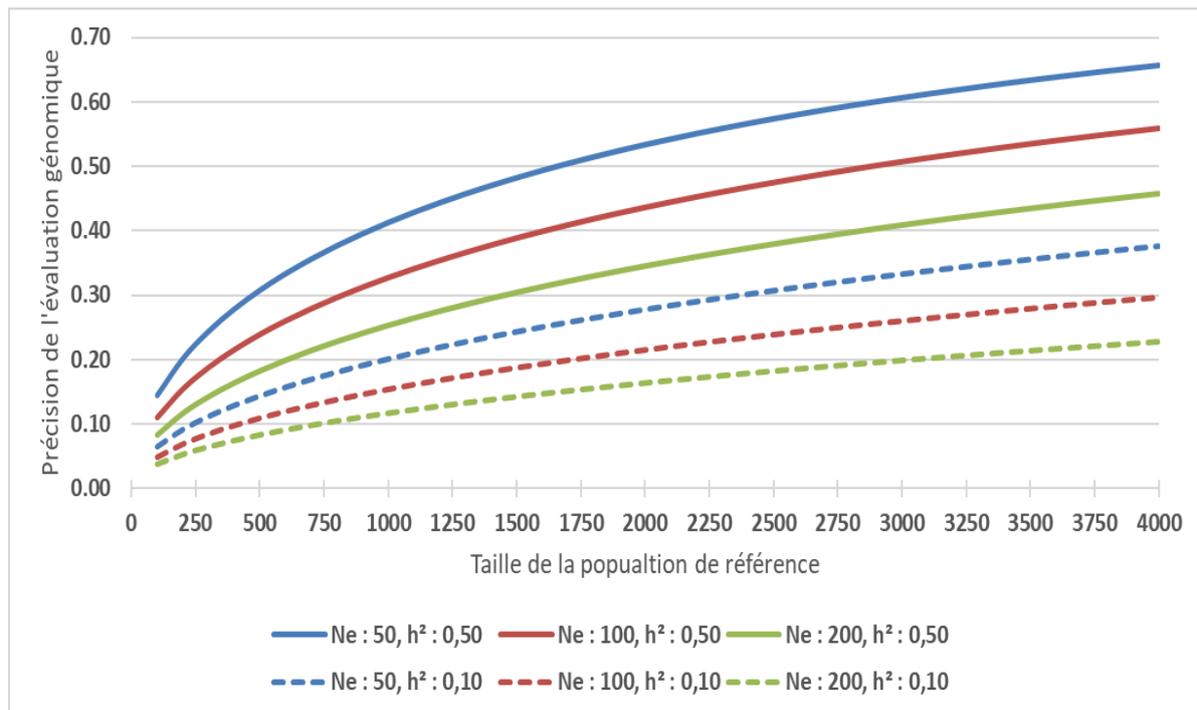
La SG peut aussi permettre d'appliquer des intensités de sélection pour des caractères coûteux à phénotyper sur les candidats à la sélection en permettant d'augmenter le nombre d'individus

génotypés (sans phénotypage) ou en faisant une présélection sur la base des données génomiques avant le phénotypage.

### 2.3. Facteurs impactant la précision de l'évaluation génomique

Daetwyler et al (2008) et Goddard (2009) ont proposé des formules déterministes pour prédire l'efficacité de la précision de la SG en fonction de la taille de la population de référence, de l'héritabilité du caractère et du nombre de segments chromosomiques indépendants du génome ( $M_e$ ), une entité théorique qui représente le nombre d'haplotypes indépendants dans la population de référence dont on doit estimer les effets grâce aux marqueurs.  $M_e$  est d'autant plus important que la diversité génétique de la population est grande et peut donc être prédite en fonction de l'effectif efficace ( $N_e$ ) de la population qui représente le nombre théorique de reproducteurs nécessaires pour garder la même diversité génétique à la génération suivante sous l'hypothèse d'une population idéale de Wright-Fisher (taille constante, croisement panmictique, contributions équivalentes des reproducteurs à la génération suivante, ni mutation, migration ou sélection). Diverses formules proposent donc une approximation de  $M_e$  en fonction de  $N_e$ , mais toutes aboutissent à des prédictions de l'efficacité de la sélection bien souvent éloignées de la réalité (Brard and Ricard 2015).

La [Figure 3](#) illustre comment la précision de la SG varie selon la formule de Goddard (2009) en fonction de  $N_e$ , la taille de la population de référence et l'héritabilité du caractère. Nous pouvons voir qu'à taille de population de référence et héritabilité constantes, la précision est d'autant plus élevée que l'effectif efficace  $N_e$  est faible. En effet, plus l'effectif efficace est faible, moins le nombre de segments indépendants dont il faut estimer précisément les effets est important : une population de référence de faible taille peut alors être suffisante pour obtenir une bonne précision de la SG.



**Figure 3 :** Précision attendue d'après (Goddard 2009) de l'évaluation génomique en fonction de la taille de la population de référence, de l'héritabilité ( $h^2$ ) du caractère et de l'effectif efficace de la population ( $N_e$ ).

Par ailleurs, plus l'effectif efficace est faible, moins nombreux sont les marqueurs nécessaires pour estimer les segments chromosomiques indépendants ( $M_e$ ). En effet, plus une population a de segments indépendants, plus grand est le nombre de marqueurs nécessaire pour couvrir l'ensemble des segments indépendants. Une étude de simulation d'un schéma de sélection basé sur 3 000 collatéraux de 100 familles à partir de 100 pères et 100 mères (1 père par mère soit environ 30 descendants par famille) avec un caractère d'héritabilité 0,4 montre peu d'évolution de la précision entre une densité de marqueurs de 0,5 et de 1  $N_e$  par Morgan (Sonesson et Meuwissen 2009). Une autre étude avec un schéma d'accouplement de 50 pères et 50 mères où chaque père est accouplé à 20 mères avec 1 descendant par famille montre une augmentation importante de la précision jusqu'à une densité de 4  $N_e$  par Morgan, puis une légère augmentation jusqu'à 8  $N_e$  par Morgan (Solberg et al. 2008). Le  $N_e$  est donc un paramètre important à estimer pour évaluer les potentialités de l'utilisation de puces basses densités.

On peut aussi observer sur la [Figure 3](#) que la précision théorique de l'évaluation génomique augmente quand la population de référence augmente (à  $N_e$  constant) car l'augmentation de la population de référence permet d'augmenter la quantité d'information permettant une meilleure estimation des effets des  $M_e$ . On remarque une augmentation très rapide de la précision avec

l'augmentation de la population de référence mais cette augmentation plafonne lorsque la population de référence devient suffisamment grande. On note aussi que plus un caractère est héritable, plus la précision de la SG est importante à taille efficace et population de référence constantes. En effet, si un caractère est peu héritable, il est nécessaire d'avoir plus individus pour bien estimer les effets des Me dont la variabilité génétique est bruitée par la variabilité environnementale.

### 3. La sélection des salmonidés

#### 3.1. Schémas de sélection historiques en salmonidés

##### 3.1.1. Sélection massale

La sélection massale individuelle est simple et peu coûteuse à mettre en place et relativement efficace pour améliorer un caractère bien héritable et mesurable chez les candidats à la sélection. Ainsi, la croissance, la morphologie, la date de ponte par exemple peuvent être améliorées par sélection massale. Mais, pour les caractères létaux, la sélection massale est aussi possible en utilisant des critères indirects de mesure qui eux peuvent être évalués sur les animaux vivants. Par exemple, la teneur en lipides des filets peut être estimée par micro-ondes (Douirin et al. 1998) ou le rendement à l'éviscération estimé par des mesures échographiques (ultrasons) (Haffray et al. 2018). Cette méthode de sélection est particulièrement efficace en aquaculture, car la pression de sélection peut être forte grâce à la fécondité importante des espèces aquacoles. Mais sans la connaissance du pedigree des individus, un risque d'augmentation de la consanguinité existe pouvant limiter le progrès génétique à long terme (Felsenstein 1965). Pour maîtriser le risque d'augmentation de consanguinité, il est possible de conduire une sélection par « walk-back » par l'assignation de parenté d'un effectif adapté de candidats grâce à la reconstruction du pedigree par l'assignation de parenté avec des marqueurs génétiques de type microsatellites ou SNPs (Doyle et Herbinger ; 1994) dans le but d'équilibrer les représentations de l'ensemble des parents fondateurs de la génération en cours de sélection (Vandeputte et Haffray 2014).

### 3.1.2. Schéma de sélection de type « norvégien » en salmonidés

La sélection généalogique en aquaculture a débuté en 1972 (Gjedrem 2010) en Norvège pour le saumon Atlantique et la truite arc-en-ciel, avant d'être étendue aux principales espèces aquacoles (e.g., Vandeputte and Haffray 2014). La sélection généalogique permet i) d'accroître la précision de l'évaluation des caractères mesurés sur le candidat en intégrant l'information familiale ; ii) de sélectionner sur des caractères létaux mesurés sur des collatéraux des candidats ; et iii) de maîtriser l'évolution de la consanguinité. Les alevins à l'éclosion étant trop petits pour être marqués physiquement et ainsi identifiables, les descendants d'une famille (descendants d'un père et d'une mère) sont élevés dans un même bassin jusqu'à ce que leur marquage soit possible à l'aide d'un transpondeur électronique. Pour réaliser ces élevages en familles séparées, il est nécessaire d'investir dans des structures d'élevage très coûteuses avant la mise en commun, après marquage, des différentes familles dans un unique bassin (jusqu'à 800 familles pour le saumon chez AquaGen). Cependant, l'élevage en bassins par famille peut amener à surestimer la composante génétique en ne permettant pas de dissocier les effets génétiques additifs des effets environnementaux et maternels. Ce dispositif est très coûteux à mettre en place pour de petites et moyennes entreprises. Dans les années 1990, l'INRA a évalué chez la truite fario *Salmo trutta* l'intérêt d'une procédure de sélection massale efficace sur la croissance en maîtrisant les effets maternels non-génétiques de différences de tailles des œufs (Chevassus et al. 2004) et les effets de compétition entre les animaux (plusieurs remises en compétition). Cette PROCÉDURE de Sélection Par Epreuves Répétées (PROSPER) a montré une réponse réalisée de +130% du poids à 1 an après quatre générations (Chevassus et al. 2004).

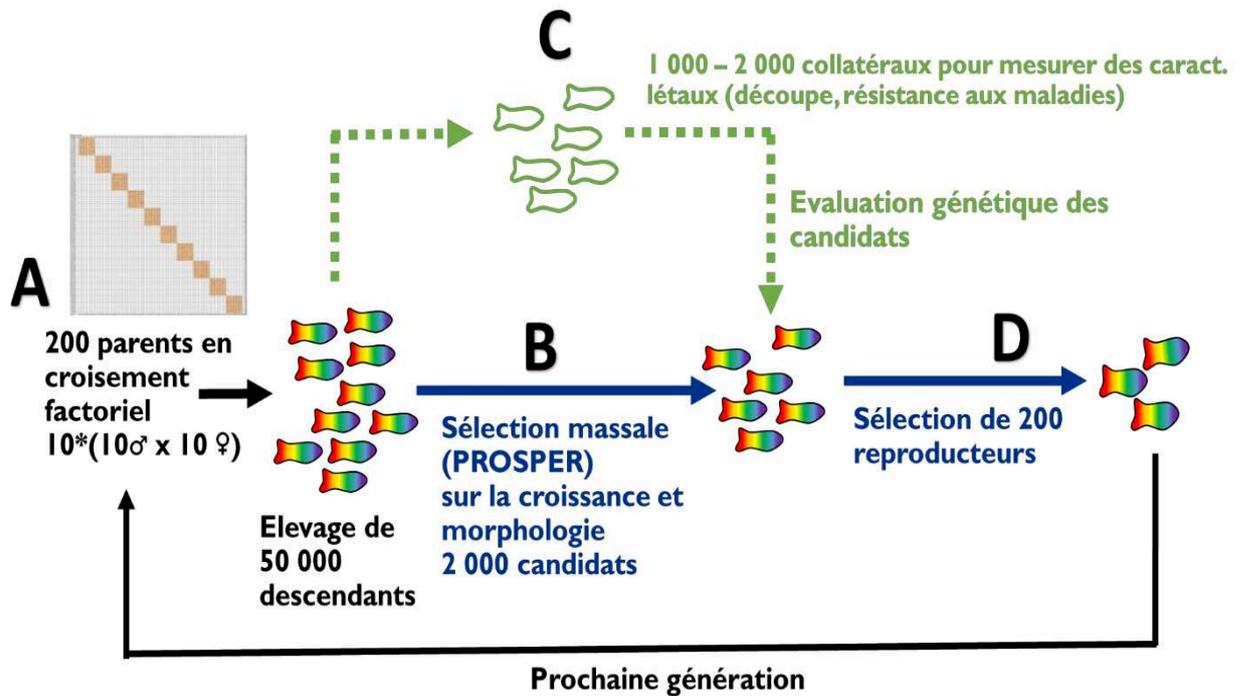
### 3.1.3. Schéma de sélection truticole français

Suite à l'estimation de la réponse à une première génération de sélection de type PROSPER par l'INRA, le Syndicat des Sélectionneurs Avicoles Français (SYSAF) a été missionné en 1990 par le Ministère de l'Agriculture pour évaluer et proposer les modalités de transfert de la procédure de sélection PROSPER. En 1991, le SYSAF est devenu le Syndicat des Sélectionneurs Avicoles et Aquacoles Français (SYSAAF) en étendant son appui technique aux salmonidés pour conseiller techniquement les entreprises souhaitant transférer la procédure de sélection de type PROSPER. Initialement conduite sans connaissance des généalogies, la

croissance est le premier caractère sélectionné. Puis des critères indirects ont été introduits pour sélectionner des caractères létaux : teneurs en lipides estimée avec le Fatmeter (micro-ondes, Douirin et al. 1998) et les rendements à l'éviscération et au filetage estimés avec des mesures échographiques et une notation morphologique des candidats (Haffray et al. 2018).

Suite au projet de recherche Microsat-Truite (Haffray et al. 2001), les entreprises françaises ont introduit dans leurs programmes de sélection massale de type PROSPER une gestion des appariements assistée par empreintes génétiques (avec des marqueurs microsatellites). La validation répétée et rapide en situation de sélection commerciale de la faisabilité de l'utilisation des empreintes génétiques a justifié une deuxième phase d'expérimentation afin d'évaluer la faisabilité de l'introduction d'une indexation BLUP à partir de performances mesurées sur apparentés phénotypés pour des caractères de rendements de découpe (éviscération, filetage) ou de qualité de chair (couleur, adiposité) dans le cadre des projets PROSPER<sup>+</sup> et Quality-Truite (Haffray et al, 2007a ; Haffray et al. 2007b).

Depuis 2004, l'assignation de parenté assistée par empreintes génétiques est utilisée en routine pour une indexation BLUP de caractères continus ou discontinus (Vandeputte et Haffray 2014). Cela permet des évaluations génétiques utilisant les informations pedigree qui améliorent la précision de la sélection en combinant phénotypes individuels et phénotypes des apparentés, en particulier pour les caractères non mesurables sur les candidats à la sélection comme les caractères létaux mesurés chez des collatéraux (rendement de découpe, qualité de la chair, caractères de reproduction ou liés au sexe). Il serait aussi possible de prendre en compte les interactions entre génotypes et environnements en considérant des collatéraux élevés dans divers milieux. Suite à des travaux de simulation, ce principe de sélection successivement massale puis généalogique a été étendu à des caractères à seuil comme la résistance à des pathologies (Chapuis et al. 2010).



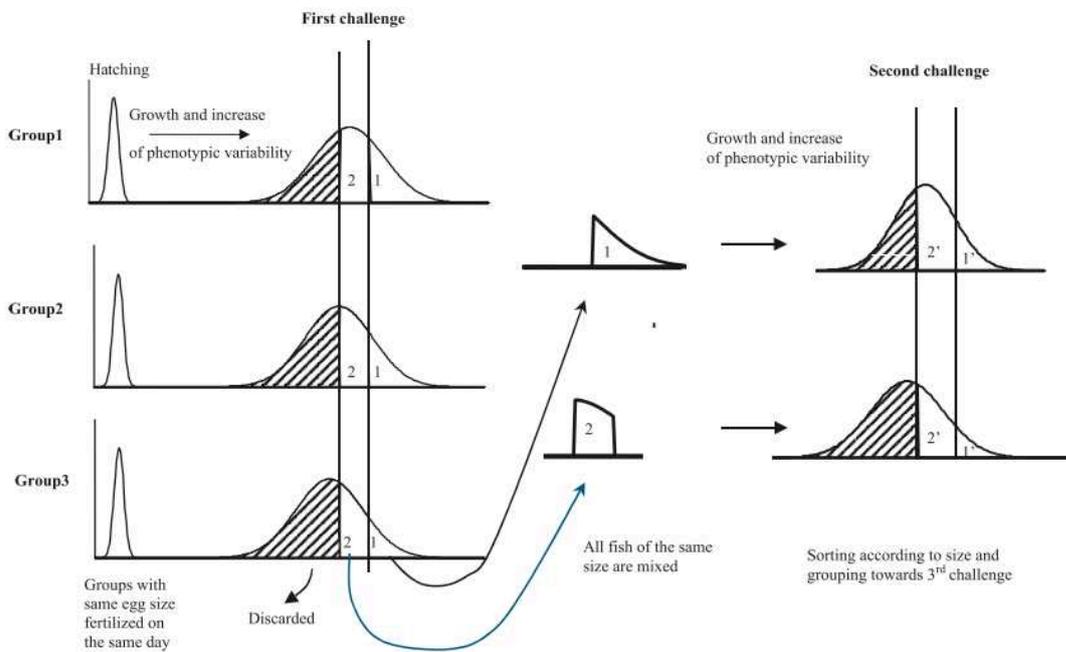
**Figure 4** : Description du schéma de sélection actuellement appliqué par les entreprises de sélection trutticole françaises (d'après Haffray, 2018)

Les schémas actuels de sélection chez la truite arc-en-ciel en France peuvent être décrits en 4 étapes :

A – Création de nombreuses familles selon plusieurs petits factoriels  $10*(10♂ \times 10♀)$  pour créer jusqu'à un millier de familles simultanément par fécondation artificielle à partir d'un plan de croisement limitant le risque de croisement entre individus pleins-germains, et équi-représentation des effectifs de descendants entre les familles. Regroupement des pontes pondues le même jour par groupes de taille moyenne d'œufs similaire (Figure 5). Les différents groupes sont ensuite alimentés en fonction de la taille des œufs ou du jour de ponte et de leurs croissances relatives afin de faire converger les moyennes de poids des différents groupes vers un même objectif de poids avant les tris de sélection (Figure 5) et de limiter les biais liés aux effets maternels initiaux.

B – Sélection massale selon la Procédure de Sélection Par Epreuves Répétées (PROSPER) avec 3 à 4 tris successifs sur la longueur avec remise en compétition des candidats entre groupes (Figure 5), permettant de limiter les effets des interactions sociales potentiellement induits par des effets non-génétiques (Chevassus et al. 2004). Cette sélection sur la croissance est suivie par une sélection massale sur les rendements

(ultrasons et morphologie externe) et/ou sur la teneur en lipides (micro-ondes) (Dourin et al. 1998 ; Haffray et al. 2018).



**Figure 5 :** Illustration de la méthode de sélection massive intra-groupe PROSPER (Chevassus et al. 2004)

C – Assignation de parenté des candidats présélectionnés de façon massive et indexation sur des caractères létaux (résistance à une pathologie, caractères de découpe) mesurés sur des collatéraux dérivés avant le 1<sup>er</sup> tri de sélection massive des candidats.

D - Sélection des reproducteurs et proposition d'accouplement accouplements pour une espérance de progrès génétique fixée en minimisant l'augmentation de consanguinité par une procédure d'optimisation des contributions (OCS : optimum contribution selection) appelée « Sélection à Parenté Minimum » (SPM). Cette méthode utilise un algorithme de recuit simulé pour résoudre le problème d'optimisation sous contraintes (Colleau et al. 2009; Chapuis et al. 2016)

Ce mode de sélection combinant successivement sélection massive et sélection généalogique de type BLUP sur apparentés a été conduit à partir de 2004 par les entreprises.

### 3.2. Éléments sur les intérêts potentiels de la sélection génomique chez les salmonidés

Le principal intérêt de la SG dans les schémas de sélection aquacole pourrait être l'augmentation de la précision de l'évaluation des valeurs génétiques pour les caractères mesurés chez des collatéraux des candidats à la sélection. Ces caractères, rendement de découpe, qualité de la chair, sont des composantes importantes des objectifs de sélection actuels.

La SG pourrait aussi permettre d'envisager une sélection sur de nouveaux caractères difficiles et/ou coûteux à mesurer. Ainsi l'un des défis majeurs pour la durabilité des filières aquacoles est l'amélioration des caractères de résistance aux maladies, pour lesquels le potentiel de sélection génétique est important (Gjedrem 2015) et dont la sélection génomique peut apporter un gain majeur d'efficacité, ces caractères étant difficiles à sélectionner dans les schémas généalogiques (Ødegård et al. 2014). Les premières études visant à évaluer l'efficacité de la SG en salmonidés ont été publiées dès 2015 pour le saumon Atlantique et 2016 pour la truite arc-en-ciel. Ces études portent principalement sur les caractères de production et de résistance aux maladies (Tableau 1). Les gains de précision estimés pour la SG par rapport à la sélection sur pedigree varient selon les caractères, de 8% (survie à l'IPN) à plus de 300% (rendement en filet) en fonction des travaux (Tableau 1). Les gains sont d'autant plus importants que les précisions sont faibles en sélection généalogique classique.

Les premières puces à ADN avec plusieurs milliers de marqueurs SNPs datent de 2014 en aquaculture. Elles concernent le saumon Atlantique pour des puces à haute densité (HD) (112K (Houston et al. 2014), 200K (Yáñez et al. 2014)) et 2015 chez la truite arc-en-ciel avec une puce moyenne densité 57K Axiom® Trout Genotyping Array de ThermoFisher (Palti et al. 2015).

Études	Esp.	Caractères étudiés	h <sup>2</sup>	N° SNP	PR	BLUP	SG	Gain (%)
Tsai et al. 2015	SA	Poids (1an)	0,48	112K	425	0,56	0,66	18
Tsai et al. 2015	SA	Poids (1an)	0,48	10K	425	0,56	0,66	18
Tsai et al. 2015	SA	Poids (1an)	0,48	5K	425	0,56	0,66	18
Tsai et al. 2015	SA	Longueur (1 an)	0,51	112K	425	0,58	0,70	21
Tsai et al. 2015	SA	Longueur (1 an)	0,51	10K	425	0,58	0,70	21
Tsai et al. 2015	SA	Longueur (1 an)	0,51	5K	425	0,58	0,69	19
Bangera et al. 2017b	SA	Rendement éviscéré	0,46	47K	2031	0,33	0,39	18
Bangera et al. 2017b	SA	Couleur filet	0,31	47K	2031	0,47	0,66	40
Bangera et al. 2017b	SA	Filet (%)	0,02	47K	2031	0,32	0,35	9
Bangera et al. 2017b	SA	Teneur en lipides dans le filet (%)	0,28	47K	2031	0,20	0,22	10
Tsai et al. 2017	SA	Poids (1 an)	-	25K	500	0,58	0,69	19
Yoshida et al. 2018	SA	Poids (25 mois)	-	37K	771	0,66	0,73	11
Gonzalez-Pena et al. 2016	TA	Poids (10 mois, mesure sur candidat)	0,37	42K	2467	0,66	0,70	6
Gonzalez-Pena et al. 2016	TA	Poids de la carcasse	0,65	42K	2467	0,28	0,55	96
Gonzalez-Pena et al. 2016	TA	Poids du filet	0,44	42K	2467	0,25	0,50	100
Gonzalez-Pena et al. 2016	TA	Rendement filets (%)	0,36	42K	2467	0,13	0,55	323
Robledo et al. 2018	SA	Score de résistance à l'Amoebic Gill Disease à <i>Neoparamoeba perurans</i> (AGD)	0,25	7K	1144	0,51	0,62	22
Robledo et al. 2018	SA	Charge qPCR à l'AGD	0,36	7K	1144	0,60	0,70	17
Bangera et al. 2017a	SA	Survie à <i>Piscirickettsia salmonis</i>	0,19	49K	2080	0,58	0,64	10
Bangera et al. 2017a	SA	Statut mort/vivant à <i>P. salmonis</i>	0,26	49K	2080	0,45	0,51	13
Tsai et al. 2017	SA	Résistance à <i>Lepeophtheirus salmonis</i>	-	25K	500	0,48	0,60	25
Tsai et al. 2016	SA	Résistance à <i>L. salmonis</i> (pop 1)	0,33	33K	425	0,48	0,60	25
Tsai et al. 2016	SA	Résistance à <i>L. salmonis</i> (pop 2)	0,22	33K	470	0,41	0,45	10
Vallejo et al. 2017	TA	Résistance <i>Flavobacterium psychrophilum</i> (survie)	0,37	35K	1473 (7900 p.)*	0,34	0,63	85
Vallejo et al. 2017	TA	Résistance <i>F. psychrophilum</i> (statut mort/vivant)	0,35	35K	1473 (7900 p.)*	0,36	0,66	83
Vallejo et al. 2018	TA	Résistance <i>F. psychrophilum</i> (statut mort/vivant)	0,35	9K	1473 (7900 p.)*	0,36	0,64	78
Vallejo et al. 2018	TA	Résistance <i>F. psychrophilum</i> (statut mort/vivant)	0,35	3K	1473 (7900 p.)*	0,36	0,61	69
Yoshida et al. 2018	TA	Résistance à <i>P. salmonis</i> (statut mort/vivant)	0,40	27K	1475	0,78	0,89	13
Yoshida et al. 2018	TA	Résistance à <i>P. salmonis</i> (statut mort/vivant)	0,56	27K	1475	0,69	0,77	13
Yoshida et al. 2019	TA	Survie au virus de la Nécrose Pancréatique Infectieuse, IPN)	0,40	38K	577 (2154 p.)*	0,49	0,53	8
Yoshida et al. 2019	TA	Statut mort/vivant à la IPN	0,35	38K	577 (2154 p.)*	0,50	0,56	12

**Tableau 1 :** Liste des différentes études sur l'efficacité de la sélection génomique par rapport à une sélection sur pedigree en salmonidés.

Esp : espèces (TA : truite arc-en-ciel, SA : saumon Atlantique) ; h<sup>2</sup> : héritabilité du caractère ; N° SNP : nombre de marqueurs utilisés dans l'étude ; PR : nombre d'individus dans la population de référence (\*, la population est constituée d'individu génotype et phénotypé + p. individus seulement phénotypés); BLUP : précision d'une évaluation sur pedigree, SG : précision de l'évaluation génomique ; gain : gain de précision en pourcentage de la SG par rapport à la sélection BLUP.

Des travaux de recherche internationaux (voir ci-dessous), conduits en partenariat avec le SYSAAF, dans le cadre du projet européen FishBoost (<http://www.fishboost.eu>), sur le bar et la daurade (Aslam et al. 2018; Palaiokostas et al. 2018) et le projet 57K chez la truite arc-en-ciel (Fraslin et al. 2019), ont permis d'évaluer l'intérêt de l'utilisation de la sélection génomique. Sur la base de l'ensemble de ces travaux, le projet de thèse sur l'évaluation de la faisabilité de la sélection génomique chez la truite arc-en-ciel est né.

#### 4. Les objectifs de la thèse

Ma thèse CIFRE (Conventions Industrielles de Formation par la Recherche) s'inscrit dans le cadre d'une collaboration de longue date entre le SYSAAF et l'INRAE pour améliorer l'efficacité et la durabilité des schémas de sélection piscicoles. Elle s'appuie sur les données recueillies au sein des projets « 57K-Truite » financé par FranceAgrimer (n°2015-0638) et « SG-Truite » financé par le Fonds Européen pour les Affaires Maritimes et de la Pêche (FEAMP) (RFEA47 0016FA 1000016), projets en partenariat fort entre le SYSAAF, ses adhérents et l'INRAE.

Dans ce contexte, l'objectif de la thèse est d'évaluer la faisabilité et l'intérêt de la sélection génomique par rapport à la sélection développée en France combinant successivement sélection massale et sélection sur apparentés (Haffray, 2018). En étudiant les principaux caractères d'intérêt aquacole et des objectifs de sélection prioritaires (reproduction, production et résistance à une maladie) pour chacune des entreprises de sélection françaises, il s'agit d'évaluer les méthodes et outils permettant de mettre en place une SG techniquement et économiquement viable.

Le premier chapitre décrivant les travaux de thèse (chapitre 2) concerne l'analyse des premières données obtenues avec la puce 57K dans le projet 57K-Truite. Les objectifs de cette partie ont été de : (i) vérifier la qualité des marqueurs de la puce 57K dans les populations françaises (ii) estimer le déséquilibre de liaison entre marqueurs dans les populations françaises pour estimer si une SG pouvait potentiellement être efficace avec cet outil (iii) étudier la diversité génétique des populations françaises de truites arc-en-ciel en sélection. Les principaux résultats sont présentés dans un article publié dans le journal *Genetics Selection Evolution*. L'étude est prolongée par une analyse de la diversité génétique des données obtenues dans SG-Truite qui

permet de confirmer les premiers résultats obtenus et d'étudier dans deux populations l'évolution de la consanguinité après 1 à 2 génération(s) supplémentaire(s) de sélection.

Les chapitres 3 à 5 traitent séparément les analyses génétiques et génomiques conduites pour chacune des trois populations des sélectionneurs du projet SG-Truite. Le chapitre 3 traite des caractères de reproduction femelle évalués dans la lignée sélectionnée par Viviers de Sarrance, le chapitre 4 des caractères de découpe et qualité de la chair mesurés dans une des lignées d'Aqualande et le chapitre 5 de la résistance à l'IPN évaluée dans la lignée en sélection de Bretagne-Truite. Chaque chapitre fait l'objet d'une étude de l'architecture des caractères en estimant les héritabilités et les corrélations génétiques entre les caractères étudiés. Des études d'association pangénomique (GWAS) sont rapportées d'une part pour les caractères de reproduction (chapitre 3) pour lesquels aucune publication n'a été rapportée à ce jour dans la littérature et d'autre part pour la résistance à IPN où deux QTLs expliquant 27 et 30% de la variance phénotypique chez la truite arc-en-ciel.

Pour les chapitres 3 à 5, l'efficacité de la sélection génomique est estimée et comparée à l'efficacité de la sélection sur pedigree. Dans les chapitres 3 et 4, nous nous sommes intéressés à l'optimisation technico-économique de la SG en identifiant des voies visant à limiter les coûts de la SG comme l'utilisation de puces LD ou le cumul d'information de cohortes différentes en population de référence.

Enfin, le chapitre 6 résume et discute les principaux résultats de la thèse afin de proposer des éléments d'optimisation pour mettre en œuvre des schémas de sélection revisités en introduisant la sélection génomique dans le but d'être techniquement plus efficaces, mais aussi économiquement viables.

## 5. Références

AGRESTE. Enquête Aquaculture 2018 - Recensement Pisciculture, culture d'algues et de cyanobactéries 2018 [internet : <https://agreste.agriculture.gouv.fr/agreste-web/disaron/ChdAgri2003/detail/>]. Vol. 3, Chiffres et données. 2020 May.

AGRESTE. Recensements 2008 de la salmoniculture et de la pisciculture marine et des élevages d'esturgeons [internet : <https://agreste.agriculture.gouv.fr/agreste-web/disaron/Dos11/detail/>]. Vol. 11, Agreste Les Dossiers. 2011.

Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M. Genome-wide identification

- of loci associated with growth in rainbow trout. *BMC Genomics*. 2020;21(1):209. <https://doi.org/10.1186/s12864-020-6617-x>
- Aslam ML, Carraro R, Bestin A, Cariou S, Sonesson AK, Bruant JS, et al. Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *BMC Genet*. 2018;19(1):1–14. <https://doi.org/10.1186/s12863-018-0631-x>
- Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM. Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics*. 2017a;18(1):121. <https://doi.org/10.1186/s12864-017-3487-y>
- Bangera, Thorland, Hillestad, Vela, Johansen, Refstie, et al. Accuracy of genomic predictions for yield traits , flesh colour and fillet fat in Atlantic salmon ( *Salmo salar* ). In: Proceedings of the World Congresson Genetics Applied to Livestock Production. 2017b. p. 11.627.
- Boichard D, Ducrocq V, Croiseau P, Fritz S. Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus - Biol*. 2016;339(7–8):274–7. <https://doi.org/10.1016/j.crvi.2016.04.007>
- Brard S, Ricard A. Is the use of formulae a reliable way to predict the accuracy of genomic selection? *J Anim Breed Genet*. 2015;132(3):207–17. <https://doi.org/10.1111/jbg.12123>
- Chapuis H, Pincent C, Colleau JJ. Optimizing selection with several constraints in poultry breeding. *J Anim Breed Genet*. 2016;133(1):3–12. <https://doi.org/10.1111/jbg.12178>
- Chapuis H, Vandeputte M, Dupont-Nivet, M. Haffray P, Quillet E. Selection for an improved disease resistance using factorial mating designs and molecular based pedigrees in fish: a simulation study. In: 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany. 2010. p. 1–6.
- Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Aquatrace Consortium, et al. A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquac Int*. 2016;24(5):1287–307. <https://doi.org/10.1007/s10499-016-9985-0>
- Chevassus B, Quillet E, Krieg F, Hollebecq M-G, Mambrini M, Fauré A, et al. Enhanced individual selection for selecting fast growing fish: the “PROSPER” method, with application on brown trout ( *Salmo trutta fario* ). *Genet Sel Evol*. 2004;36(6):643–61. <https://doi.org/10.1051/gse:2004022>
- Colleau JJ, Tual K, De Preaumont H, Regaldo D. A mating method accounting for inbreeding and multi-trait selection in dairy cattle populations. *Genet Sel Evol*. 2009;41(7):1–10. <https://doi.org/10.1186/1297-9686-41-7>
- Daetwyler HD, Villanueva B, Bijma P, Woolliams JA. Inbreeding in genome-wide selection. *J*

- Anim Breed Genet. 2007;124(6):369–76. <https://doi.org/10.1111/j.1439-0388.2007.00693.x>
- Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3(10). <https://doi.org/10.1371/journal.pone.0003395>
- Doublet AC, Croiseau P, Fritz S, Michenet A, Hozé C, Danchin-Burge C, et al. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genet Sel Evol*. 2019;51(1):1–13. <https://doi.org/10.1186/s12711-019-0495-1>
- Douirin C, Haffray P, Vallet JL, Fauconneau B. Détermination de la teneur en lipides des filets de truite arc-en-ciel *Oncorhynchus mykiss* par le Torry Fish Fat Meter. *Sci Aliments*. 1998;18:527–35.
- Doyle, R. W., Herbinger, C., M., 1994. Walk-back: The use of DNA fingerprinting for high-intensity, within-family selection in fish breeding. *Proceedings 5th World Congress, Genetics Applied to Livestock Production*. Volume: 19
- FAO. The State of World Fisheries and Aquaculture 2020. FAO; 2020. <https://doi.org/10.4060/ca9229en>
- Felsenstein J. The effect of linkage on directional selection. *Genetics*. 1965;52(2):349–63.
- Fraslin C, Brard-Fudulea S, D’Ambrosio J, Bestin A, Charles M, Haffray P, et al. 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*. 2019;50(3):293–7. <https://doi.org/10.1111/age.12777>
- Gjedrem T. The first family-based breeding program in aquaculture. *Rev Aquac*. 2010;2(1):2–15. <https://doi.org/10.1111/j.1753-5131.2010.01011.x>
- Gjedrem T. Disease resistant fish and shellfish are within reach: A review. *J Mar Sci Eng*. 2015;3(1):146–53. <https://doi.org/10.3390/jmse3010146>
- Goddard M. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136(2):245–57. <https://doi.org/10.1007/s10709-008-9308-0>
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). *Front Genet*. 2016;7(NOV). <https://doi.org/10.3389/fgene.2016.00203>
- Guillaume F, Boichard D, Ducrocq V, Fritz S. Utilisation de la sélection génomique chez les bovins laitiers. *INRAE Prod Anim*. 2011;24(4):363–8. <https://doi.org/10.20870/productions-animales.2011.24.4.3269>

- Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 2007;177(4):2389–97. <https://doi.org/10.1534/genetics.107.081190>
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12. <https://doi.org/10.1186/1471-2105-12-186>
- Haffray P., Dupont-Nivet M., Amigues Y., Vandeputte M., Pincet C., Deshayes C. and Chevassus B., 2001. Amélioration de la gestion des reproducteurs en trutticulture : Validation de l'intérêt des empreintes génétiques. Compte rendu final de la convention de recherche 033/00/C-OFIMER/CIPA de l'OFIMER.
- Haffray P., Chapuis H., Pincet C., Jégaden M., Coudurier B., Doerflinger J., Petit V., Vandeputte M. et Dupont-Nivet M., 2007a. PROSPER+. Evaluation et amélioration de la procédure de sélection PROSPER : optimisation des méthodes actuelles et introduction d'indexations sur collatéraux. Rapport final Ministère de l'Agriculture et de la Pêche, Génétique animale. Action innovante n° 2005 PA01 et IFOP n°2005/159.
- Haffray, P., Chapuis, H., Coudurier, B., Pincet, C., Petit, V., Mazeiraud, E., Doerflinger, J., Pantano, T., Chatain, B., Choubert, G., Bugeon, J., Vandeputte, M., Dupont-Nivet, M., 2007b. QUALITYTRUITE : Introduction d'une évaluation généalogique qualitative en sélection de type PROSPER. Rapport final des conventions OFIMER n°037/05/C et IFOP n°206/2005.
- Haffray P, Enez F, Bugeon J, Chapuis H, Dupont-Nivet M, Chatain B, et al. Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. *Aquaculture*. 2018;490(March):350–4. <https://doi.org/10.1016/j.aquaculture.2018.03.003>
- Haffray P. 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. 2019.
- Henderson CR. Best linear unbiased estimation and prediction under a selection model published by : international biometric society stable. *Biometrics*. 1975;31(2):423–47.
- Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics*. 2014;15(1):1–13. <https://doi.org/10.1186/1471-2164-15-90>
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 2009;92(9):4656–63. <https://doi.org/10.3168/jds.2009-2061>

- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819–29. <https://doi.org/11290733>
- Meuwissen THE, Sonesson AK, Gebregiwerigis G, Woolliams JA. Management of Genetic Diversity in the Era of Genomics. *Front Genet*. 2020;11(August):1–16. <https://doi.org/10.3389/fgene.2020.00880>
- Neto RVR, Yoshida GM, Lhorente JP, Yáñez JM. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Mol Genet Genomics*. 2019;294(3):563–71. <https://doi.org/10.1007/s00438-018-1518-2>
- Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwisse THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front Genet*. 2014;5(NOV):1–8. <https://doi.org/10.3389/fgene.2014.00402>
- Ozaki A, Sakamoto T, Khoo S, Nakamura K, Coimbra MRM, Akutsu T, et al. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol Gen Genet*. 2001;265(1):23–31. <https://doi.org/10.1007/s004380000392>
- Palaiokostas C, Cariou S, Bestin A, Bruant JS, Haffray P, Morin T, et al. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet Sel Evol*. 2018;50(1):1–11. <https://doi.org/10.1186/s12711-018-0401-2>
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K SNP array for rainbow trout. *Mol Ecol Resour*. 2015;15(3):662–72. <https://doi.org/10.1111/1755-0998.12337>
- Robledo D, Matika O, Hamilton A, Houston RD. Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3 Genes, Genomes, Genet*. 2018;8(4):1195–203. <https://doi.org/10.1534/g3.118.200075>
- Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. Genomic selection using different marker types and densities. *J Anim Sci*. 2008;86(10):2447–54. <https://doi.org/10.2527/jas.2007-0010>
- Sonesson AK, Meuwissen TH. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol*. 2009;41(1):1–9. <https://doi.org/10.1186/1297-9686-41-37>
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, et al. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol*. 2016;48(1):1–11. <https://doi.org/10.1186/s12711-016-0226-9>

- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015;16(1):1–9. <https://doi.org/10.1186/s12864-015-2117-9>
- Tsai HY, Matika O, Edwards SMK, Antolín-Sánchez R, Hamilton A, Guy DR, et al. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 Genes, Genomes, Genet*. 2017;7(4):1377–83. <https://doi.org/10.1534/g3.117.040717>
- Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, et al. 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*. 2017;49(1):1–13. <https://doi.org/10.1186/s12711-017-0293-6>
- Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Sixin L, Parsons JE, et al. 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*. 2018;(February):1–12. <https://doi.org/10.1111/jbg.12335>
- Vandeputte M, Haffray P. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front Genet*. 2014;5(DEC):1–8. <https://doi.org/10.3389/fgene.2014.00432>
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23. <https://doi.org/10.3168/jds.2007-0980>
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)*. 2012;94(2):73–83. <https://doi.org/10.1017/S0016672312000274>
- Yáñez JM, Naswa S, López ME, Bassini L, Cabrejos ME, Gilbey J, et al. Development of a 200K SNP Array for Atlantic Salmon: Exploiting Across Continents Genetic Variation. *Proc World Congr Genet Appl to Livest Prod*. 2014;Species Br:263.
- Yoshida GM, Carvalheiro R, Lhorente JP, Correa K, Figueroa R, Houston RD, et al. Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture*. 2018;491(March 2017):147–54. <https://doi.org/10.1016/j.aquaculture.2018.03.004>
- Yoshida GM, Carvalheiro R, Rodríguez FH, Lhorente JP, Yáñez JM. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics*. 2019;111(2):127–32. <https://doi.org/10.1016/j.ygeno.2018.01.008>

## Chapitre 2: Diversité génétique, effectifs efficaces et consanguinité dans les populations françaises de truite arc-en-ciel

Ce chapitre a fait l'objet d'un article publié dans *Genetics Selection Evolution : Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding*. 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) <https://doi.org/10.1186/s12711-019-0468-4>

### 1. Introduction

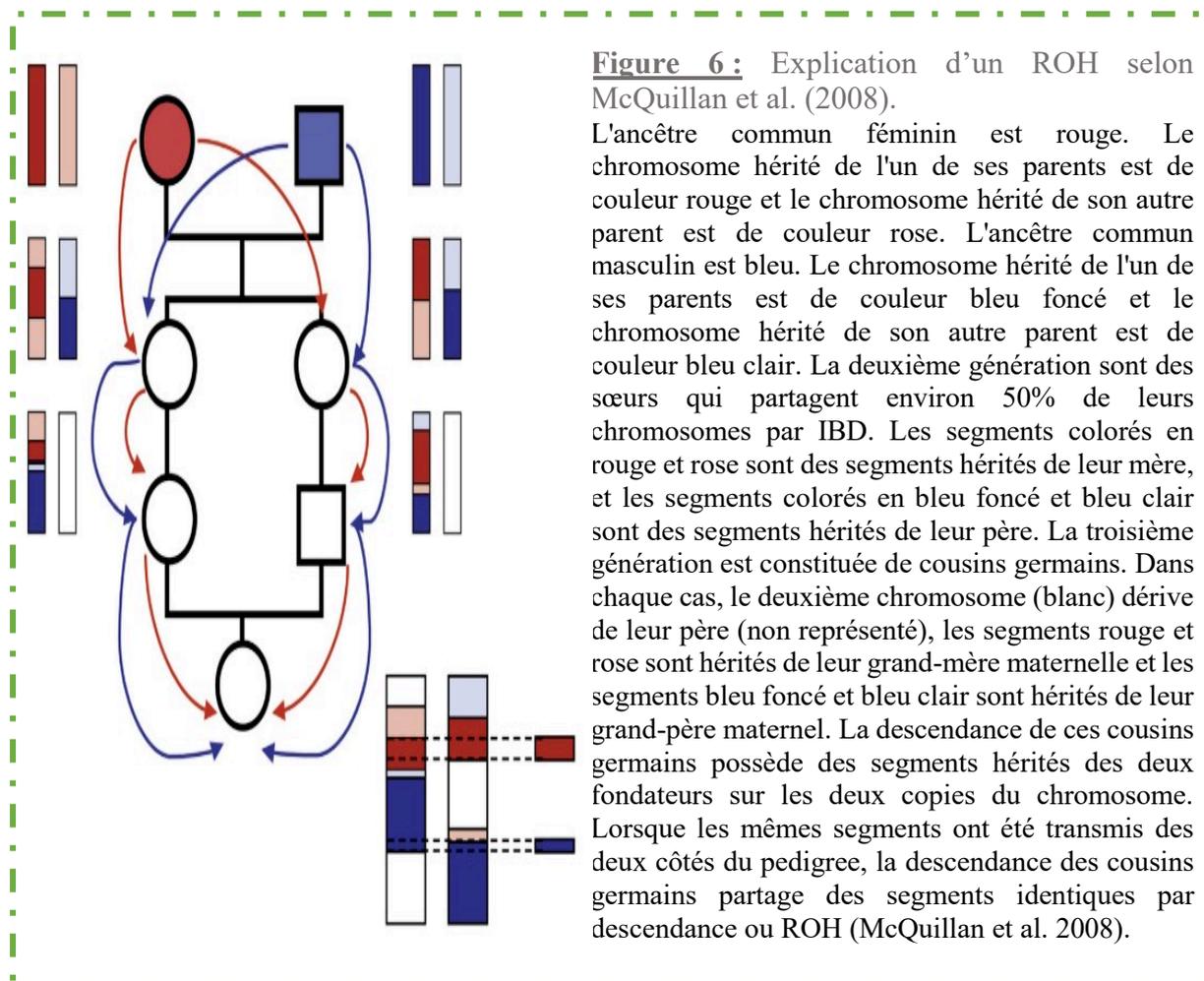
Les populations de truite arc-en-ciel sont fermées et sélectionnées depuis plusieurs décennies. La sélection rationnelle avec des méthodes basées sur des résultats scientifiques est une pratique relativement récente par rapport aux animaux domestiques terrestres et a débuté au niveau mondial au milieu des années 70 en Norvège. Une sélection intense peut conduire à une diminution rapide de la diversité génétique des populations. Or la diversité génétique est importante pour maintenir le progrès génétique à long terme et la capacité d'adaptation de la population à des changements potentiels d'environnement notamment dans un contexte de changement climatique ainsi que pour éviter la dépression de consanguinité (Felsenstein 1965). Il est donc utile d'étudier la diversité génétique des populations de truite arc-en-ciel des sélectionneurs français. Ces populations présentent entre 5 à 9 générations de sélection rationnelle réalisées avec des méthodes préconisées par l'INRAE. L'objectif de ce chapitre est de présenter une analyse de la diversité génétique au sein et entre six lignées françaises de truite arc-en-ciel expérimentales ou commerciales. Nous nous sommes focalisés sur divers indicateurs issus des données moléculaires : indice de fixation ( $F_{ST}$ ), déséquilibre de liaison (DL), effectif efficace ( $N_e$ ) et coefficient de consanguinité estimé à partir des segments d'homozygotie ( $F_{ROH}$ ).

Le  $F_{ST}$  permet d'estimer la différenciation génétique entre les différentes populations étudiées. C'est une information importante pour savoir s'il est nécessaire, dans le cadre de la sélection génomique, d'avoir une population de référence par sélectionneur ou s'il est possible de créer une population de référence mutualisée constituée d'animaux des différentes populations. Une population mutualisée permettrait de diminuer les coûts, mais nécessite une proximité génétique suffisante entre les populations pour que la précision des équations de prédiction soit suffisante.

Pour calculer l'effectif efficace ( $N_e$ ), nous nous sommes basés sur le déséquilibre de liaison entre les marqueurs (estimé par  $r^2$ ) qui définit l'association préférentielle entre deux allèles de deux loci. Outre l'importance de la connaissance du  $r^2$  pour estimer  $N_e$ ,  $r^2$  permet aussi de prédire la précision de la sélection génomique (Goddard, 2009). L'effectif efficace représente le nombre théorique de reproducteurs nécessaires pour garder la même diversité génétique à la génération suivante sous l'hypothèse d'une population idéale de Wright-Fisher (taille constante, croisements panmictiques, contributions équivalentes des reproducteurs à la génération suivante, absence de mutation, migration ou sélection). Celui-ci est un estimateur couramment utilisé, notamment par la FAO qui estime une population en danger lorsque l'effectif efficace est inférieur à 50 (FAO 1998). En effet, si le nombre de reproducteurs est faible, la dérive génétique est forte amenant à une perte de la diversité génétique très rapidement. On trouve aussi fréquemment une recommandation d'une augmentation du coefficient de consanguinité inférieure à 1% par génération pour limiter la perte de variabilité génétique. Ce qui correspond à un  $N_e$  au moins égal à 50 ( $N_e = (2*\Delta F)^{-1} : (2*0,01)^{-1}=50$ ) (FAO 1998).

De plus, nous avons utilisé un autre estimateur basé sur de longs fragments du génome où les SNPs consécutifs sont à l'état homozygote, les Runs of homozygosity ou ROH (McQuillan et al. 2008). Ces ROHs représentent des haplotypes identiques transmis des parents aux descendants. La [Figure 6](#) illustre et explique plus en détail ce concept. Une étude montre, sur des données simulées, une estimation beaucoup plus précise de la consanguinité estimée à partir des ROHs par rapport à une consanguinité estimée avec le pedigree (Forutan et al. 2018).

La première partie de ce chapitre est une description de la puce 57K utilisée durant la thèse. Elle est suivie d'un article publié dans *Genetics Selection Evolution* concernant la diversité génétique des populations expérimentales et sélectionnées de truite arc-en-ciel en France (projet 57K-Truite), puis d'une étude approfondie de la consanguinité à partir des ROHs dans les 3 populations commerciales dans lesquelles l'architecture génétique des caractères et l'efficacité de la SG (projet SG-Truite) ont été étudiées dans le cadre de la thèse.



## 2. Analyse de l'outil génomique utilisé dans les travaux de thèse et validation des SNPs

La puce 57K (puce MD) utilisée dans cette thèse est la puce à ADN publique Axiom® Trout Genotyping array (Palti et al. 2015). C'est la première puce commerciale pour la truite arc-en-ciel contenant plusieurs dizaines de milliers de marqueurs (57 501 SNP). Cette puce a été conçue à partir de :

- 30 032 SNPs provenant d'une base de données obtenues par le séquençage des régions flanquantes de sites de restriction (Restriction Associated DNA sequencing : RADs) chez des lignées isogéniques de l'Université de l'État de Washington et chez des lignées isogéniques de l'INRAE (Palti et al. 2014),

- 39 159 SNPs à partir d'un reséquençage de 12 truites d'AquaGen (sélectionneur norvégien),
- 10 964 SNPs d'autres projets (Palti et al. 2015).

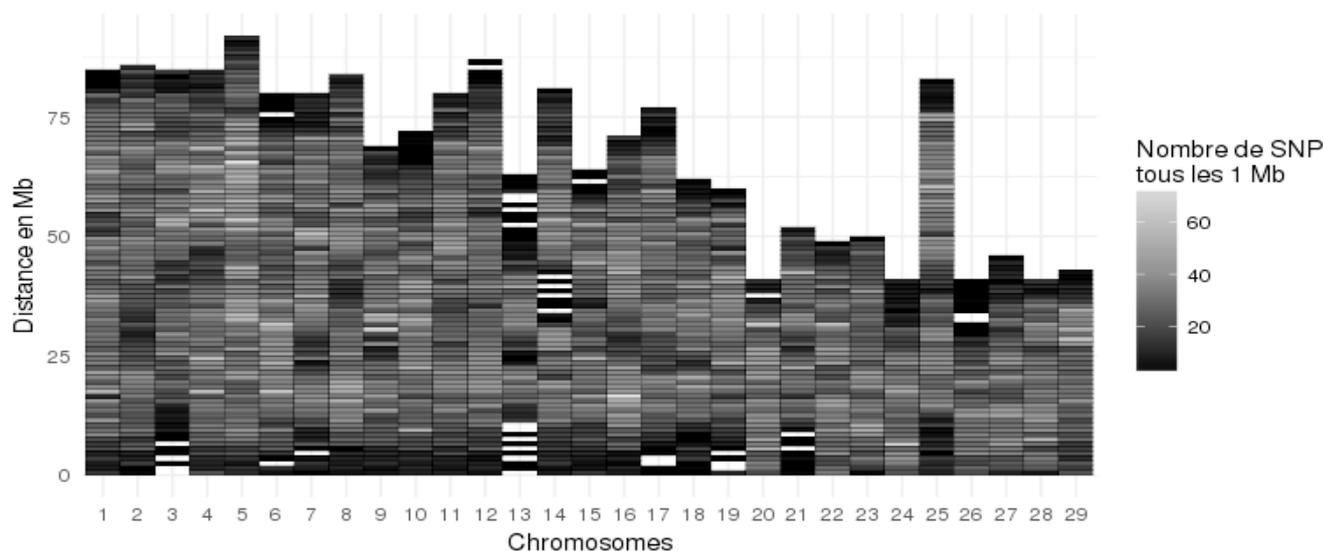
L'assemblage du génome de la lignée américaine SWANSON de truite arc-en-ciel est basé sur 29 chromosomes (Omy1 à Omy29). Dans les populations françaises, deux groupes de liaison distincts (RT 4 et RT 25) composent le chromosome 25 de l'assemblage. Aussi dans le cadre de cette thèse, nous avons considéré que le RT 4 seul constitue le chromosome 25 (0,751851 – 38,963071 Mb du Omy 25 sur l'assemblage) français et le RT 25 a été indiqué comme chromosome 30 (39,174703 – 82,450733 Mb du Omy 25 sur l'assemblage) dans notre travail (sauf dans l'article sur la diversité génétique où le chromosome 30 est appelé chromosome 25b).

Les différentes étapes listées ci-dessous ont été nécessaires pour disposer d'une liste de marqueurs fiable :

- Un blast des 57 501 sondes des SNPs sur l'assemblage du génome de la truite arc-en-ciel , Omyk\_1.0 ( accession GCF\_002163495.1 sur le site web du NCBI ; Pearse et al. 2018) a été réalisé par les bioinformaticiens, Mathieu Charles et Maria Bernard, de GABI avec une homologie supérieure à 97%. Pour 50 820 SNPs, une position unique sur l'assemblage a été trouvée. La [Figure 7](#) montre la répartition des 50 820 marqueurs le long du génome. En moyenne, il y a un marqueur tous les 40 kb. Toutefois, la répartition des marqueurs est plutôt hétérogène le long du génome avec une densité très importante notamment sur le chromosome 5. La densité est également plus forte vers les régions centromériques et moindre aux télomères. Plusieurs zones de taille importante (>1Mb) ne sont couvertes par aucun SNP sur les chromosomes 13 et 14 par exemple ([Figure 7](#)).
- Ensuite, nous avons effectué un contrôle qualité des 50820 SNPs avec le logiciel Axiom Analysis Suite 2.0 (Affymetrix 2016) avec 3418 génotypes d'individus provenant de 6 lignées de truite arc-en-ciel françaises (4 lignées commerciales et 2 lignées expérimentales INRAE) pour éliminer 5 360 marqueurs « poubelles » ([Figure 8](#)), un call-rate inférieur à 97% (2351 SNPs éliminés), les 1 689 SNPs monomorphes dans toutes les populations, les 2 995 SNPs avec aucun homozygote pour l'allèle mineur et enfin les 43 SNPs éliminés pour une fréquence de l'allèle mineur (MAF) inférieure à

0,001. Après ces étapes de contrôle qualité, 38 425 SNPs fiables ont été identifiés pour les analyses.

- Un dernier contrôle sur les 38 425 SNPs avec une  $p$ -value < 0,001 pour le test d'écart à l'équilibre d'Hardy Weinberg sur plusieurs lignées a écarté 75 SNPs.



**Figure 7 :** Densité des 50 820 marqueurs SNPs de la puce MD avec une position unique sur le génome pour chaque 1 Mb des 29 chromosomes de l'assemblage américain. En blanc, zone avec une densité de 0 marqueur.

### Explication des différents types de marqueurs rencontrés

**Marqueur polymorphe :** Présence de l'allèle homozygote majeur (AA), allèle hétérozygote (AB) et de l'allèle homozygote mineur (BB). (Figure 8-A)

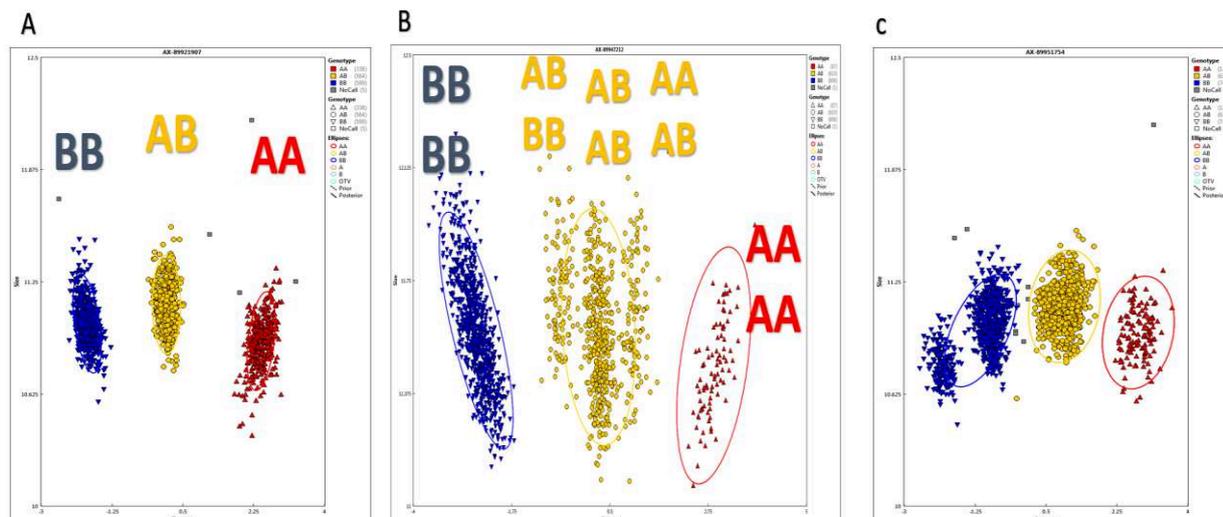
**Marqueur monomorphe :** marqueur avec uniquement un seul allèle (AA ou BB). Il est possible qu'un marqueur soit polymorphe dans une population et monomorphe dans une autre.

**Marqueur sans allèle homozygote mineur :** Présence de l'allèle homozygote majeur (AA) et de l'allèle hétérozygote (AB). Il est possible qu'un haplotype avec la présence de l'allèle mineur (BB) soit délétère.

**Marqueurs dupliqués :** La séquence d'ADN ciblée par la sonde est dupliquée (ou plus) dans le génome. Plus de 3 génotypes différents peuvent être observés (AA-AA, AA-AB, AA-BB, BB-BB, BB-AB, BB-AA ou AB-AB, AB-AA, AB-BB si deux copies de la séquence du SNP). Il est impossible de localiser le marqueur de manière unique (Figure 8-B)

**Marqueurs avec polymorphisme de sonde :** Il est possible qu'une mutation soit présente dans une population sur la séquence cible de la sonde utilisée pour identifier un SNP, ce qui engendre des erreurs de génotypage entre les 3 génotypes possibles. (Figure 8-C)

**Marqueurs « poubelles » :** Impossibilité d'identifier les différents clusters et les allèles avec le logiciel (polymorphisme d'amorce, marqueurs dupliqués, problème technique, qualité des amorces...). (Figure 8- B et C)



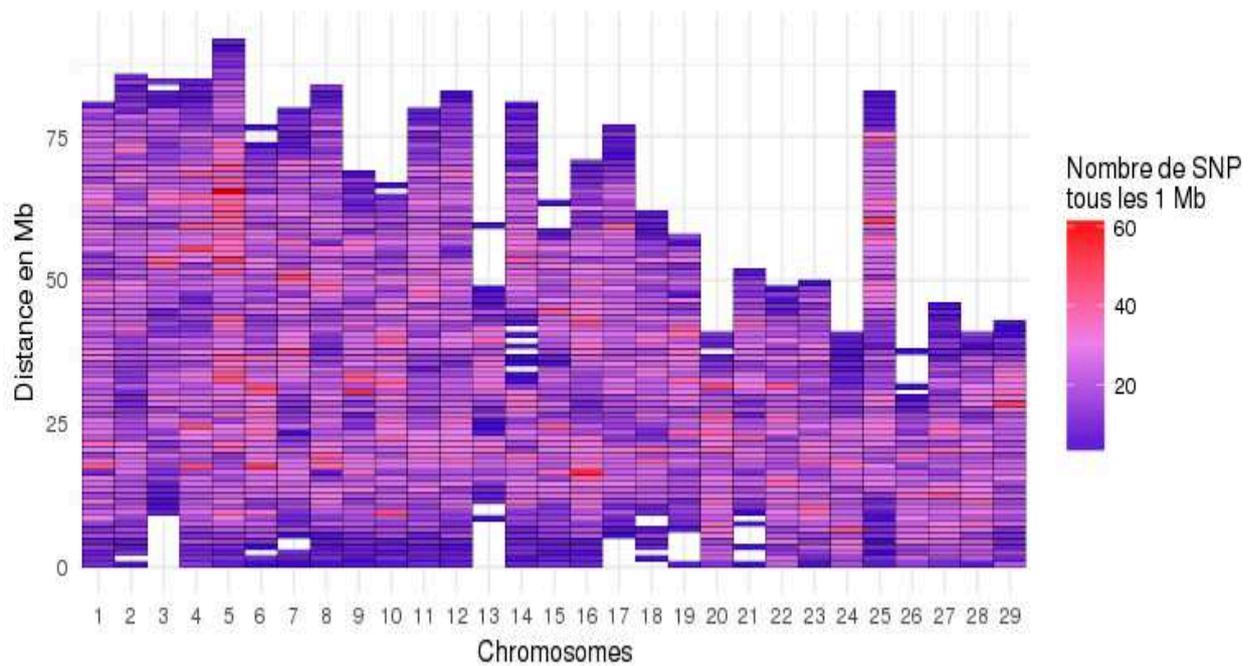
**Figure 8 :** Visualisation du problème de polymorphisme de sonde ou duplication à partir du logiciel Axiom Analysis Suite 2.0.

A : marqueur polymorphe correctement génotypé. B : marqueur avec un polymorphisme de sonde ou duplication ( 5 clusters observés au lieu de 3) . C : marqueur avec polymorphisme de sonde et 4 clusters observés.

Après les différentes étapes de tri, nous retenons une liste de 38 350 marqueurs fiables. Les études suivantes s'appuieront sur cette liste avec un dernier contrôle de qualité (sur Hardy Weinberg et la MAF) effectué au cas par cas selon les analyses.

Pour les marqueurs de confiance, on garde en moyenne 75,3% par chromosome des 50 820 marqueurs (avec un écart type de 1,8%). La **Figure 9** montre la densité des 38 350 marqueurs fiables le long du génome. La répartition des marqueurs reste hétérogène d'un chromosome à un autre, ainsi que le long des chromosomes avec une densité plus importante aux centromères par rapport aux télomères.

De nouvelles zones non couvertes par des SNPs apparaissent par exemple sur les chromosomes 13 et 14. Le tri des marqueurs a accentué l'hétérogénéité de la couverture du génome.



**Figure 9 :** Répartition des 38 350 marqueurs fiables par Mb pour chacun des 29 chromosomes de l'assemblage américain.  
En blanc, zone avec une densité de 0 marqueurs.

### 3. Article sur la diversité génétique des lignées françaises de truite arc-en-ciel

L'étude publiée dans *Genetics Selection Evolution* analyse la diversité génétique au sein et entre deux lignées françaises de truite arc-en-ciel (*Oncorhynchus mykiss*) expérimentales de l'INRAE et quatre lignées commerciales. La sélection génétique est une pratique relativement récente chez les espèces aquacoles par rapport aux animaux terrestres. Néanmoins, la variabilité génétique des lignées de salmonidés d'élevage, sélectionnées depuis plusieurs générations, doit être évaluée. En effet, une diminution significative de la variabilité génétique due à une intensité de sélection élevée aurait pu se produire, mettant potentiellement en péril le progrès génétique à long terme ainsi que les capacités d'adaptation des populations en cas de changement d'environnement. Ainsi, il est important d'évaluer l'état actuel de la diversité génétique pour adapter les pratiques de sélection si nécessaire. L'étude est basée sur une puce à ADN de densité moyenne (environ 30 000 SNPs utiles) et divers indicateurs génétiques moléculaires: indice de fixation ( $F_{ST}$ ), déséquilibre de liaison (DL), taille effective de la population ( $N_e$ ) et coefficient de consanguinité estimé à partir des segments ROH ( $F_{ROH}$ ).

Nos résultats ont montré un niveau modéré de différenciation génétique entre les lignées sélectionnées avec des  $F_{ST}$  variant de 0,08 à 0,15. Le DL diminue rapidement au cours des 100 premiers kb, mais il reste élevé sur de longues distances, conduisant à de faibles estimations du  $N_e$  dans la dernière génération. Nous avons trouvé des  $N_e$  variant de 24 à 68 selon les lignées et les méthodologies considérées. Ces résultats concordent avec des estimations de consanguinité (à partir des ROHs) qui varient de 10,0% dans une lignée expérimentale non sélectionnée à 19,5% dans une lignée commerciale sélectionnée, et qui sont clairement supérieures aux estimations correspondantes chez les ruminants ou les porcs. De plus, de fortes variations de DL et de consanguinité ont été observées le long du génome. Ceux-ci peuvent s'expliquer par des différences dans les taux de recombinaisons locales ainsi que par la fixation d'allèles favorables pour des gènes clés dans les processus de domestication ou de sélection.

Au moment de sa publication, cet article était le premier rapportant une étude basée sur les ROHs pour une espèce aquacole. La consanguinité apparaît à des niveaux modérés à élevés dans les six lignées françaises de truite arc-en-ciel, en raison des effets fondateurs au début des programmes de reproduction, d'un succès reproductif variable entre reproducteurs et aussi de la sélection effectuée pour les lignées commerciales. La gestion efficace de la consanguinité est un objectif majeur des programmes de sélection pour garantir que les populations puissent s'adapter aux futurs objectifs de sélection. Les informations moléculaires peuvent être utilisées pour gérer la consanguinité localement en autorisant une consanguinité forte dans les zones d'intérêt du génome et en la limitant sur les régions à fort impact sur les phénomènes de dépression de consanguinité.

RESEARCH ARTICLE

Open Access



# Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding

Jonathan D'Ambrosio<sup>1,2</sup>, Florence Phocas<sup>1\*</sup> , Pierrick Haffray<sup>2</sup>, Anastasia Bestin<sup>2</sup>, Sophie Brard-Fudulea<sup>3</sup>, Charles Poncet<sup>4</sup>, Edwige Quillet<sup>1</sup>, Nicolas Dechamp<sup>1</sup>, Clémence Fraslin<sup>1,2</sup>, Mathieu Charles<sup>1</sup> and Mathilde Dupont-Nivet<sup>1</sup>

## Abstract

**Background:** Selective breeding is a relatively recent practice in aquaculture species compared to terrestrial livestock. Nevertheless, the genetic variability of farmed salmonid lines, which have been selected for several generations, should be assessed. Indeed, a significant decrease in genetic variability due to high selection intensity could have occurred, potentially jeopardizing the long-term genetic progress as well as the adaptive capacities of populations facing change(s) in the environment. Thus, it is important to evaluate the impact of selection practices on genetic diversity to limit future inbreeding. The current study presents an analysis of genetic diversity within and between six French rainbow trout (*Oncorhynchus mykiss*) experimental or commercial lines based on a medium-density single nucleotide polymorphism (SNP) chip and various molecular genetic indicators: fixation index ( $F_{ST}$ ), linkage disequilibrium (LD), effective population size ( $N_e$ ) and inbreeding coefficient derived from runs of homozygosity (ROH).

**Results:** Our results showed a moderate level of genetic differentiation between selected lines ( $F_{ST}$  ranging from 0.08 to 0.15). LD declined rapidly over the first 100 kb, but then remained quite high at long distances, leading to low estimates of  $N_e$  in the last generation ranging from 24 to 68 depending on the line and methodology considered. These results were consistent with inbreeding estimates that varied from 10.0% in an unselected experimental line to 19.5% in a commercial line, and which are clearly higher than corresponding estimates in ruminants or pigs. In addition, strong variations in LD and inbreeding were observed along the genome that may be due to differences in local rates of recombination or due to key genes that tended to have fixed favorable alleles for domestication or production.

**Conclusions:** This is the first report on ROH for any aquaculture species. Inbreeding appeared to be moderate to high in the six French rainbow trout lines, due to founder effects at the start of the breeding programs, but also likely to sweepstakes reproductive success in addition to selection for the selected lines. Efficient management of inbreeding is a major goal in breeding programs to ensure that populations can adapt to future breeding objectives and SNP information can be used to manage the rate at which inbreeding builds up in the fish genome.

\*Correspondence: florence.phocas@inra.fr

<sup>1</sup> GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

Full list of author information is available at the end of the article



## Background

Rainbow trout (*Oncorhynchus mykiss*) is native to the Pacific drainages of North America and also to Kamchatka in Russia. This fish was introduced at the end of the nineteenth century to waters on all continents except Antarctica, for recreational angling and aquaculture purposes. Rainbow trout is one of the main species of fish reared in cold freshwater around the world, particularly in Europe, North America and Chile [1]. For several decades, the rainbow trout farming industry has been endeavoring to continually increase production efficiency and sales by increasing rearing densities, improving diets, water quality and recirculation technology, controlling sexual maturation and gender, or developing genetically superior lines of fish for improved growth, fillet quality and disease resistance. Recent access to the genome sequence of rainbow trout [2], genetic maps [3–5], and a medium-throughput genotyping chip [6] of 57,501 single nucleotide polymorphisms (SNPs) offer new perspectives for research and organization of trout breeding programs, which may be more effective than those that are historically based on phenotypic or genealogical selection of the broodstock.

Selective breeding can contribute to a significant decrease in the genetic variability of farmed populations, jeopardizing long-term genetic progress as well as reducing the adaptive capacities of populations in the event of a change in the environment [7]. Greater genetic variability within a population increases the likelihood that some of its individuals will have alleles that are better adapted to environmental fluctuations and are likely to survive and to transmit to their offspring alleles and favorable genetic characteristics.

Selection, mutation, migration between populations and genetic drift constitute the different evolutionary forces that can create linkage disequilibrium (LD), e.g. preferential association of alleles at different loci. The analysis of LD plays a central role in many areas of population genetics, including: the determination of genetic maps, ascertainment of levels of recombination at the population level, and estimation of effective population sizes ( $N_e$ ). The  $N_e$  of a population is a concept that was developed by Wright [8] and defined as the size of an idealized population undergoing the same rate of genetic drift as the population under study. A number of methods to estimate  $N_e$  from demographic, pedigree, or molecular data have been proposed (e.g. Leroy et al. [9]). Most of the molecular estimates are derived from LD or temporal methods that give indirect estimators of  $N_e$  [10] through the use of a genetic index: the squared correlation  $r^2$  of alleles at different gene loci using a single sample in the LD method or the standardized variance in allele frequency between two temporal samples in the

temporal approach. The temporal method is based on the theory that  $N_e$  is the only parameter that is needed to determine rates of change in allele frequency at neutral loci in a population in Hardy–Weinberg equilibrium [11]. The LD method of estimating  $N_e$ , developed by Sved [12] and modified by Hill [13], is based on the principle that in closed finite populations in approximate drift–mutation–recombination equilibrium and constant census size, associations between alleles at different neutral loci are a function of the population's  $N_e$ . Therefore, if an estimate of the rate of recombination between loci is available,  $N_e$  can be derived from the expected level of allele association across loci  $E(r^2)$ .

The maintenance of genetic diversity within a population is achieved by maximizing  $N_e$ , or equivalently, by minimizing the increase in inbreeding across generations. A molecular estimate of the inbreeding coefficient can be based on measuring long stretches of consecutive homozygous genotypes in each individual, the so-called runs of homozygosity (ROH; McQuillan et al. [14]). Long homozygous regions throughout the genome result from mating between close relatives, reduction in population size, and selection. Thus, population structure and selection effects can be assessed based on the distribution and location of ROH. Several studies have shown that characterizing inbreeding based on ROH provides a better measure of individual autozygosity because parents transmit identical haplotypes to their offspring than estimating overall inbreeding based on pedigree information, because kinships between base animals are not accounted for in pedigree files [15, 16].

In the present study, we used medium-density SNP chips to analyze the genetic diversity within and between six lines of rainbow trout. The objective of this research was to evaluate the impact of selection practices on genetic diversity through basic population indices and individual molecular genetics statistics. Our results may help to evaluate breeding practices in the light of the genetic evolution of farmed fish populations under selection. In addition, it provides commercial breeders a better understanding of the genetic composition of their selected lines and allows them to design and implement effective genomic breeding programs.

## Methods

### Background of the selected lines

The INRA synthetic line was initially developed by intercrossing several domesticated lines of rainbow trout in order to create a population with a large genetic variability. The line was constituted from a mixture of French farmed populations with some new introductions from Denmark and USA in the early 1980s. The population was then closed to outside germplasm, and has since

been bred without any intentional selection in order to maintain genetic diversity using a full factorial mating design between 2-year-old breeding animals (about 60 dams and 80 sires, each year). Individuals genotyped for the current study were from the 2006 birth year class ( $SY_n$ ), and from the 2016 birth year class (SY) (Fig. 1).

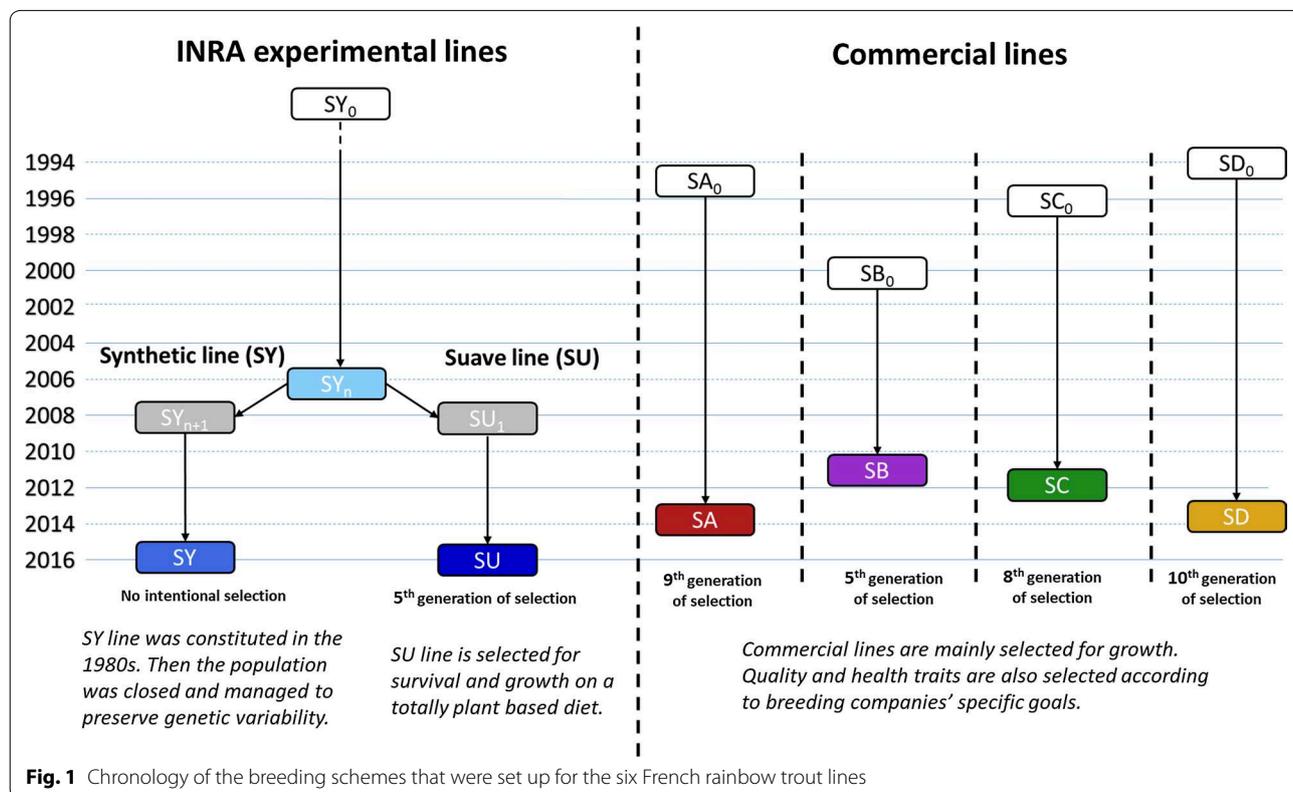
In early 2008, the first generation of the Suave ( $SU_1$ ) line was spawned from a full factorial cross among  $SY_n$  founder parents (32 dams and 44 sires). The SU line in the current study is the 5th generation of selection for survival and growth on a totally plant-based diet provided from first feeding [17]. Each generation of the SU line is created from a full factorial design between at least 40 dams and 45 sires. The selection is a sequential phenotypic selection: among the surviving animals, fish with the longest body length are selected through three to four selection events during the first year. The proportion of selected fish at each generation is around 4 to 5%. After the first three generations of selection, survival of SU fish was increased by 15% compared to SY fish and their final weight (at 197 days post-fertilization) was increased by 48% due to a 19% increase in feed intake when fish were fed a plant-based diet [18].

Breeding schemes for the four commercial lines (SA, SB, SC and SD) have been based on closed populations that were selected for at least five generations at the time

the cohorts were sampled for the study. The samples represented the 5th, 8th, 9th and 10th generation of selection in SB, SC, SA and SD breeding companies, respectively (Fig. 1). The history and the composition of these lines before the initiation of their structured selection program are unknown, however all four lines were created using sex-reversed males from INRA.

The four commercial lines are reproduced by an artificial fertilization protocol to balance the contribution of each parent by partial factorial mating of 6 to 10 dams and 10 sires according to the PROSPER procedure [19]. Each spawn is subdivided in 10 subgroups, with each subgroup being fertilized by a different sire before recombining the dam subgroups. Each of the 10 sires is used to fertilize subgroups from 6 to 10 different dams. At eyed stage, similar numbers of eggs from each dam are mixed together to balance the maternal contributions in the fry rearing tank (effectively also balancing the contribution of each sire). This procedure is expected to minimize inbreeding, since a very large number of families per generation are then created (>600) limiting the risk of mating related fish from the same full-or half-sib families [20].

Commercial lines are selected for growth traits (mainly weight and length at 18 months) by optimized within-group mass selection and a 3 to 10% selection pressure



**Fig. 1** Chronology of the breeding schemes that were set up for the six French rainbow trout lines

according to the PROSPER procedure [19]. To maintain the four commercial lines, a minimum number of broodstock (> 150) is selected at each generation. In addition, for the SA, SB and SC lines, a sib-based selection is performed based on carcass quality traits (carcass and fillet yields) assisted by ultrasound prediction [21, 22]. The management of inbreeding within these three lines has been based on DNA parentage assignment [23], and has since been improved over the last 10 years by using an optimal pedigree-based selection strategy [24].

### Genotypes

The 57,501 SNPs (57K SNP) Axiom® Trout Genotyping array [6] was used to genotype 302 females from six French lines of rainbow trout at the INRA genotyping Platform Gentyane. Animals were sampled to represent the genetic diversity of their birth cohorts by avoiding full-sib relationships for the three commercial lines with pedigree information available (SA, SB, SC). Among the genotyped animals, eight individuals with more than 45% identity-by-state (IBS) with another individual were removed from the INRA and SD populations for which pedigrees are unknown. In addition, four animals with less than 95% of the SNPs genotyped were removed from the study. After editing, 290 genotyped animals were considered in the analysis, including 48, 48, 49, 48, 32, 32 and 33 fish from SA, SB, SC, SD, SU, SY<sub>n</sub> and SY lines, respectively.

### Subset of the 57K SNPs validated for the study

Quality control of SNPs was performed in several steps. First, all 57,501 SNP probes were positioned with a BLASTn® procedure on the second genome assembly at the chromosome level Omyk\_1.0 [25, 26] and only 50,820 SNPs with a unique position were retained. Although *Oncorhynchus mykiss* is a diploid species, there is still residual tetraploidy in the 2.2 Gb of the rainbow trout genome due to the fourth round salmonid whole-genome duplication event that occurred approximately 96 million years ago [2, 27]. In addition, intraspecific variation in diploid chromosome number (2n varying between 58 and 64) exists in rainbow trout [28]. The American reference genome is based on a set of 29 pairs of chromosomes. However, all six French lines are expected to carry a set of 30 pairs of chromosomes, the American Omy25 being split into two chromosomes (French Omy25a and Omy25b).

Second, we used the Axiom Analysis Suite 2.0 software [29] to control the quality of the remaining markers on a large dataset of 3418 genotyped individuals from the six French rainbow trout lines (including the 302 individuals of the current study). Edits consisted in discarding 7711 SNPs with probe polymorphism or a call rate lower than

97%; 2995 SNPs for which no homozygous individual was observed for the minor allele on the full genotyped set; and 1689 SNPs that were monomorphic in all lines.

Third, regarding the genotypes for the remaining 38,425 SNPs, we performed a final quality control using PLINK v1.9 software [30] on each line set considered in the current study. We discarded SNPs with a very significant deviation from Hardy–Weinberg equilibrium (p value < 0.0001) in one or several populations. Thus, only 38,350 SNPs remained for the analysis of genetic diversity between populations (PCA and  $F_{ST}$  analysis).

Finally, we only retained the SNPs that had a sufficiently high minor allele frequency (MAF) to obtain reasonable results for each of the analyses. For the ROH studies (MAF ≥ 1%), 34,077–37,340 SNPs were kept depending on the lines. For LD calculation (MAF ≥ 5%), 31,190–34,723 SNPs were considered depending on the lines (see Additional file 1: Table S1).

### Genetic structure of the population

Observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) for a population under Hardy–Weinberg equilibrium were derived with the PLINK v1.9 software [30]. Levels of genetic variation in the different lines were compared using a Kruskal–Wallis non-parametric test on  $H_o$  values. Genetic differentiation between populations was measured with pairwise  $F_{ST}$  estimates [31], using the VCFtools v0.1.13 software [32]. In addition, a principal component analysis (PCA) was performed with the R package Adegenet [33] to visualize the genetic structure of the lines.

### Linkage disequilibrium

To estimate LD, we used the squared correlation based on genotypic allele counts (number of non-reference alleles at each locus) using the PLINK v1.9 software [30]. This  $r^2$  value does not necessitate phasing, it is very similar to but not identical to the  $r^2$  estimate derived from haplotype frequencies [34].

Pairwise LD between adjacent SNPs and pairwise LD between all SNPs in a 60-Mb long window were derived for each chromosome and line. Mean  $r^2$  values were calculated for each chromosome and line by considering the following average distances between SNPs: 10 kb with a 20 kb-window; 50 kb and 100 kb with a 50 kb-window; 1, 3, 5, 10 and 30 Mb within a 100-kb window for each distance.

### Estimates of effective population size $N_e$

Two metrics were considered to estimate  $N_e$ : ( $N_{e_t}$ ) was based on LD and could be derived for all populations, whereas ( $N_{e_T}$ ) was based on temporal changes in allele

frequency and could therefore only be calculated for the SY and SU lines.

For all populations,  $N_e$  was derived using the formula proposed by Sved [12] and based on the expected LD:  $E(r^2) \approx \frac{1}{4cN_e+1}$  in which  $c$  is the distance in Morgan between SNPs. Estimating  $r^2$  by sampling from the populations produces another source of error [13]; consequently, for sample size  $S$ , the formula is  $E(r^2) \approx \frac{1}{1+4cN_e} + \frac{1}{S}$ , which can be rearranged as: 
$$N_e \approx \frac{1 + 1/S - E(r^2)}{4c[E(r^2) - 1/S]}.$$

$N_e$  was estimated by transforming physical distance between SNPs into genetic distance  $c$ , based on the genetic map recently established for a French rainbow trout population: 10 cM corresponded to six Mb on average across the 30 chromosomes [5].

We derived  $N_e$  at  $t$  past generations ( $N_{e_t}$ ) based on the equation  $t = 1/2c$  based on the coalescent theory and the assumptions of Wright–Fisher model [35], considering  $E(r^2)$  as the mean  $r^2$  across all the chromosomes and all SNP pairs across a distance of  $c$ .

To derive mean  $r^2$ , the window around  $c$  was determined for any  $t$  generation by considering the interval  $]t - 0.5; t + 0.5]$ ; for instance at  $t = 1$ ,  $c$  averaged 33.3 Mb between SNPs with possible values ranging from 20 Mb ( $t = 1.5$ ) to 60 Mb ( $t = 0.5$ ). For all populations,  $N_{e_t}$  was calculated up to the 10th generation back. Standard errors for these  $N_e$  estimates were derived according to equations (5) and (6) of Hill [13] considering the chromosome average  $V(r^2)$  over all SNP pairs across a distance of  $c$  per chromosome.

Second,  $N_e$  was derived by considering the temporal approach [11] and the formula proposed by Nei and Tajima [36]:  $N_{e_T} = \frac{t}{2\hat{F}_k - 1/S_0 - 1/S_t}$ , with  $\hat{F}_k$  the standardized variance of allele frequency (corresponding to the Wright inbreeding coefficient),  $S_0$  and  $S_t$  sample sizes at generation 0 and  $t$ , respectively.

$\hat{F}_k$  was derived as  $\hat{F}_k = 2 * (x - y)^2 * \left[ \frac{1}{x+y} + \frac{1}{2-x-y} \right]$  as proposed by Pollak [37] and considering only two alleles per SNP, where  $x$  is the major allele frequency at generation 0 and  $y$  is the allele frequency at generation  $t$ . Then,  $\hat{F}_k$  was averaged for all SNPs to estimate  $N_{e_T}$ .

### Runs of homozygosity

ROH were identified for each fish within all lines using PLINK v1.9 [30]. ROH were defined by sliding windows with a minimum length of one Mb containing at least 30 homozygous SNPs. The maximum gap between two consecutive homozygous SNPs in a run was set to the default value of one Mb. To ensure that the low SNP density did not falsify ROH length, a minimum density of one SNP

every 100 kb was also set (the median and average distances between adjacent SNPs across all the genomes were 31 and 56 kb, respectively). No more than five SNPs with missing genotypes were allowed per window and up to one possible heterozygous genotype was permitted per ROH. These parameters are common practice when deriving ROHs in animal livestock populations with 2–3 Gb genome sizes and 50 K SNP chips. In most studies, the minimum number of SNPs to constitute a ROH is in-between 20 and 50 [38–40]. According to the formula proposed by Purfield et al. [15], the minimum number of SNPs that constituted a ROH should be about 35 SNPs in each line in order to minimize the number of ROH that may occur only by chance in the SNP panel (accepting a 5% false positive rate).

Mean number of ROH ( $N_{ROH}$ ), number of SNP per ROH ( $S_{ROH}$ ), length of ROH ( $L_{ROH}$ ) and percentage of ROH segment longer or equal to 10 Mb were calculated per individual and line. To identify the genomic regions most commonly associated with ROH, the percentage of individuals with a SNP in a ROH segment was calculated by counting the number of times the SNP was detected in a ROH within the population. This count was plotted against the position of the SNP along the chromosome.

### Estimates of total and recent inbreeding

Inbreeding coefficients ( $F_{ROH}$ ) were calculated as the sum of ROH lengths of an individual divided by the total length of the autosomal genome covered by SNPs. The total size of the autosomal genome covered by SNPs was calculated within each line and chromosome. Genome regions with a gap between two adjacent SNPs larger than 1 Mb (authorized gap to derive a ROH) were deducted from the total size of the genome covered by SNPs. In addition, recent inbreeding ( $F_{ROH>10Mb}$ ) was derived as the sum of the lengths of ROH segments longer than 10 Mb in order to estimate inbreeding occurring within the last three generations (10 Mb = 0.166c and  $t = 1/2c$ ; therefore  $t = 3$  generations for 10 Mb). Chromosome  $F_{ROH}$  was also derived as the sum of the lengths of ROH in a given chromosome of an individual divided by the total length of the chromosome genome covered by SNPs and then averaged for each line.

## Results

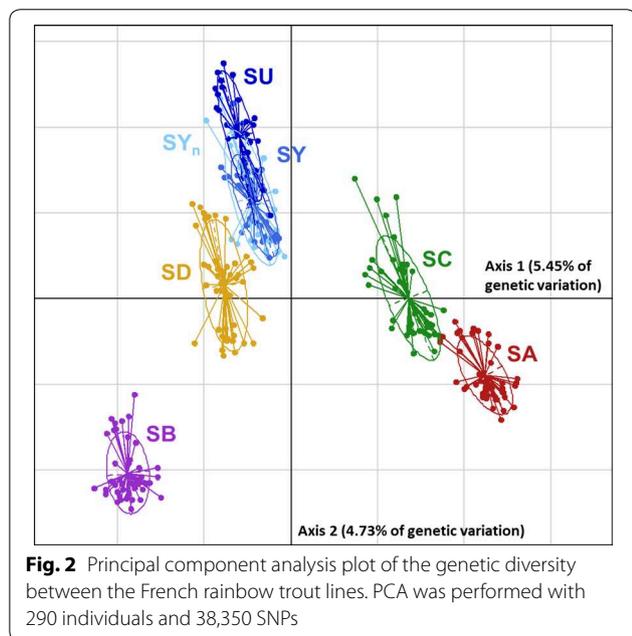
### Population genetic statistics and structure

When considering SNPs irrespective of their MAF, the observed heterozygosity ( $H_o$ ) ranged from 33.5 to 35.3% (Table 1). On average, values for  $H_o$  went up to 36.5 to 38.0% when only SNPs with a MAF  $\geq 5\%$  within the population were considered (Table 1). Frequencies at almost all loci were in agreement with Hardy–Weinberg expectations in each line, although,

**Table 1 Observed (Ho) and expected (He) heterozygosity for each rainbow trout line**

Line	N	All SNPs		SNPs with a MAF ≥ 5%	
		Ho	He	Ho	He
SA	48	33.50*	32.47	36.54*	35.41
SB	48	34.24*	33.34	36.94	35.96
SC	49	33.92*	33.17	36.92	36.10
SD	48	35.01	34.18	37.47	36.58
SU	32	35.33	34.10	37.98	36.65
SY <sub>n</sub>	32	35.30	34.90	37.49	37.06
SY	33	35.23	34.87	37.47	37.05

\* Significant p value (< 1%) for Kruskal–Wallis test for Ho value in comparison to Ho value for SY line



**Fig. 2** Principal component analysis plot of the genetic diversity between the French rainbow trout lines. PCA was performed with 290 individuals and 38,350 SNPs

on average across all the SNPs the expected heterozygosity (He) tended to be one point below that of Ho in all selected populations. Kruskal–Wallis tests showed

that there was a very significant difference in Ho (p value < 0.00001) between the lines when considering all SNPs and still some significant variation (p value < 1%) when considering SNPs with a MAF ≥ 5%. Comparing the Ho values two-by-two showed that the commercial SA, SB and SC lines had significantly lower heterozygosity values than the SD and INRA experimental lines.

Results showed moderate genetic differentiation between lines. Overall, 10% of the total genetic variation is explained by the first two PCA axes (Fig. 2). Except for the null differentiation between the two cohorts of the INRA unselected line (Table 2),  $F_{ST}$  values ranged from 0.02 (SY<sub>n</sub>–SU) to 0.15 (SA–SB). All commercial lines were moderately distant from each other with  $F_{ST}$  ranging from 0.09 to 0.15. The SD line was the commercial line that was genetically the closest to the INRA SY line, whereas the SA and SB were the most distant.

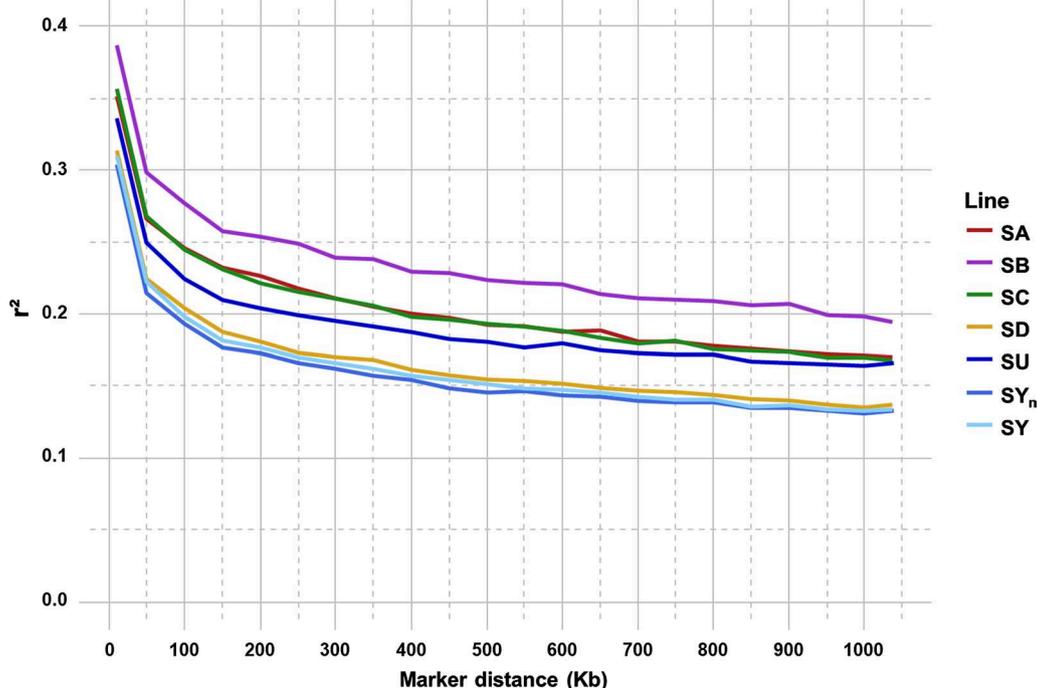
**Linkage disequilibrium analysis**

As expected, average  $r^2$  tended to decrease with increasing distance between pairs of SNPs in all the populations studied, the most rapid decline being over the first 100 kb (Fig. 3). On average, LD decreased from 0.34 (0.30–0.39) to 0.25 (0.22–0.30), 0.23 (0.19–0.28), 0.16 (0.13–0.20), 0.10 (0.09–0.12) and 0.07 (0.07–0.09) for distances between markers of 10 kb, 50 kb, 100 kb, 1 Mb, 5 Mb and 10 Mb, respectively (see Additional file 1: Table S2). The unselected lines SY<sub>n</sub> and SY and the commercial line SD had the lowest LD with an average  $r^2$  equal to 0.22 at 50 kb, whereas the selected lines SU, SC, SA and SB had corresponding  $r^2$  values of 0.25, 0.27, 0.27 and 0.30, respectively (see Additional file 1: Table S2).

Average  $r^2$  for the 50 kb-distant markers varied strongly between chromosomes: from 0.18 for Omy22 to 0.37 for Omy5 with an average value of 0.24 across chromosomes (see Additional file 2: Figure S1). Within chromosome,  $r^2$  varied also between lines: for Omy5, average  $r^2$  for 50 kb-distant markers varied from 0.34 for SY to 0.41 for the SU line; for Omy22,  $r^2$  varied from 0.16 for SD to 0.22 for the SA and SB lines. For some other chromosomes, in

**Table 2 Pairwise  $F_{ST}$  between lines of French rainbow trout**

Line	SA	SB	SC	SD	SU	SY <sub>n</sub>
SB	0.150					
SC	0.115	0.139				
SD	0.113	0.092	0.101			
SU	0.128	0.128	0.113	0.08		
SY <sub>n</sub>	0.101	0.096	0.086	0.057	0.024	
SY	0.105	0.102	0.091	0.065	0.038	0.007



**Fig. 3** Linkage disequilibrium ( $r^2$ ) decay with physical distance between markers in each of the six French rainbow trout lines

particular for the sex chromosome Omy29, variation in average  $r^2$  was more important across lines: from 0.18 for SD to 0.31 for SB.

#### Estimates of effective population size

The  $N_e$  of all lines showed a decreasing trend over the last 10 generations with a steeper slope for the INRA experimental lines and SD line (Fig. 4) than for the four other lines.  $N_e$  stabilized during the last three generations for the SA, SB and SC lines and even started to increase for the last two generations for the SA and SB lines.

In the last generation ( $t = 1$ ),  $N_e$  ranged from only 24 for the SU line to 48 for the SA and SY lines, with intermediate values of 37, 38, 39 and 42 for the SB, SD, SC and  $SY_n$  lines, respectively. For the experimental lines SU and SY, estimates of  $N_e$  in the last generation ( $t = 1$ ) due to variations in allele frequency were higher than estimates based on LD: 66 versus 24 and 216 versus 48 for the SU and SY lines, respectively (results not shown).

#### Analysis of runs of homozygosity

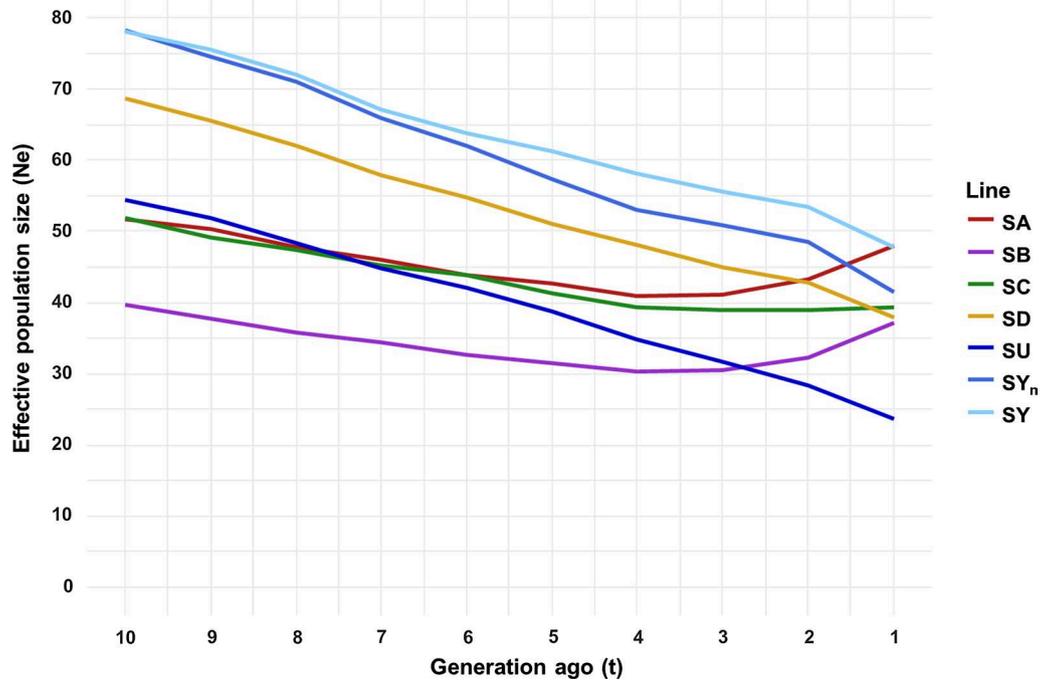
Statistics concerning the average number and size of ROH segments per individual according to the lines are reported in Table S3 (see Additional file 1: Table S3). The average number of ROH per individual varied from 46 for the  $SY_n$  to 68 for the SA line. Individuals with

the smallest number (25) and the largest number (90) of ROH belonged to the  $SY_n$  and SA lines, respectively. The average size of ROH varied from 3.84 Mb for  $SY_n$  to 5.38 Mb for SB. The proportion of long ROH ( $\geq 10$  Mb) varied from 7% for  $SY_n$  to nearly 14% for SB.

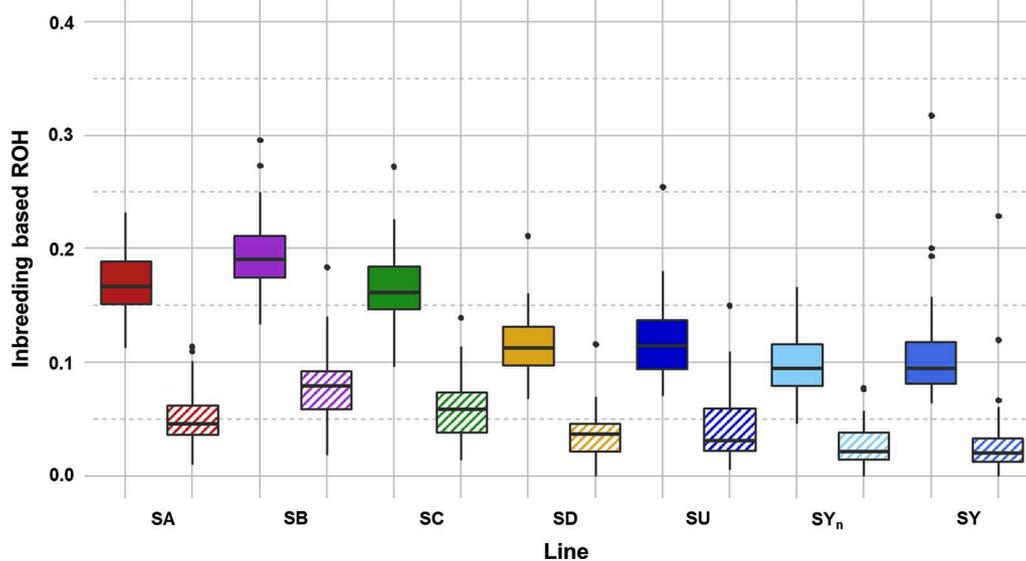
#### Estimates of total and recent inbreeding

Average  $F_{ROH}$  varied from 10.0% for  $SY_n$  to 19.5% for SB [Fig. 5 and Table S3 (see Additional file 1: Table S3)]. Individual  $F_{ROH}$  varied from 4.6 to 31.7% (Fig. 5). Recent inbreeding ( $F_{ROH>10Mb}$ ) ranged from 2.9 for  $SY_n$  to 7.9% for SB [Fig. 5 and Table S3 (see Additional file 1: Table S3)]. Individual  $F_{ROH>10Mb}$  varied from 0 to 22.9% for an individual in SY (Fig. 5). An increase in inbreeding of 1 point (from 10 to 11% between  $SY_n$  and SY) was observed in the unselected SY line in five generations. All selected lines had higher average inbreeding levels than the unselected SY line, the highest levels for total (>16%) and recent (>5%) inbreeding being found for the SA, SB and SC lines that are selected based on pedigree information since seven, four and six generations, respectively.

Regarding  $F_{ROH}$  at the chromosome level (Fig. 6), the mean  $F_{ROH}$  across lines ranged from 11% for Omy13 to 20% for Omy5 with an average of 14% across chromosomes. The highest mean  $F_{ROH}$  (30%) was observed for SA on Omy23 and the lowest value (4%) for  $SY_n$  on



**Fig. 4** LD-based estimates of effective population size ( $N_e$ ) in the French rainbow trout lines over the last ten generations



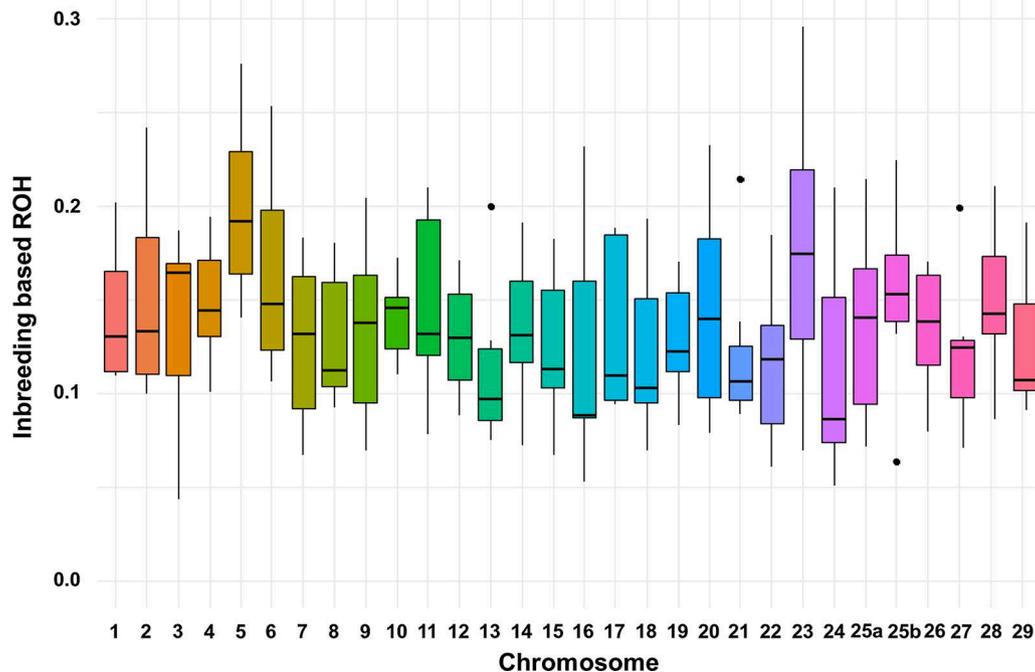
**Fig. 5** Box plots of total inbreeding ( $F_{ROH}$ ) and recent inbreeding ( $F_{ROH>10}$ ) for each rainbow trout line. Plain box: total inbreeding ( $F_{ROH}$ ); hatched box: recent inbreeding ( $F_{ROH>10}$ )

Omy3. Average  $F_{ROH}$  is sometimes relatively close from one line to another (for instance for Omy10, the lowest value is 11% for SD and the highest value is 17% for SB), whereas variation in  $F_{ROH}$  is much larger for Omy23, from 7% for SY to 30% for SA.

## Discussion

### Population differentiation

Most of our rainbow trout populations were moderately differentiated according to the qualitative guidelines



**Fig. 6** Box plots of the line averages of inbreeding coefficients ( $F_{ROH}$ ) derived from ROH per chromosome

proposed by Wright [41] for the interpretation of  $F_{ST}$ : 0–0.05 for little genetic differentiation, 0.05–0.15 for moderate genetic differentiation, 0.15–0.25 for large genetic differentiation, and above 0.25 for very large genetic differentiation, respectively.

Little genetic differentiation ( $F_{ST} < 0.03$ ) was observed between the SU and SY experimental lines, which is reasonable given that the SU selected line was derived from the  $SY_n$  line. The SD commercial line was also genetically close to the SY INRA lines, partly due to the fact that about 50% of the male founders of the SD line came from the INRA sex-reversed males, 25 years ago. Although the other commercial lines are more distant from the INRA line than the SD line (Fig. 2), it is also important to note that all of these four lines were created using sex-reversed males from INRA. Among all the selected lines,  $F_{ST}$  values were quite similar to the observed differentiation levels between Chinese or Western pigs breeds [42], European cattle breeds [39, 43] or sheep breeds [44, 45]. As far as we know, only one  $F_{ST}$  study has been performed for rainbow trout based on a panel of 99 SNPs [46] and found a global  $F_{ST}$  of 0.13 across eight commercial lines and a pairwise  $F_{ST}$  between any two lines ranging from 0.056 to 0.195. Some older works based on microsatellites [47–49] or allozymes [50] also indicated a moderate degree of differentiation among wild population and/or farmed stock rainbow trout lines.

One way to assess the genetic diversity within populations is through observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity. We believe that these parameters should be derived without setting a threshold MAF value in order to keep all SNPs in the analysis and avoid any bias. In this way, we estimated an  $H_o$  of about 34% ( $\pm 1\%$ ); however, most of studies in the literature give results for a subset of SNPs with MAF above 5%. We also calculated  $H_o$  under this MAF threshold and found values close to 37% in all lines, which are very similar to the values estimated in pig lines [51] or cattle breeds [39]. Based on allozyme markers, Cárcamo et al. [50] indicated that commercial lines of rainbow trout showed a similar range of variation in heterozygosity to that of wild populations.

In our study, we observed a slight heterozygosity excess ( $H_o > H_e$ ) with a large set of neutral loci for all lines. This observation may indicate a recent (<100 generations) bottleneck [52] that is likely due to the recent domestication and selection process.

#### Linkage disequilibrium

The LD at short distances between markers ( $\leq 100$  kb) is very similar to  $r^2$  values reported for dairy cattle breeds [53] and slightly higher than values reported for beef cattle breeds or sheep breeds [45]; however, LD estimates for our trout lines at short distances are lower than in most pig breeds [51, 54]. At longer distances between markers (e.g.  $\sim 1$  Mb), LD in French rainbow trout lines

(0.13–0.20) is clearly higher than in ruminant breeds, but very similar to the LD observed in most pig breeds. This high LD, even at long distances, enables accurate genomic predictions for rainbow trout populations even with low density SNP chips [55].

Similar values of  $r^2$  using long distance windows (5 Mb) were observed between the French rainbow trout lines and an American commercial line of rainbow trout [55], with a similar large variation in average  $r^2$  among chromosomes (see Additional file 2: Figure S1), i.e. very high  $r^2$  values on Omy5 and very low values on Omy21 and Omy22. The higher than average LD on Omy5 is likely caused by a large chromosomal double-inversion of 56 Mb [26], which prevents recombination in fish. This double-inversion contains key genes involved in photosensory processes, circadian rhythm, adiposity, and sexual differentiation [26]. This inversion is shown to mediate sex-specific migration through sex-dependent dominance. Quantitative trait loci (QTL) for spawning date and body weight have been detected on Omy5 [56] and may also explain the high LD because of selection of favorable haplotypes for these important traits for breeding. Furthermore, Omy5 is one major exception to the strong association observed between large metacentric chromosomes and high female:male recombination rate ratios [57]. In addition to having an equal female:male recombination rate ratio (average 9:1 ratio for metacentric chromosomes), there is evidence that the short arm of Omy5 is the homeologous linkage group to the Omy29 [58], sex chromosome for which we observed a large variation of  $r^2$  values among lines.

### Effective population size

The estimates of  $N_e$  for the selected lines were consistent with the reports in other aquaculture species such as the Pacific abalone [59], catfish [60], Atlantic salmon [61] for which  $N_e$  was less than 50 after a few generations of mass selection. The  $N_e$  across many livestock breeds is less than 100 [9, 51, 62]. Nevertheless, large and ongoing genetic gains for production traits are typically achieved in livestock, and there are no apparent signs of reaching selection plateaus [63]. Across both domestic and wild populations, the minimal  $N_e$  to avoid inbreeding depression in the short term has been estimated to be at least 50 [64]. For instance, in a long-term (120 generations) selection experiment in mice, several lines were kept at an average  $N_e$  of 60 with no apparent inbreeding problems [65].

In the ideal scenario, population genetics theory recommends keeping equal numbers of males and females and maintaining a constant population size over time. However in most livestock breeding programs, it is often impossible to maintain a 1:1 sex ratio, which greatly

affects the  $N_e$  of a population. In the trout lines under study, the dams-sires ratios are relatively well balanced with values ranging from 6:10 to 10:10. In fish, an additional explanation for the small  $N_e$  may be asymmetric reproduction, e.g. high variance in individual reproductive success or survival rate of broodstock [66]. This last phenomenon has been described as “sweepstakes reproductive success” (SRS [67]), which maintains much less genetic diversity than expected on the basis of a large census size, and would increase the impact of inbreeding depression for aquaculture farms. Christie et al. [68] found in their review of salmon species that early-generation hatchery fish averaged only half the reproductive success of their wild-origin counterparts when spawning in the wild. For *Oncorhynchus mykiss*, the ratio of  $N_e$  to the estimated census population size ( $N$ ) was estimated to range from 0.09 to 0.18 [47], 0.17 to 0.40 [69], and a 0.45 [48], with a large variance in reproductive success among individuals being the key factor to reduce the  $N_e/N$  ratio in salmonid species [68]. Fish breeding from multigenerational hatchery program without pedigree information resulted in a decrease in  $N_e$ , not only by decreasing the mean reproductive success but also by increasing the variance in reproductive success among breeding parents, whereas there was no reduction in  $N_e$  found in fish breeds for a single generation in a local hatchery [69].

Intense directional selection of relatively few animals often results in a skewed genetic contribution and may explain why  $N_e$  values less than 100 are observed for many livestock populations. In our study, the impact of intense selection can be quantified by comparing the evolution of  $N_e$  between the SU and SY lines, but also by taking into account that  $N_e$  is affected by the smaller number of broodstock used for SU than for SY breeding. Based on LD values,  $N_e$  at  $t = 1$  generation was estimated to be equal to 24 and 48 for those two lines, respectively. The discrepancy was even greater when considering estimates of  $N_e$  based on allele frequency variation, 66 for the SU line and 216 for the SY line.

Due to the existence of two major inversions [26] on Omy5 and Omy20 that may influence our LD-based analysis (and to a lower extent the ROH), we removed these chromosomes from the derivation of  $N_e$  and  $F_{ROH}$ . Values for total  $F_{ROH}$  and recent inbreeding were almost unchanged, but  $N_e$  estimates were very sensitive to the absence of Omy5 and Omy20 although  $r^2$  values were averaged per chromosome before deriving  $N_e$ . Removing these chromosomes increases our estimates of  $N_e$  across all lines, on average by 40% at generation  $t = 1$  and by 14% at generation  $t = 10$  (see Additional file 1: Table S4). However, LD-based estimates of  $N_e$  remained small for all the lines compared to census sizes that were

close to 200 for all commercial lines, 140 for the SY line and 80 for the SU line. Regardless of the line considered in the study (selected or not), we observed that in general  $N_e$  tended to decrease in the last 10 generations, which may be explained by drift and SRS phenomena. It should also be noted that in the three selected lines (SA, SB and SC) for which a mating optimization procedure to limit inbreeding was set up in the last 10 years by SYSAAF,  $N_e$  stabilized (SC) or even slightly increased (SA and SB) in the last three generations. This suggests that some mating optimization protocol (genomic-based pedigree) should be similarly introduced to limit the inbreeding increases in the SD, SU and SY lines.

In addition to quantifying the important decreases in  $N_e$  due to selection, our results underscored the difficulty in acquiring accurate estimates for this parameter, since estimates based on the same molecular information can be doubled or tripled depending on the evolutionary model considered [70]. It is well-established that better estimates of inbreeding levels and  $N_e$  are obtained using molecular rather than genealogical information [16], at least when considering more than 10K SNPs [71]. In most cases, a full-depth pedigree is unavailable and most base populations include selected or partially inbred founders in the pedigree files. These unknown relationships between animals may lead to strong underestimation of the pedigree-based estimates of inbreeding [16]. In addition, by considering complete pedigree back to a base population in simulation studies, Liu et al. [72] and Fortun et al. [16] reported that pedigree-based estimates of inbreeding were lower than true inbreeding in selected populations. The pedigree-based inbreeding assumes neutral loci, i.e. that the two alleles at the same locus on two homologous chromosomes have an equal chance of being selected. In reality, for some loci, the two alleles may have different effects on a naturally or artificially selected trait, which leads to unequal selection probabilities between the two alleles.

Most methods applied to infer  $N_e$  from genomic population data rely on the Wright–Fisher model's assumptions of low fecundity non-skewed offspring distributions. Although proven to be robust to violations of most of these assumptions, these methods drastically failed to approximate the genealogies of species with high SRS, whereby few individuals contribute most of the offspring to the next generation [73]. As stated by Montano [70], the development of statistical tools based on models that consider SRS will substantially improve estimates of population demographic parameters.

### Runs of homozygosity

The majority of metrics to estimate  $N_e$  assumes that the value remains constant across the genome. However,  $N_e$

varies across the genome, such that some regions have an increased loss of diversity compared with others [74, 75]. The  $N_e$  is expected to vary across the genome as a consequence of genetic hitchhiking due to selection [76] and negative selection acting on deleterious mutations (i.e., background selection [77]). The action of selection, particularly in regions of the genome with low rates of recombination, is expected to reduce  $N_e$ , leading to lower levels of genetic diversity and reduced effectiveness of selection. We could have derived  $N_e$  per chromosome based on chromosome-specific LD estimates; however, we preferred to study the heterogeneity of genetic diversity throughout the genome using ROH because  $N_e$  estimates have been proven unreliable for fish populations (see previous section), and quantifying the absolute level of inbreeding (in total and for each chromosome) in our rainbow trout lines was per se an objective of the study. ROH can be used to directly estimate inbreeding depression [78] and can identify the chromosome regions that are responsible for inbreeding depression along the genome [79].

While for selected rainbow trout lines,  $F_{\text{ROH}}$  varied between 12 and nearly 20%, estimates were a bit lower (~11%) for the INRA unselected line. These results are consistent with our  $N_e$  estimates for three to ten generations ago: the SY and SD lines had the largest  $N_e$  and the lowest  $F_{\text{ROH}}$  whereas SB had the smallest  $N_e$  and the highest  $F_{\text{ROH}}$  (Pearson correlation  $F_{\text{ROH}}$  and  $N_{e_{t=10}}$ :  $-0.90$ ,  $p$  value  $< 0.005$ ). It is difficult to compare our results to other broodstock programs because the few published estimations of levels of inbreeding for fish commercial populations are all based on pedigree information, and thus are likely to be underestimated due to unknown relationships and inbreeding in the base populations. In rainbow trout, Pante et al. [80] estimated inbreeding levels that varied between 3 and 10% at the 6th generation of selection for three commercial populations. A hierarchical mating system was used (one male mated to two to three females) with a very variable number of families from one generation to another; full-sib and half-sib matings were avoided and matings that would yield inbreeding coefficients of 12.5% or more were restricted. In Coho salmon, Myers et al. [81] reported inbreeding levels of about 15–16% in two lines after the 9th and 10 generations of the Domsea selection program based on a circular mating design with 60 families produced at each generation [82]. More recently, in two different lines of Coho salmon, Yáñez et al. [83] estimated inbreeding levels at 5 and 7% after 6 and 7 generations of selection, respectively, under a hierarchical design (one male mated to three to five females) with 100 families produced at each generation and inbreeding controlled by avoiding half- and full-sib matings.

While empirical information on genome-wide diversity in model species and livestock has been collected, we still lack a clear picture for farmed fish of the heterogeneity of genetic diversity across the genome. As far as we know, our study is the first to calculate ROH and estimate  $F_{ROH}$  in any fish species.  $F_{ROH}$  estimates from this study were higher than estimates for terrestrial livestock regardless of their breeding management (in breeds for ruminants or in lines for pigs). Estimates varied from 3 to 9% in dairy cattle breeds [15, 38, 39], from 2 to 11% in sheep breeds [84, 85] and from 3 to 11% in pig lines [51, 86, 87].

Although the SY line is not under selective breeding, inbreeding is relatively high (11%) and has increased by 1% in six generations (see Additional file 1: Table S3). This is probably partly due to a founder effect as well as to SRS, as previously discussed for  $N_e$ . In spite of the existence of mating plans for selected lines with known pedigree (SA, SB and SC), recent inbreeding is relatively high in these lines (5–8%). Nevertheless, we may be able to manage their genetic diversity along the entire genome. Because no ROH fragments are fixed within the populations (see Additional file 2: Figure S2), we can develop selection and mating strategies to specifically limit inbreeding in genome regions for which an important proportion of individuals (>50% for instance) share the same ROH.

It is worthwhile to underscore that  $F_{ROH}$  estimates may vary depending on the parameters considered to define ROH segments. These parameters have to be tuned according to marker density and pan-genomic heterogeneity and the level of recombination along the genome. There is no clear consensus in the literature on how to choose these parameters and most studies used Plink default values. In a recent study [16], a gene-dropping simulation was performed and inbreeding estimates based on ROH and pedigree data were compared to true inbreeding. Inbreeding based on ROH was estimated using different software and different threshold parameters using 50 K chip data. While pedigree inbreeding underestimated true inbreeding, using ROH with a minimum window size of 20 to 50 SNPs provided the closest estimates to true inbreeding regardless of the software [16, 88]. In our study, we observed that inbreeding estimates were similar for window sizes of 30 and 50 SNPs, and estimates were almost identical when considering recent inbreeding (results not shown). Nevertheless the results were sensitive to the maximum gap between two successive SNPs (1000 kb vs. 250 kb) and, to a lesser extent, to the MAF threshold (see Additional file 1: Table S3). Due to the heterogeneity of the marker density along the genome, long ROH were fragmented into smaller ones when considering a maximum gap of 250 kb, leading to underestimated  $F_{ROH}$  in our lines. The

largest differences were for recent inbreeding estimates,  $F_{ROH \geq 10^7}$  that were less than 1% for the gap threshold of 250 kb, but varied between 3 and almost 8% for the 1000 kb threshold (see Additional file 1: Table S3).

As shown in previous studies on the sheep genome [45, 85, 89], we observed a strong variation in  $F_{ROH}$  values among chromosomes, these results being consistent with variation in LD among chromosomes. Chromosome ends are less represented in ROH segments (results not shown), probably due to the very low marker density in these areas. Hotspot areas for ROH were observed for each chromosome (see Additional file 2: Figure S2). In general, these areas and the proportions of individuals within a population that shared ROH hotspots in those areas were highly variable from one line to another. Nevertheless a few hotspot areas were commonly detected across lines, such as the position at ~48 Mb of Omy10 (see Additional file 2: Figure S3). At this specific position, is located a QTL for bacterial cold disease resistance that was detected in an American rainbow trout population [90] and in a French commercial line [91]. When hotspot areas are common to different lines, they may be due to local low recombination rates [85] or to key genes for domestication or production [39, 40, 85]. These hotspot areas could then help identify important genes for domestication or production.

The accumulation of inbreeding is heterogeneous across the genome, such that certain regions are being inbred at a faster rate than other regions of the genome. Currently, SNP information is predominantly used to predict breeding values for genomic selection in livestock populations, now including farmed fish (catfish, tilapia and salmonids). However, this information could also be used to manage the rate at which inbreeding builds up in the genome of livestock populations. Characterizing and efficiently managing inbreeding levels is a major goal to ensure that populations can adapt to future breeding goals while maintaining genetic diversity and avoiding the accumulation of detrimental effects associated with inbreeding. Identifying the main regions of the genome that contribute to inbreeding depression may help to manage inbreeding not only at the individual level but directly at the genome level.

## Conclusions

Our findings provide the breeding companies a better understanding of the genetic diversity in their rainbow trout lines in order to implement efficient breeding programs. The availability of a genome-wide SNP chip allowed us to characterize genetic diversity between and within lines. Lines are moderately differentiated across commercially developed trout lines, with similar  $F_{ST}$  values as those reported across ruminant breeds or pig lines.

Within each line, effective population size seems rather small and inbreeding levels are higher than in terrestrial livestock selected populations. This may be explained by founder effects, sweepstakes reproductive success, and intense selection in some lines. The impact of these significant levels of inbreeding on rainbow trout performance should be quantified in order to assess potential inbreeding depression phenomena and risks for future genetic gains. The levels of molecular inbreeding derived from the identification of homozygous genomic segments could also be used to directly identify the regions that are responsible for inbreeding depression along the genome. We could then expect a more efficient purging allowing for higher relatedness between selected individuals without inbreeding drawbacks in breeding programs.

## Additional files

**Additional file 1. Table S1:** Number of SNPs per minimum allele frequency (MAF) category in each line. **Table S2:** Average  $r^2 \pm$  SD between SNPs according to different distances. **Table S3:** Sensitivity analysis of ROH estimates to MAF and maximal distance gap between two SNPs and derived inbreeding coefficient. **Table S4:** Estimates of effective population size (standard errors in brackets) for each line with or without Omy5 and Omy20.

**Additional file 2. Figure S1:** Line mean linkage disequilibrium at 50 kb for each chromosome. **Figure S2:** Proportion of individuals per line with a SNP in a ROH along the genome. **Figure S3:** Proportion of individuals per line with a SNP in a ROH along chromosome 10.

## Acknowledgements

We thank the staff of the INRA experimental facilities (PEIMA, Sizon) and the French breeding companies who are members of SYSAAF (Aqualande, Bretagne Truite, Les Fils de Charles Murgat and Viviers de Sarrance) for their participation in the fish rearing and for providing biological samples for genotyping individuals. We are very grateful to D. Laloë and G. Restoux for valuable advice on principal component analysis and  $N_e$  estimation, respectively.

## Authors' contributions

JDA managed the validation of genotypes, wrote some scripts, estimated and interpreted the population genetic parameters and wrote the manuscript. FP wrote some scripts, supervised the analysis and the interpretation of the results and wrote the manuscript. PH and AB managed the project, supervised data collection and contributed to interpreting the results. SB proposed some scripts. CP performed the genotyping of all animals. EQ contributed to the design of the study. ND wrote some scripts. CF and MC blasted the markers on the rainbow trout genome and edited duplicate markers. MDN initiated the project, managed the data collection and preparation for genotyping, and contributed to the design of the study and interpretation of the results. All authors read and approved the final manuscript.

## Funding

This study was supported by the European Maritime and Fisheries Fund and FranceAgrimer as part of "57K-Truite" Project (No. 2015-0638) and "SG-Truite" Project (RFEA47 0016 FA 1000016, No. 2017-0239).

## Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because they belong to commercial breeding companies but are available from the corresponding author on reasonable request and with permission of the relevant companies.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. <sup>2</sup> SYSAAF Section Aquacole, Campus de Beaulieu, 35000 Rennes, France. <sup>3</sup> SYSAAF Section Avicole, Centre INRA Val de Loire, 37380 Nouzilly, France. <sup>4</sup> GDEC, INRA, Université Clermont-Auvergne, 63039 Clermont-Ferrand, France.

Received: 1 October 2018 Accepted: 22 May 2019

Published online: 06 June 2019

## References

- Cowx IG. Cultured aquatic species information programme—*Oncorhynchus mykiss*. Rome: FAO Fisheries and Aquaculture Department. 2005. [http://www.fao.org/fishery/culturedspecies/Oncorhynchus\\_mykiss/en](http://www.fao.org/fishery/culturedspecies/Oncorhynchus_mykiss/en). Accessed 2 May 2019.
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun*. 2014;5:3657.
- Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC Genet*. 2012;13:15.
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in Rainbow trout (*Oncorhynchus mykiss*). *Front Genet*. 2016;7:203.
- Fraslin C, Dechamp N, Bernard M, Krieg F, Hervet C, Guyomard R, et al. Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. *Genet Sel Evol*. 2018;50:60.
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K SNP array for rainbow trout. *Mol Ecol Resour*. 2015;15:662–72.
- Felsenstein J. The effect of linkage on directional selection. *Genetics*. 1965;52:349–63.
- Wright S. Evolution in mendelian populations. *Genetics*. 1931;16:97–159.
- Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E, Danchin-Burge C. Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genet Sel Evol*. 2013;45:1.
- Waples RS. Tiny estimates of the  $N_e/N$  ratio in marine fishes: Are they real? *J Fish Biol*. 2016;89:2479–504.
- Waples RS. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*. 1989;121:379–91.
- Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*. 1971;2:125–41.
- Hill WG. Estimation of effective population size from data on linkage disequilibrium. *Genet Res (Camb)*. 1981;38:209–16.
- McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet*. 2008;83:359–72.
- Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet*. 2012;13:70.
- Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics*. 2018;19:98.
- Le Boucher R, Dupont-Nivet M, Vandeputte M, Kerneis T, Goardon L, Labbé L, et al. Selection for adaptation to dietary shifts: towards sustainable breeding of carnivorous fish. *PLoS One*. 2012;7:e44989.

18. Callet T, Médale F, Larroquet L, Surget A, Aguirre P, Kerneis T, et al. Successful selection of rainbow trout (*Oncorhynchus mykiss*) on their ability to grow with a diet completely devoid of fishmeal and fish oil, and correlated changes in nutritional traits. *PLoS One*. 2017;12:e0186705.
19. Chevassus B, Quillet E, Krieg F, Hollebecq M-G, Mambrini M, Fauré A, et al. Enhanced individual selection for selecting fast growing fish: the 'PROSPER' method, with application on brown trout (*Salmo trutta fario*). *Genet Sel Evol*. 2004;36:643–61.
20. Dupont-Nivet M, Vandeputte M, Haffray P, Chevassus B. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture*. 2006;252:161–70.
21. Haffray P, Bugeon J, Rivard Q, Quittet B, Puyo S, Allamelou JM, et al. Genetic parameters of in vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 2013;410–411:236–44.
22. Haffray P, Enez F, Bugeon J, Chapuis H, Dupont-Nivet M, Chatain B, et al. Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. *Aquaculture*. 2018;490:350–4.
23. Haffray P, Pincet C, Rault P, Coudurier B. Domestication et amélioration génétique des cheptels piscicoles français dans le cadre du SYSAAF. *Prod Anim*. 2004;17:243–52.
24. Chapuis H, Pincet C, Colleau JJ. Optimizing selection with several constraints in poultry breeding. *J Anim Breed Genet*. 2016;133:3–12.
25. Omyk\_1.0. [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_002163495.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1/).
26. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadia-Cardoso A, et al. Sex-dependent dominance maintains migration supergene in rainbow trout. *bioRxiv*. 2018. <https://doi.org/10.1101/504621>.
27. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc Biol Sci*. 2014;281:20132881.
28. Thorgaard GH. Chromosomal differences among rainbow trout populations. *Copeia*. 1983;1983:650–62.
29. Affymetrix. Axiom™ Analysis Suite 2.0: UserGuide. 2016. [https://media.affymetrix.com/support/downloads/manuals/axiom\\_analysis\\_suite\\_user\\_guide.pdf](https://media.affymetrix.com/support/downloads/manuals/axiom_analysis_suite_user_guide.pdf). Accessed 2 May 2019.
30. Chang CC, Chow CC, Tellier L CAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
31. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* (NY). 1984;38:1358–70.
32. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
33. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011;27:3070–1.
34. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968;38:226–31.
35. Hayes BJ, Visscher PM, Mcpartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*. 2003;13:635–43.
36. Nei M, Tajima F. Genetic drift and estimation of effective population size. *Genetics*. 1981;98:625–40.
37. Pollak E. A new method for estimating the effective population size from allele frequency changes. *Genetics*. 1983;104:531–48.
38. Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *PLoS One*. 2015;10:e0124157.
39. Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, et al. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genet Sel Evol*. 2017;49:83.
40. Mastrangelo S, Ciani E, Sardina MT, Sottile G, Pilla F, Portolano B. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. *Anim Genet*. 2018;49:71–81.
41. Wright S. *Evolution and the genetics of populations, volume 4. Variability within and among natural populations*. Chicago: The University of Chicago Press; 1984.
42. Ai H, Huang L, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One*. 2013;8:e56001.
43. Gautier M, Laloë D, Moazami-Goudarzi K. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One*. 2010;5:e13038.
44. Mastrangelo S, Di Gerlando R, Tolone M, Tortorici L, Sardina MT, Portolano B. Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. *BMC Genet*. 2014;15:108.
45. Al-Mamun HA, Clark SA, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genet Sel Evol*. 2015;47:90.
46. Liu S, Palti Y, Martin KE, Parsons JE, Rexroad CE. Assessment of genetic differentiation and genetic assignment of commercial rainbow trout strains using a SNP panel. *Aquaculture*. 2017;468:120–5.
47. Heath DD, Busch C, Kelly J, Atagi DY. Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Mol Ecol*. 2002;11:197–214.
48. Rexroad CE III, Vallejo RL. Estimates of linkage disequilibrium and effective population size in rainbow trout. *BMC Genet*. 2009;10:83.
49. Gross R, Lulla P, Paaver T. Genetic variability and differentiation of rainbow trout (*Oncorhynchus mykiss*) strains in northern and Eastern Europe. *Aquaculture*. 2007;272:S139–46.
50. Carcamo CB, Diaz NF, Winkler FM. Genetic diversity in Chilean populations of rainbow trout, *Oncorhynchus mykiss*. *Lat Am J Aquat Res*. 2015;43:59–70.
51. Grossi DA, Jafarikia M, Brito LF, Buzanskas ME, Sargolzaei M, Schenkel FS. Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs. *BMC Genet*. 2017;18:6.
52. Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*. 1996;144:2001–14.
53. Hozé C, Fouilloux MN, Venot E, Guillaume F, Dassonneville R, Fritz S, et al. High-density marker imputation accuracy in sixteen French cattle breeds. *Genet Sel Evol*. 2013;45:33.
54. Wang L, Sørensen P, Janss L, Ostensen T, Edwards D. Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. *BMC Genet*. 2013;14:115.
55. Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Sixin L, Parsons JE, et al. 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*. 2018;135:263–74.
56. O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM. Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *J Hered*. 2003;94:273–84.
57. Danzmann RG, Cairney M, Davidson WS, Ferguson MM, Gharbi K, Guyonard R, et al. A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome*. 2005;48:1037–51.
58. Phillips RB, Nichols KM, DeKoning JJ, Morasch MR, Keatley KA, Rexroad C, et al. Assignment of rainbow trout linkage groups to specific chromosomes. *Genetics*. 2006;174:1661–70.
59. Chen N, Luo X, Lu C, Ke C, You W. Effects of artificial selection practices on loss of genetic diversity in the Pacific abalone, *Haliotis discus hannai*. *Aquac Res*. 2017;48:4923–33.
60. Garcia ALS, Bosworth B, Waldbieser G, Misztal I, Tsuruta S, Lourenco DAL. Development of genomic predictions for harvest and carcass weight in channel catfish. *Genet Sel Evol*. 2018;50:66.
61. Barria A, López ME, Yoshida G, Carvalheiro R, Lhorente JP, Yáñez JM. Population genomic structure and genome-wide linkage disequilibrium in farmed Atlantic salmon (*Salmo salar* L.) using dense SNP genotypes. *Front Genet*. 2018;9:649.
62. Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT, Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*. 2016;10:746–54.
63. Hill WG, Kirkpatrick M. What animal breeding has taught us about evolution. *Annu Rev Ecol Evol Syst*. 2010;41:1–19.
64. FAO. Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk. 1998. <http://www.fao.org/3/a-w9361e.pdf>. Accessed 2 May 2019.

65. Holt M, Meuwissen T, Vangen O. Long-term responses, changes in genetic variances and inbreeding depression from 122 generations of selection on increased litter size in mice. *J Anim Breed Genet*. 2005;122:199–209.
66. Lallias D, Boudry P, Lapègue S, King JW, Beaumont AR. Strategies for the retention of high genetic variability in European flat oyster (*Ostrea edulis*) restoration programmes. *Conserv Genet*. 2010;11:1899–910.
67. Hedgecock D, Pudovkin AI. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Mar Sci*. 2011;87:971–1002.
68. Christie MR, Ford MJ, Blouin MS. On the reproductive success of early-generation hatchery fish in the wild. *Evol Appl*. 2014;7:883–96.
69. Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. *Mol Ecol*. 2007;16:953–66.
70. Montano V. Coalescent inferences in conservation genetics: Should the exception become the rule? *Biol Lett*. 2016. <https://doi.org/10.1098/rsbl.2016.0211>.
71. Wang J. Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol*. 2016;107:4–13.
72. Liu H, Sørensen AC, Meuwissen THE, Berg P. Allele frequency changes due to hitch-hiking in genomic selection programs. *Genet Sel Evol*. 2014;46:8.
73. Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*. 2006;172:2621–33.
74. Gossmann TI, Woolfit M, Eyre-Walker A. Quantifying the variation in the effective population size within a genome. *Genetics*. 2011;189:1389–402.
75. Jiménez-Mena B, Hospital F, Bataillon T. Heterogeneity in effective population size and its implications in conservation genetics and animal breeding. *Conserv Genet Resour*. 2016;8:35–41.
76. Smith JM, Haigh J. Hitch-hiking effect of a favorable gene. *Genet Res (Camb)*. 1974;23:23–35.
77. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993;134:1289–303.
78. Silló L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J Anim Breed Genet*. 2013;130:349–60.
79. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet Sel Evol*. 2014;46:71.
80. Pante MJR, Gjerde B, McMillan I. Inbreeding levels in selected populations of rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*. 2001;192:213–24.
81. Myers JM, Heggelund PO, Hudson G, Iwamoto RN. Genetics and broodstock management of coho salmon. *Aquaculture*. 2001;197:43–62.
82. Iwamoto RN, Towner R, Myers J, Hudson G, Munsell P. Coho salmon broodstock development: a case study of the Domsea Coho salmon (1977 to 2015). *Bull Jpn Fish Res Educ Agency*. 2017;45:107–12.
83. Yáñez JM, Bassini LN, Filp M, Lhorente JP, Neira R, Ponzoni RW, Neira R. Inbreeding and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding nucleus in Chile. *Aquaculture*. 2013;420–421:515–9.
84. Mastrangelo S, Portolano B, Di Gerlando R, Ciampolini R, Tolone M, Sardina MT. Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. *Animal*. 2017;11:1107–16.
85. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One*. 2017;12:e0176780.
86. Zanella R, Peixoto JO, Cardoso FF, Cardoso LL, Biegelmeyer P, Cantão ME, et al. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genet Sel Evol*. 2016;48:24.
87. Yang B, Cui L, Perez-Enciso M, Traspov A, Crooijmans RPMA, Zinovieva N, et al. Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet Sel Evol*. 2017;49:71.
88. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci USA*. 2007;104:19942–7.
89. Chitneedi PK, Arranz JJ, Suarez-Vega A, García-Gómez E, Gutiérrez-Gil B. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. *Anim Genet*. 2017;48:436–46.
90. Vallejo RL, Liu S, Gao G, Fragomeni BO, Hernandez AG, Leeds TD, et al. Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. *Front Genet*. 2017;8:156.
91. Frasin C, Brard-Fudulea S, D'Ambrosio J, Bestin A, Charles M, Haffray P, et al. 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*. 2019. <https://doi.org/10.1111/age.12777>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



## Additional File 1 pour l'article dans GSE

Figure S1. Line mean linkage disequilibrium at 50 Kb for each chromosome.

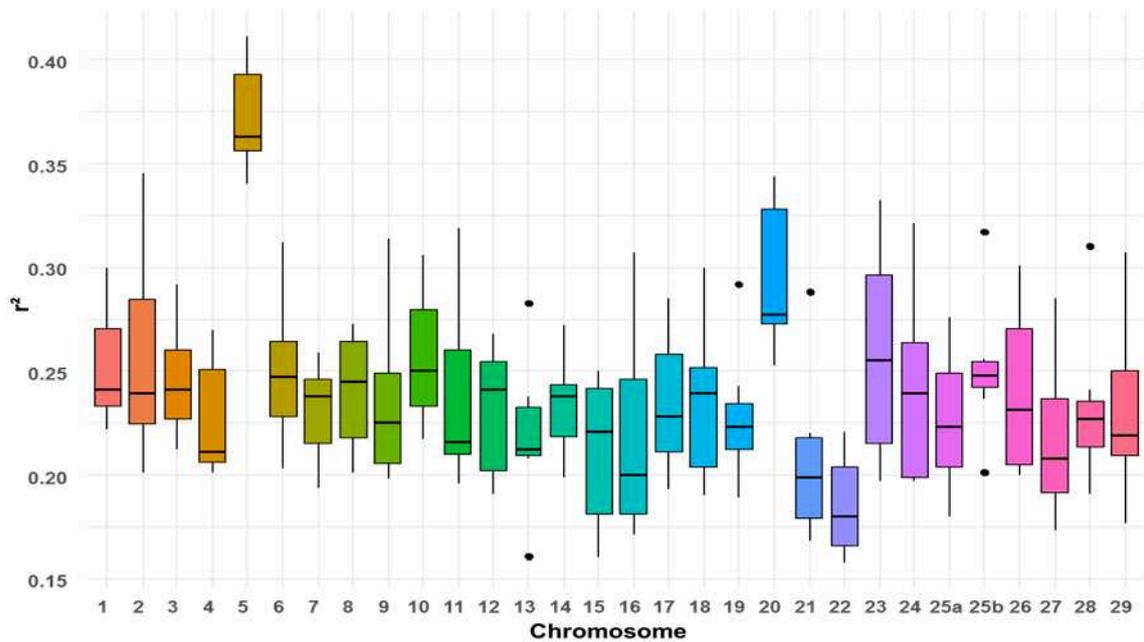
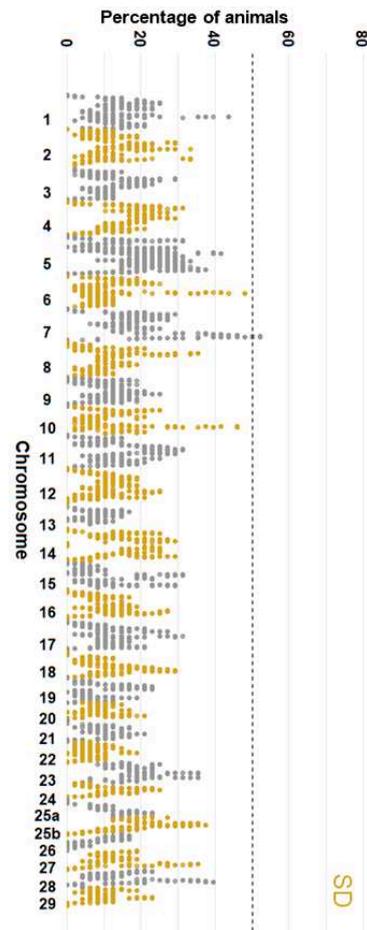
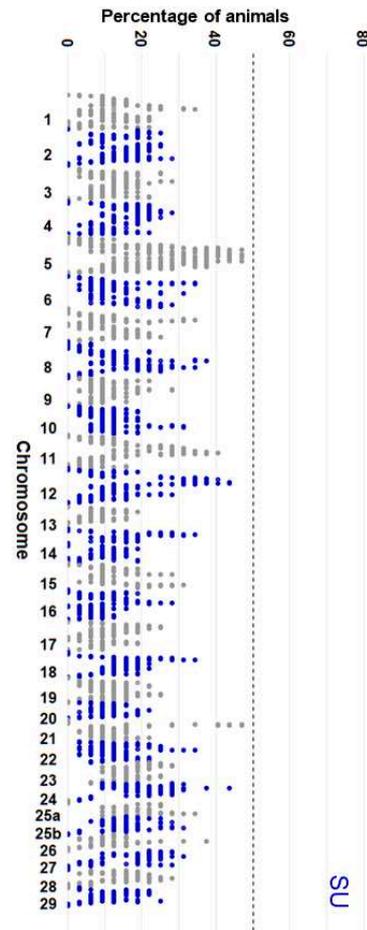
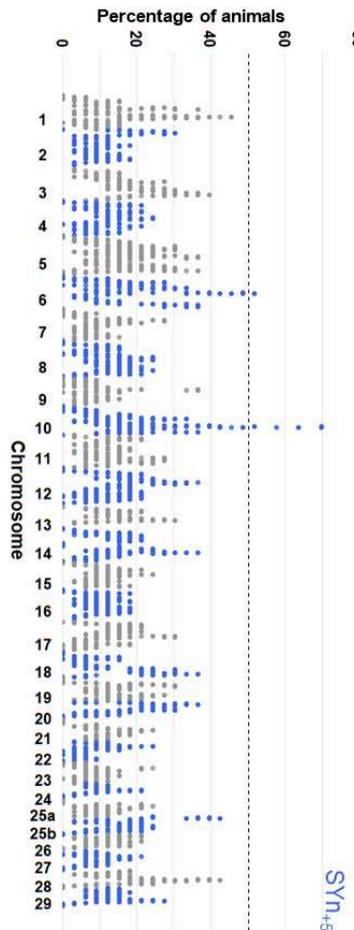
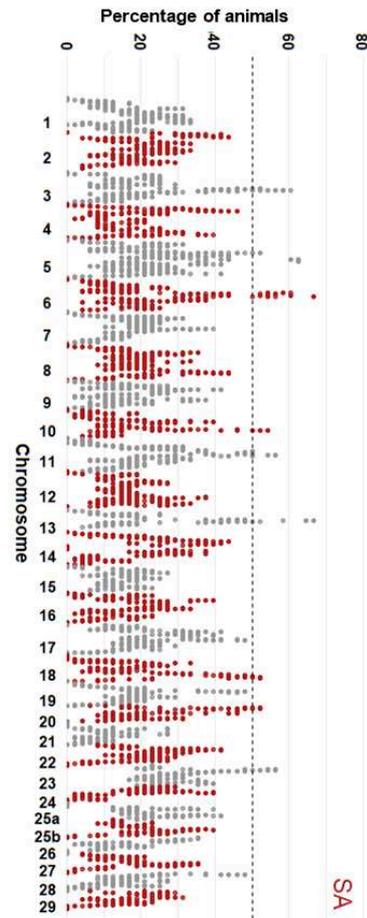
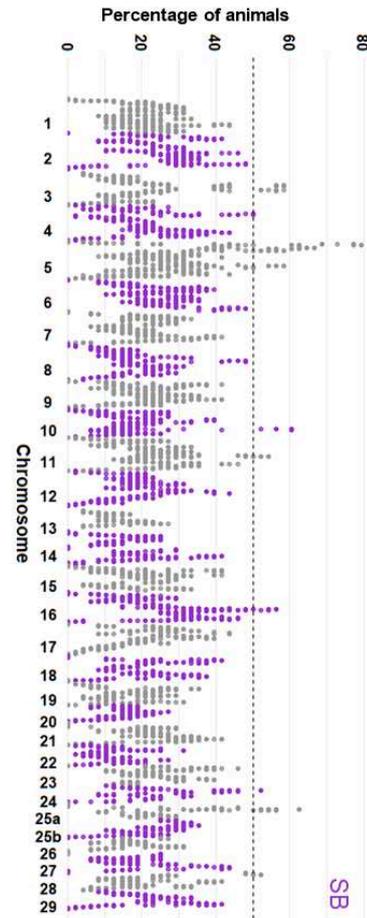
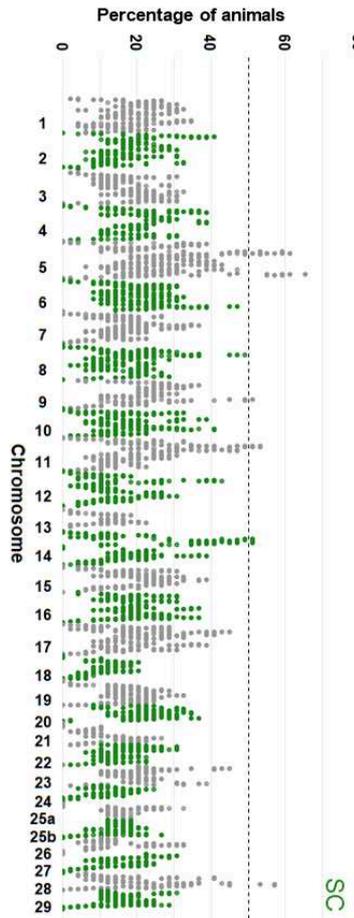


Figure S2. Percentage of individuals within line with a SNP in a ROH along the genome.



**Table 1S.** Number of SNP per minimum allele frequency (MAF) category in each line.

Line	MAF < 1%	1% ≤MAF <5%	MAF ≥5%	MAF ≥ 25%
SA	2,884	3,432	32,034	16,263
SB	4,273	2,887	31,190	16,716
SC	3,162	3,286	31,902	16,915
SD	1,363	2,810	34,175	18,726
SU	2,132	3,025	33,201	17,865
SY <sub>n</sub>	1,010	2,622	34,723	19,173
SY <sub>n+5</sub>	1,389	2,621	34,341	19,173

**Table 2S.** Average  $r^2 \pm$  SD between SNPs according to different distances.

Distance (Kb)	SA	SB	SC	SD	SU	SY <sub>n</sub>	SY <sub>n+5</sub>
10	0.35 ± 0.34	0.39 ± 0.36	0.36 ± 0.35	0.31 ± 0.32	0.34 ± 0.33	0.30 ± 0.32	0.31 ± 0.32
50	0.27 ± 0.29	0.30 ± 0.31	0.27 ± 0.29	0.22 ± 0.26	0.25 ± 0.27	0.22 ± 0.25	0.22 ± 0.25
100	0.25 ± 0.27	0.28 ± 0.30	0.25 ± 0.27	0.20 ± 0.24	0.22 ± 0.25	0.19 ± 0.23	0.20 ± 0.23
1000	0.17 ± 0.21	0.20 ± 0.24	0.17 ± 0.21	0.14 ± 0.18	0.16 ± 0.21	0.13 ± 0.18	0.13 ± 0.18
3000	0.12 ± 0.17	0.14 ± 0.19	0.12 ± 0.17	0.10 ± 0.15	0.13 ± 0.18	0.10 ± 0.15	0.10 ± 0.15
5000	0.10 ± 0.16	0.12 ± 0.18	0.10 ± 0.16	0.09 ± 0.14	0.11 ± 0.17	0.09 ± 0.15	0.09 ± 0.14
10000	0.07 ± 0.13	0.08 ± 0.14	0.07 ± 0.14	0.07 ± 0.13	0.09 ± 0.15	0.07 ± 0.13	0.07 ± 0.12
30000	0.04 ± 0.11	0.05 ± 0.11	0.05 ± 0.12	0.05 ± 0.10	0.07 ± 0.13	0.06 ± 0.11	0.05 ± 0.10

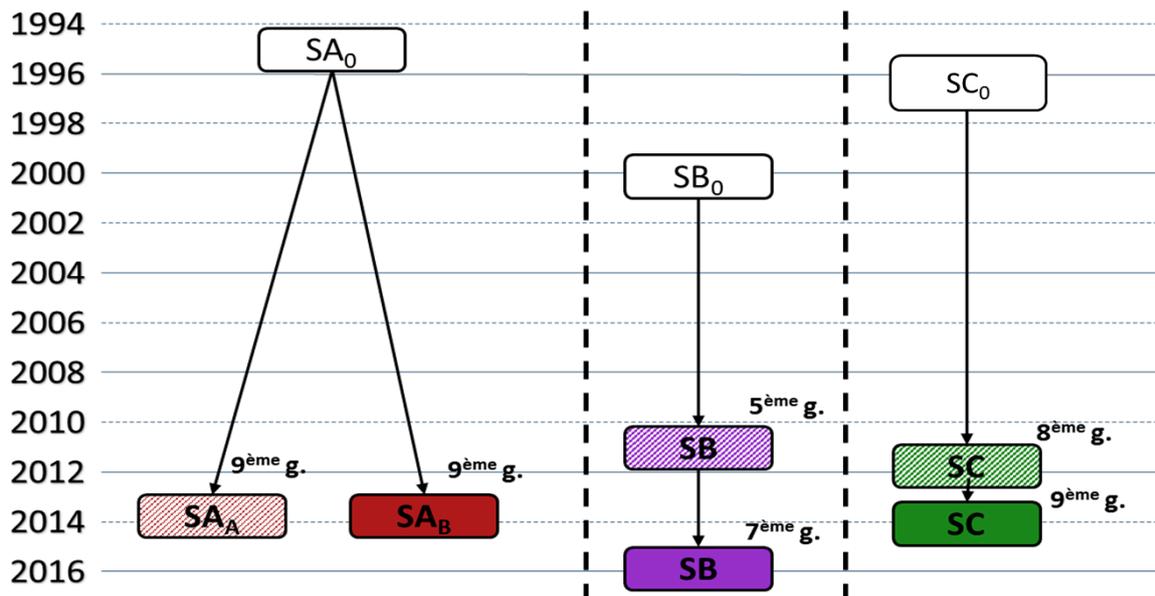
**Table 3S.** Sensitivity analysis of ROH estimates to MAF and maximal distance gap between two SNPs and derived inbreeding coefficient.

ROH with (MAF1%, max gap 1000kb)						
Line	Number of segments	Number of SNPs per segment	Size of segments (Mb)	% ROH>10Mb	F <sub>ROH</sub>	F <sub>ROH&gt;10Mb</sub>
SA	68.2 (± 9.4)	96 (± 91)	4.51 (± 4.19)	8.86	0.171 (± 0.027)	0.051 (± 0.024)
SB	65.2 (± 7.3)	110 (± 107)	5.38 (± 5.21)	13.57	0.195 (± 0.034)	0.079 (± 0.032)
SC	65 (± 9.7)	98 (± 100)	4.60 (± 4.71)	10.08	0.166 (± 0.034)	0.058 (± 0.027)
SD	50.2 (± 9)	93 (± 99)	4.14 (± 4.42)	7.67	0.116 (± 0.027)	0.036 (± 0.020)
SU	49.9 (± 9.6)	97 (± 114)	4.40 (± 5.24)	9.33	0.122 (± 0.040)	0.045 (± 0.035)
SYn	46.2 (± 10.5)	87 (± 87)	3.84 (± 3.97)	6.96	0.100 (± 0.030)	0.029 (± 0.019)
SYn+5	49.7 (± 9.5)	89 (± 98)	4.01 (± 4.52)	7.32	0.112 (± 0.049)	0.034 (± 0.042)
ROH with (MAF 1%, max gap 250kb)						
Line	Number of segments	Number of SNPs per segment	Size of segments (Mb)	% ROH>10Mb	F <sub>ROH</sub>	F <sub>ROH&gt;10Mb</sub>
SA	77.9 (± 11.1)	75 (± 57)	3.15 (± 2.32)	1.82	0.160 (± 0.026)	0.013 (± 0.011)
SB	80.6 (± 10.9)	78 (± 57)	3.30 (± 2.33)	2.51	0.176 (± 0.032)	0.016 (± 0.013)
SC	76.7 (± 12.5)	75 (± 56)	3.11 (± 2.29)	2.05	0.156 (± 0.032)	0.013 (± 0.011)
SD	56.1 (± 11)	76 (± 63)	3.04 (± 2.45)	2.53	0.110 (± 0.026)	0.013 (± 0.012)
SU	56.6 (± 14.1)	77 (± 67)	3.14 (± 2.69)	3.04	0.115 (± 0.040)	0.015 (± 0.013)
SYn	49.5 (± 12.4)	73 (± 59)	2.89 (± 2.28)	1.77	0.092 (± 0.030)	0.008 (± 0.010)
SYn+5	53.5 (± 15)	74 (± 63)	2.98 (± 2.45)	2.66	0.103 (± 0.047)	0.013 (± 0.019)
ROH with (MAF 5%, max gap 1000kb)						
Line	Number of segments	Number of SNPs per segment	Size of segments (Mb)	% ROH>10Mb	F <sub>ROH</sub>	F <sub>ROH&gt;10Mb</sub>
SA	61 (± 8.8)	93 (± 85)	4.8 (± 4.37)	9.57	0.163 (± 0.027)	0.050 (± 0.023)
SB	58.1 (± 7.5)	108 (± 101)	5.75 (± 5.43)	15.20	0.186 (± 0.034)	0.079 (± 0.032)
SC	58 (± 8.9)	96 (± 95)	4.91 (± 4.85)	11.29	0.159 (± 0.034)	0.058 (± 0.027)
SD	46.4 (± 8.6)	91 (± 94)	4.35 (± 4.55)	8.31	0.112 (± 0.027)	0.035 (± 0.020)
SU	45 (± 9.2)	95 (± 108)	4.69 (± 5.46)	10.35	0.117 (± 0.039)	0.045 (± 0.035)
SYn	42.1 (± 9.7)	86 (± 84)	4.03 (± 4.17)	7.94	0.094 (± 0.029)	0.029 (± 0.020)
SYn+5	45.1 (± 9.3)	88 (± 94)	4.24 (± 4.65)	8.13	0.106 (± 0.049)	0.033 (± 0.042)
ROH with (MAF 5%, max gap 250kb)						
Line	Number of segments	Number of SNPs per segment	Size of segments (Mb)	% ROH>10Mb	F <sub>ROH</sub>	F <sub>ROH&gt;10Mb</sub>
SA	71.4 (± 11.2)	68 (± 46)	3.06 (± 1.94)	1.17	0.147 (± 0.025)	0.007 (± 0.008)
SB	74.1 (± 11.3)	70 (± 47)	3.16 (± 2.02)	1.40	0.160 (± 0.031)	0.008 (± 0.009)
SC	68.8 (± 12.5)	68 (± 49)	3.00 (± 2.03)	1.36	0.140 (± 0.031)	0.008 (± 0.008)
SD	52.5 (± 10.7)	71 (± 56)	3.04 (± 2.29)	2.06	0.105 (± 0.025)	0.010 (± 0.009)
SU	52.2 (± 14.1)	71 (± 59)	3.07 (± 2.49)	2.33	0.107 (± 0.038)	0.011 (± 0.011)
SYn	45.4 (± 12.3)	70 (± 52)	2.89 (± 2.13)	1.45	0.086 (± 0.030)	0.006 (± 0.008)
SYn+5	48.7 (± 15.1)	71 (± 54)	2.99 (± 2.21)	2.61	0.096 (± 0.045)	0.010 (± 0.015)

#### 4. Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

Dans les chapitres 3 à 5 de la thèse, seront étudiées de nouvelles cohortes ou lignées des populations de trois des entreprises qui ont fait l'objet de l'étude de diversité génétique présentée dans l'article ci-dessus. Ici, les différents estimateurs de consanguinité ont donc été recalculés avec les génotypes obtenus dans ces nouvelles cohortes pour confirmer et préciser les résultats précédents grâce à un nombre d'individus génotypés beaucoup plus important dans chaque population.

Pour la lignée SA, les générations étudiées dans l'article publié dans GSE et dans le projet SG-Truite ne proviennent pas de la même lignée du sélectionneur. Les deux lignées SA<sub>A</sub> et SA<sub>B</sub> ont la même origine, mais sont séparées depuis 9 générations de sélection (Figure 10). Pour la lignée SC, les générations proviennent de la même lignée SC à 1 génération d'écart (Figure 10). Pour la lignée SB, les cohortes proviennent de la même lignée SB à 2 générations d'écart (Figure 10).



**Figure 10 :** Chronologie de création et de sélection des différentes cohortes des lignées commerciales étudiées durant la thèse.

Les cohortes en couleur hachurée ont été étudiées pour la diversité génétique dans l'article présenté ci-dessus (projet 57K-Truite) et les générations en couleur pleine sont étudiées dans les chapitres suivants pour l'architecture des caractères et l'efficacité de la sélection génomique (projet SG-Truite).

## Chapitre 2: Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

Pour les analyses de l'article présenté ci-dessus, nous avons 48 à 49 individus génotypés selon les entreprises avec des individus choisis pour être les moins apparentés possibles dans les cohortes du projet 57K-Truite. Pour le projet SG-Truite, l'estimation de la consanguinité à partir des ROH est faite avec l'ensemble des individus disponibles. Nous avons effectué un échantillonnage de 73 à 93 individus par cohorte en choisissant un individu par famille avec des apparentements uniquement entre les familles. Nous avons répété l'échantillonnage 15 fois pour obtenir une moyenne et un écart-type sur l'ensemble des échantillonnages pour le DL et le Ne.

### Effectif efficace

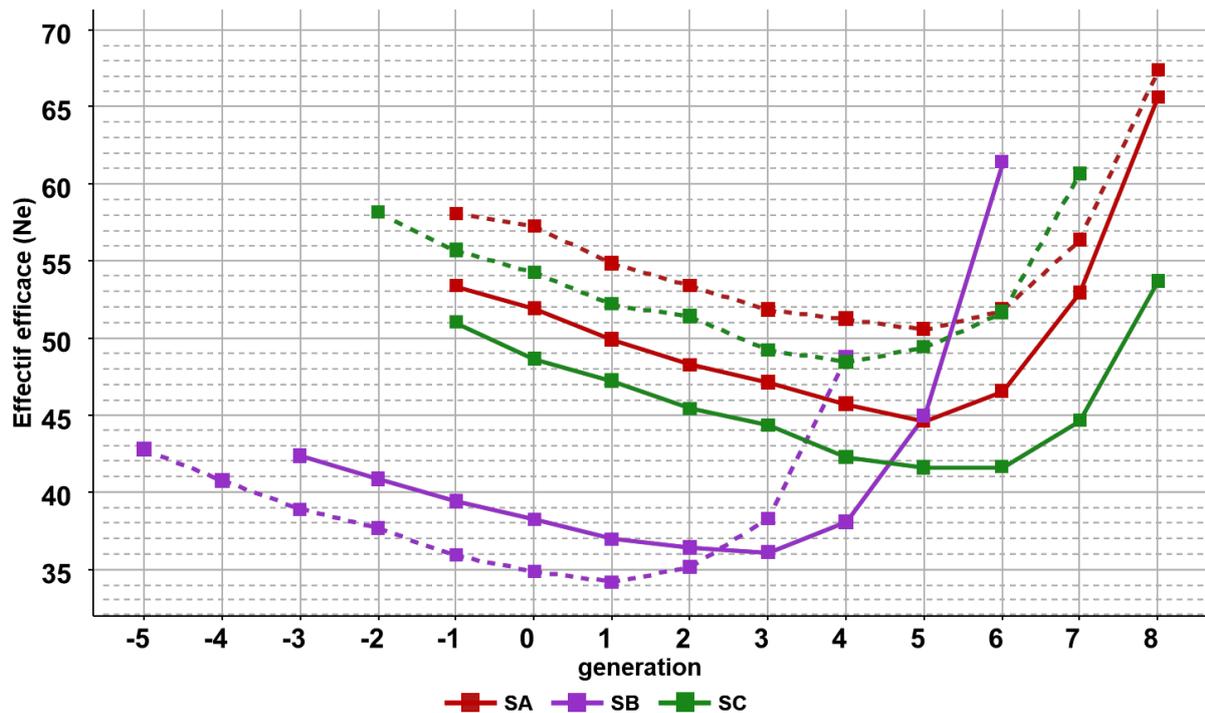
La [Figure 11](#) représente les effectifs efficaces des différentes lignées calculées (sans les chromosomes 5 et 20) dans le cadre du projet 57K-truite (article ci-dessus) et du projet SG-Truite au cours des dix dernières générations. Le déséquilibre de liaison est très élevé sur les chromosomes 5 et 20 en raison d'inversions importantes dans les deux chromosomes limitant la recombinaison et diminuant fortement les estimations de Ne. L'ensemble des résultats obtenus pour les 15 échantillonnages réalisés sont résumés dans [Annexe - tableau 1](#) et [Annexe - tableau 2](#).

L'effectif efficace de la lignée SA<sub>B</sub> est estimé à 41 et 65 respectivement avec et sans les chromosomes 5 et 20 (chr 5 et chr 20). Celui de la lignée SA<sub>A</sub> avait été quant à lui estimé à 48 et 68 avec et sans les chromosomes 5 et 20. De même, pour la lignée de SB, les Ne sont estimés à 49 et 61 (avec ou sans chr 5 et 20) pour la génération 7 alors qu'ils avaient été estimés à 37 et 49 pour la génération 5. Pour SC, les Ne estimés sont de 37 et 53, avec ou sans chr 5 et 20, pour la génération et avaient été estimés à 39 et 61 pour la génération 8.

Sur la [Figure 11](#), on note des augmentations du Ne pour les dernières générations dans toutes les lignées. Ces augmentations sont vraisemblablement liés à l'utilisation de la méthode d'optimisation des contributions (OCS) pour sélectionner les candidats retenus comme reproducteurs et les apparier en créant des croisements minimisant la consanguinité intra-famille grâce au logiciel OPTIVAR développé par le SYSAAF avec l'appui conceptuel de Jean-Jacques Colleau (INRAE).

## Chapitre 2: Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

Pour les deux lignées SA, on note que les  $N_e$  des cohortes parentales et grand-parentales des 9<sup>ème</sup> générations de sélection des deux lignées sont similaires, même si des  $N_e$  plus élevés sont estimés pour la lignée SA<sub>A</sub> dans les générations antérieures. Ceci doit être lié à une sélection opérée différemment au cours du temps entre les deux lignées (nombre de reproducteurs, intensité de sélection).



**Figure 11 :** Effectif efficace ( $N_e$ ) sur les 10 générations précédant la cohorte génotypée, calculé à partir du déséquilibre de liaison moyenné par chromosome (sans les chromosomes 5 et 20) pour les différentes lignées étudiées dans le projet 57K-Truite (ligne pointillée) et dans le projet SG-Truite (ligne pleine).

Les lignées SA (rouge) des deux projets sont issues d'une même population il y a 9 générations. Pour les lignées SB (violette) et SC (vert), les cohortes analysées dans le projet SG-Truite sont plus récentes respectivement de 2 et 1 générations des cohortes analysées dans le projet 57K-Truite.

Pour SC, l'estimation du  $N_e$  d'une même génération d'ancêtres est plus faible quand elle est effectuée à partir des données de génotypage de la génération 9 (SG-Truite) qu'à partir des données obtenues en génération 8 (57K-Truite). Cette différence est significative d'après les écarts-types estimés par échantillonnage ([Annexe - tableau 1](#) et [Annexe - tableau 2](#)). Cette différence d'estimation du  $N_e$  peut être expliquée par le non-respect des hypothèses d'une population idéale de Wright-Fisher sur la base des calculs réalisés. En effet, la sélection réalisée entre les générations 8 et 9 a dû engendrer une perte de contribution en génération 9 de certains ascendants mieux représentés dans la génération 8. Cela indique donc que malgré la méthode OCS une perte de variabilité génétique est tout de même associée à la sélection en cours.

## Chapitre 2: Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

Le même phénomène est observé pour SB (à deux générations de décalage) pour les effectifs efficaces de reproducteurs des générations les plus récentes d'ancêtres. En revanche, les estimations des effectifs efficaces calculés pour les générations d'ancêtres les plus anciennes sont un peu plus élevées quand elles sont calculées à partir des génotypes de la génération 7 que de la génération 5. L'OCS n'ayant été mise en oeuvre que depuis 2012-2013 par le SYSAAF, c'est-à-dire juste après la production de la génération 5 de la lignée SB, cette différence d'estimation traduit vraisemblablement le fait que l'OCS a réellement permis de rééquilibrer les contributions des ancêtres fondateurs en génération 7 par rapport aux générations précédentes.

### Consanguinité calculée à partir des ROHs

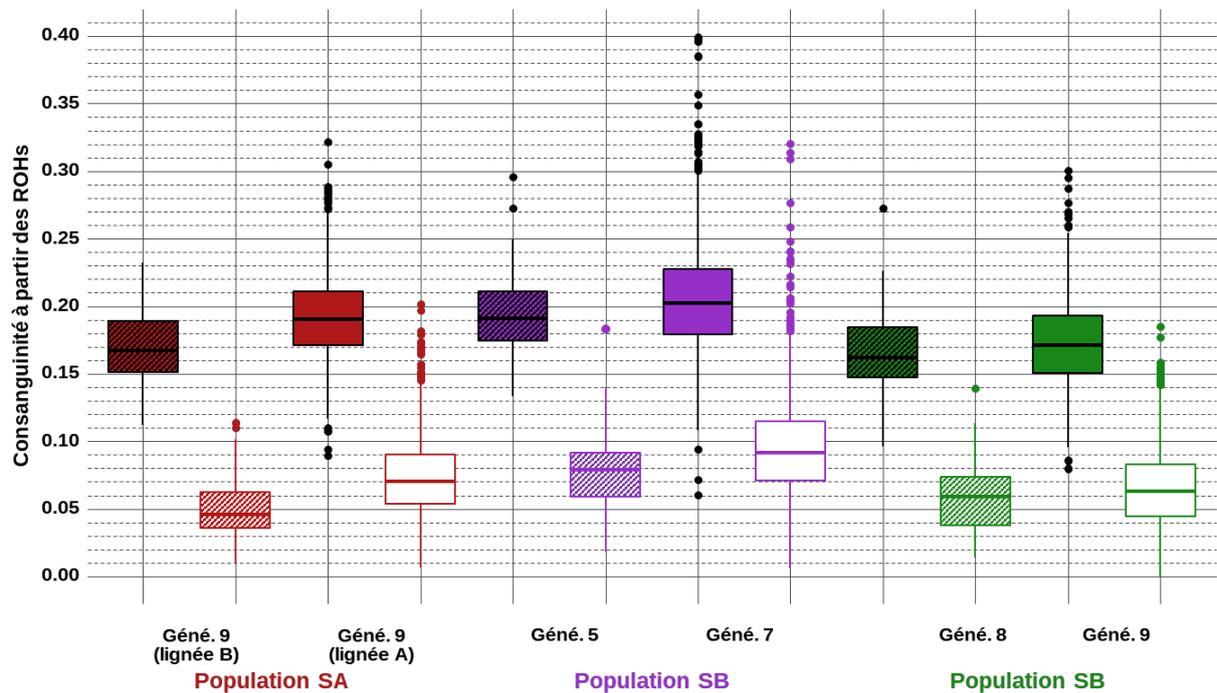
Les ROHs sont définis avec les mêmes règles que dans l'article présenté dans le chapitre 2 (taille minimale 1 Mb et 30 SNPs minimum par ROH, une densité de 1 SNP par 100 kb, distance maximale entre deux SNPs de 1 000 kb, 1 SNP hétérozygote autorisé par ROH et maximum 2 SNPs manquants par ROH). La consanguinité totale ( $F_{ROH}$ ) est calculée sur l'ensemble des ROHs de chaque individu et la consanguinité récente (environ 3 générations) est calculée à partir des ROHs supérieurs à 10 Mb ( $F_{ROH>10Mb}$ ).

Les consanguinités totale et récente calculées à partir des ROHs augmentent pour SB et SC dans les nouvelles générations étudiées dans SG-Truite (Figure 12).

Pour SB, la consanguinité estimée de la 7<sup>ème</sup> génération est de 20,6% et 9,6% pour  $F_{ROH}$  et  $F_{ROH>10Mb}$ , respectivement alors qu'elle est de 19,5% et 7,9% respectivement à la 5<sup>ème</sup> génération (Figure 12). L'augmentation est de 1,1 et 1,7 pour  $F_{ROH}$  et  $F_{ROH>10Mb}$  pour deux générations d'écart, soit un  $\Delta F$  de 0,69 et 0,91 % par génération pour les consanguinités totale et récente, respectivement. Pour SC, la consanguinité estimée de la 9<sup>ème</sup> génération est de 17,3% et 6,5% pour  $F_{ROH}$  et  $F_{ROH>10Mb}$ , respectivement alors qu'elle est de 16,6% et 5,8% respectivement à la 8<sup>ème</sup> génération (Figure 12). L'augmentation est de 0,70 et 0,75 pour  $F_{ROH}$  et  $F_{ROH>10Mb}$  pour une génération d'écart, soit un  $\Delta F$  de 0,70% et 0,71% pour les consanguinités totale et récente sur une génération.

Pour SA, la consanguinité estimée de la lignée  $SA_B$  est de 19,2% et 7,2% pour  $F_{ROH}$  et  $F_{ROH>10Mb}$ , respectivement à la 9<sup>ème</sup> génération contre 17,1% et 5,1% respectivement pour la

lignée SA<sub>A</sub>. On ne peut cependant pas comparer directement ces chiffres qui correspondent à deux lignées gérées indépendamment avec des objectifs de sélection différents.



**Figure 12 :** Boîtes à moustaches de la consanguinité totale ( $F_{ROH}$ , boîtes à moustaches pleines) et de la consanguinité récente ( $F_{ROH>10Mb}$ , boîtes à moustaches transparentes) pour les cohortes étudiées dans le projet 57K-Truite (boîtes à moustaches hachurées) et SG-Truite (boîtes à moustaches non hachurées).

Géné. x : génération x

Le SYSAAF a mis en place chez ses adhérents sélectionneurs le choix des reproducteurs avec une méthode d'optimisation de contributions (OCS) pour avoir un maximum de gain génétique tout en contraignant l'augmentation de la consanguinité (estimée sur pedigree) à un maximum d'1% par génération (voir ci-dessus). Nos résultats avec un calcul de  $\Delta F$  génomique à partir des ROH confirment les résultats de l'article présenté dans le chapitre 2 avec la stabilisation ou l'augmentation de l'effectif efficace sur les dernières générations de sélection grâce au contrôle de l'augmentation de la consanguinité. Les  $N_e$  calculés à partir des  $\Delta F$  ( $F_{ROH}$ ) d'après la formule :  $N_e = (2 * \Delta F)^{-1}$  sont de 72 pour SB et 71 pour SC. Ces estimations sont supérieures d'une dizaine d'unités aux  $N_e$  calculés à partir du DL.

Nous avons montré dans le même article que l'occurrence ( $O_{ROH}$ ) d'un SNP dans un ROH varie le long du génome avec des régions à forte  $O_{ROH}$  (hot-spots) variables d'une population à une

## Chapitre 2: Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

autre. Certaines zones avec forte  $O_{ROH}$  sont communes à plusieurs lignées comme sur le chromosome 10 entre 48 et 50 Mb (figure 2 annexe article).

On remarque une corrélation forte au sein de la même population pour les  $O_{ROH}$  (Tableau 2).

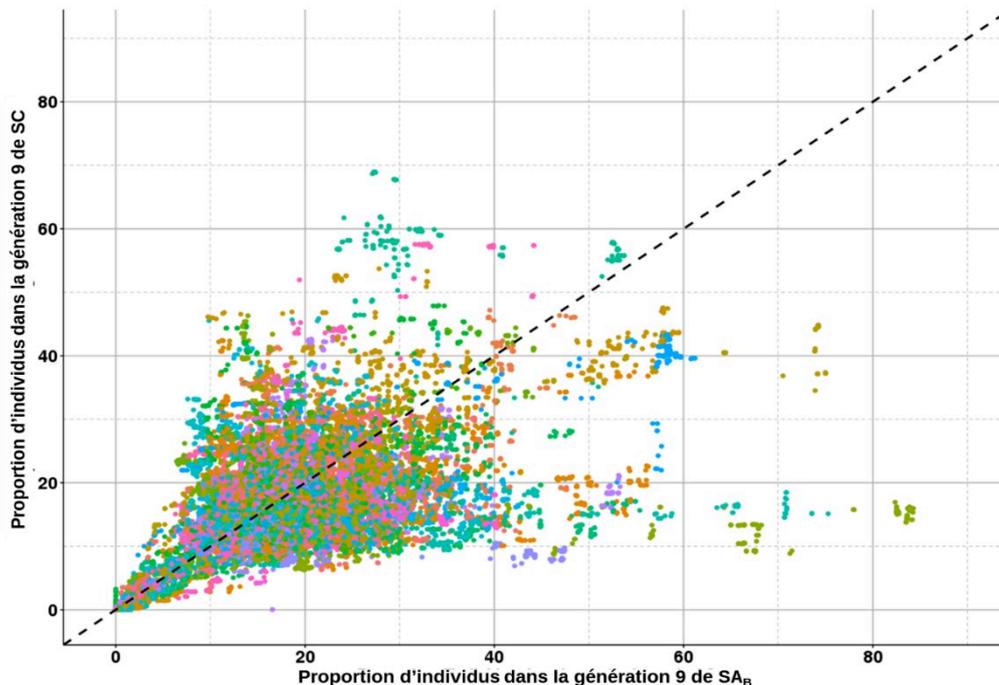
Lignée	$SA_B$ - 9g.	SB - 5g.	SB - 7g.	SC - 8g.	SC - 9g.
$SA_A$ - 9g	<b>0,56</b>	0,20	0,30	0,34	0,38
$SA_B$ - 9g		0,24	0,28	0,36	0,44
SB - 5g			<b>0,81</b>	0,21	0,26
SB - 7g				0,30	0,34
SC - 8g					<b>0,74</b>

**Tableau 2 :** Corrélations de Pearson entre la proportion d'individus pour chaque SNP dans un ROH entre les différentes cohortes étudiées.

$SA$ ,  $SB$ , et  $SC$  désignent l'origine de la lignée selon l'entreprise de sélection ; x g. désigne le nombre de générations de sélection.

Les  $O_{ROH}$  seraient donc globalement conservées d'une génération à l'autre, même si la fréquence des différents haplotypes peut évoluer avec la sélection. Les corrélations sont beaucoup plus faibles pour  $O_{ROH}$  entre les populations des différents sélectionneurs variant de 0,20 à 0,44 (Tableau 2). On remarque une  $O_{ROH}$  un peu plus forte entre les deux lignées de  $SA$  qui proviennent d'une origine commune (Tableau 2).

La Figure 8 représente une comparaison des  $O_{ROH}$  entre deux populations ( $SA_B$  et  $SC$ ).



**Figure 13 :** Proportion d'individus où les SNPs sont conjointement dans des ROHs de  $SA_B$  (axe x) et  $SC$  (axe y).

Chaque point représente un SNP et les couleurs correspondent aux différents chromosomes.

## Chapitre 2: Estimation de la consanguinité dans les populations étudiées pour le projet SG-Truite

Pour certains SNPs, nous observons de fortes  $O_{ROH}$  dans une des populations alors qu'elles sont faibles dans l'autre population (par exemple, une zone  $O_{ROH} > 80\%$  dans la lignée  $SA_B$  et  $< 20\%$  dans la lignée  $SC$ ). Mais, on remarque aussi des  $O_{ROH}$  fortes dans les deux populations ( $> 50\%$  dans les deux, [Figure 13](#)). Cela peut être lié à des zones du génome avec un faible taux de recombinaison ou bien à des zones d'intérêt pour la domestication ou pour la production qui sont alors fortement sélectionnées dans toutes les populations.

Depuis la sortie de notre article, d'autres travaux ont été publiés sur l'étude des ROHs dans des populations piscicoles. Yoshida et al. (2020) montrent un  $F_{ROH}$  de 14% et 15% pour deux populations de saumon coho en sélection pour la croissance pour la 7<sup>ème</sup> et 8<sup>ème</sup> génération de sélection (environ 10% de gain génétique par génération) montrant ainsi une forte consanguinité par rapport aux populations animales terrestres d'élevage, mais proche des populations trutticoles des sélectionneurs français. Leur étude montre aussi une consanguinité très faible (0,4%) pour une population croisée à partir des pères et des mères provenant des 2 populations sélectionnées depuis 7 et 8 générations, avec principalement des ROH de petite taille (1-2Mb). Les mêmes auteurs montrent également des corrélations faibles ( $< 0,4$ ) entre  $F_{ROH}$  et estimation de la consanguinité à partir du pedigree. Une étude chez le turbot (Morales-González et al. 2020) montre l'importance d'adapter les paramètres de calcul des ROHs par rapport à l'espèce étudiée, en raison notamment des distributions différentes des ROHs et d'un génome de taille 3,5 fois inférieure à celui de la truite arc-en-ciel. Une autre étude, également chez le turbot (Aramburu et al. 2020,) montre l'association entre les hot-spots de ROHs et des QTLs. Les ROHs peuvent donc être un puissant outil pour identifier des zones du génome d'intérêt pour la domestication ou la sélection des espèces piscicoles.

La mise en oeuvre de la SG en routine permettra d'accumuler des données génomiques et donc d'utiliser les ROHs non seulement comme un puissant outil de recherche pour répondre à différentes questions évolutives, mais aussi pour gérer la diversité génétique dans les schémas de sélection.

## 5. Résumé

Les populations commerciales françaises de truites arc-en-ciel sont sous sélection en population fermée depuis plusieurs décennies, ce qui entraîne une perte de diversité génétique, élément clé pour conserver le progrès génétique à long terme ainsi que les capacités d'adaptation à des changements d'environnement. L'objectif de ce chapitre est de présenter l'outil génomique utilisé durant la thèse, la puce MD (Axiom™ Trout Genotyping array avec 57 501 SNPs) et une analyse de la diversité génétique au sein et entre sept lignées françaises de truite arc-en-ciel expérimentale ou commerciale. Cette étude est basée sur divers indicateurs génétiques moléculaires : indice de fixation ( $F_{ST}$ ), déséquilibre de liaison (DL), taille effective de la population ( $N_e$ ) et coefficient de consanguinité dérivé des ROHs. Pour certaines lignées, nous avons les données sur différentes générations successives. Après un premier tri important, nous avons une liste de seulement 38 350 SNPs fiables sur les 57 501 disponibles sur la puce avec une répartition inégale entre chromosomes et une couverture plus importante pour les centromères par rapport aux télomères. Plusieurs zones du génome, de taille  $> 1$  Mb, ne sont pas couvertes par des SNPs.

Nos résultats montrent une différenciation génétique modérée entre les populations des sélectionneurs ( $F_{ST}$  variant de 0,08 à 0,15). Le DL décline rapidement dans les premiers 100 kb, mais se maintient à longue distance ( $> 1$  Mb). L'effectif efficace varie entre 24 et 68 selon la méthode utilisée. La consanguinité calculée à partir des ROHs varie de 10% pour une lignée expérimentale à 20,5% pour une lignée sélectionnée, ce qui est clairement supérieure aux estimations chez les ruminants et les porcs. Pour deux lignées sélectionnées avec des données sur les dernières générations, nous avons un taux d'augmentation de la consanguinité inférieur à 1% par génération et une corrélation forte ( $> 0,78$ ) pour l'occurrence des mêmes SNPs dans un ROH d'une génération à une autre. La consanguinité des lignées françaises est modérée à forte. Cela est dû à plusieurs facteurs, en particulier des effets fondateurs à la création des lignées, mais aussi à des succès reproducteurs différents ainsi qu'à la sélection pour les lignées soumises à sélection volontaire. Il serait peut être judicieux d'envisager une gestion de la consanguinité à l'échelle du génome pour améliorer la gestion de la diversité génétique dans les lignées. Une forte variation du DL et des ROHs est observée le long du génome. Cela peut être lié à des taux de recombinaisons locales différents et/ou à des variants sélectionnés pour des gènes clés pour la domestication ou pour la production. C'est la première étude en aquaculture utilisant les ROHs permettant de montrer la puissance de cet outil tant pour

apprécier finement le niveau de consanguinité que pour identifier des zones du génome associées aux processus de domestication et sélection.

## 6. Annexes

<i>L/G</i>	1	2	3	4	5	6	7	8	9	10
<b>SA<sub>A</sub> - 9 g.</b>	67 (2)	56 (1)	52 (1)	51 (1)	51 (1)	52 (1)	53 (1)	55 (1)	57 (1)	58 (1)
<b>SB - 5 g.</b>	49 (1)	38 (0)	35 (0)	34 (0)	35 (0)	36 (1)	38 (1)	39 (1)	41 (1)	43 (1)
<b>SC - 8 g.</b>	61 (2)	52 (1)	49 (1)	48 (1)	49 (1)	51 (1)	52 (1)	54 (1)	56 (1)	58 (1)
<b>SA<sub>B</sub> - 9 g.</b>	66 (3)	53 (2)	47 (1)	45 (1)	46 (1)	47 (1)	48 (1)	50 (1)	52 (1)	53 (1)
<b>SB - 9 g.</b>	61 (2)	45 (1)	38 (1)	36 (0)	36 (0)	37 (0)	38 (0)	39 (0)	41 (0)	42 (0)
<b>SC - 7 g.</b>	54 (3)	45 (1)	42 (1)	42 (1)	42 (1)	44 (1)	45 (1)	47 (1)	49 (0)	51 (1)

**Annexe - tableau 1 :** Effectif efficace calculé à partir des différentes lignées et différentes générations sans le chromosome 5 et 20.

L. : Lignée et génération, G : Nombre de génération antérieur de la génération utilisée pour estimer le Ne.

<i>L/G</i>	1	2	3	4	5	6	7	8	9	10
<b>SA<sub>A</sub> - 9 g.</b>	48 (2)	43 (1)	41 (1)	41 (0)	43 (0)	44 (0)	46 (0)	48 (0)	50 (0)	52 (0)
<b>SB - 5 g.</b>	37 (1)	32 (1)	31 (1)	30 (1)	31 (1)	33 (0)	34 (1)	36 (0)	38 (0)	40 (0)
<b>SC - 8 g.</b>	39 (1)	39 (1)	39 (1)	40 (1)	41 (1)	44 (0)	45 (0)	47 (0)	49 (0)	52 (0)
<b>SA<sub>B</sub> - 9 g.</b>	41 (1)	37 (0)	35 (0)	35 (1)	37 (1)	38 (1)	40 (1)	42 (1)	44 (1)	46 (1)
<b>SB - 9 g.</b>	49 (1)	38 (0)	33 (0)	32 (0)	33 (0)	34 (1)	35 (1)	37 (1)	38 (1)	40 (1)
<b>SC - 7 g.</b>	37 (1)	34 (0)	33 (0)	34 (0)	35 (1)	38 (1)	39 (1)	41 (1)	43 (1)	45 (1)

**Annexe - tableau 2 :** Effectif efficace calculé à partir des différentes lignées et différentes générations avec l'ensemble des chromosomes.

L. : Lignée et génération des génotypes, G : Nombre de génération antérieur de la génération utilisée pour estimer le Ne.

## 7. Références

- Affymetrix. AxiomTM Analysis Suite 2.0: UserGuide [internet : [https://media.affymetrix.com/support/downloads/manuals/axiom\\_analysis\\_suite\\_user\\_guide.pdf](https://media.affymetrix.com/support/downloads/manuals/axiom_analysis_suite_user_guide.pdf)]. 2016.
- Aramburu O, Ceballos F, Casanova A, Le Moan A, Hemmer-Hansen J, Bekkevold D, et al. 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*). *Front Genet.* 2020;11(April):1–14.

- <https://doi.org/10.3389/fgene.2020.00296>
- D'Ambrosio J, Phocas F, Haffray P, Bestin A, Brard-Fudulea S, Poncet C, et al. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genet Sel Evol.* 2019;51(1). <https://doi.org/10.1186/s12711-019-0468-4>
- FAO. Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk. [internet : <http://www.fao.org/3/a-w9361e.pdf>]. Food and Agricultural Organization. 1998.
- Felsenstein J. The effect of linkage on directional selection. *Genetics.* 1965;52(2):349–63.
- Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics.* 2018;19(1):98. <https://doi.org/10.1186/s12864-018-4453-z>
- McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. *Am J Hum Genet.* 2008;83(3):359–72. <https://doi.org/10.1016/j.ajhg.2008.08.007>
- Morales-González E, Saura M, Fernández A, Fernández J, Pong-Wong R, Cabaleiro S, et al. Evaluating different genomic coancestry matrices for managing genetic variability in turbot. *Aquaculture.* 2020;520(January):734985. <https://doi.org/10.1016/j.aquaculture.2020.734985>
- Omyk\_1.0 [internet : [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_002163495.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1/)].
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K SNP array for rainbow trout. *Mol Ecol Resour.* 2015;15(3):662–72. <https://doi.org/10.1111/1755-0998.12337>
- Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E. A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol Ecol Resour.* 2014;14. <https://doi.org/10.1111/1755-0998.12204>
- Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, et al. Sex-dependent dominance maintains migration supergene in rainbow trout. *bioRxiv.* 2018;504621. <https://doi.org/10.1101/504621>
- Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet.* 2012;13(1):70. <https://doi.org/10.1186/1471-2156-13-70>
- Yoshida GM, Cáceres P, Marín-Nahuelpi R, Koop BF, Yáñez JM. Estimates of Autozygosity Through Runs of Homozygosity in Farmed Coho Salmon. *Genes (Basel).* 2020;11(5). <https://doi.org/10.3390/genes11050490>
- Omyk\_1.0 [internet : [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_002163495.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1/)].



## Chapitre 3: Architecture génétique et sélection des caractères de reproduction femelle

Ce chapitre a fait l'objet d'un article publié dans BMC Genomics (mis en annexe) : D'Ambrosio J, Morvezen R, Brard-Fudulea S, Bestin A, Acin Perez A, Guéméné D, Poncet C., Haffray P., Dupont-Nivet M., Phocas, F. *Genetic architecture and genomic selection of female reproduction traits in rainbow trout*. BMC Genomics. 2020; 21(1):558. <https://doi.org/10.1186/s12864-020-06955-7>

### 1. Introduction

L'amélioration génétique chez la truite arc-en-ciel porte essentiellement sur les performances de croissance, de découpe et de qualité de la chair et plus récemment de la résistance à un nombre encore limité de pathologies (Chavanne et al. 2016; Abdelrahman et al. 2017). Petit à petit, des phénotypes plus complexes comme le rendement de découpe, la couleur de la chair, l'efficacité alimentaire ou l'adaptation à une alimentation végétale sont étudiés et parfois intégrés dans les schémas de sélection (Haffray et al. 2013; Callet et al. 2017).

La truite arc-en-ciel est une espèce aquacole très féconde avec environ 1500 à 2000 œufs produits par an et par kilo de femelle (Jalabert and Fostier 2010). Les principaux sélectionneurs français produisent plus de 700 millions d'œufs embryonnés vendus en France et à l'international (dans plus de 20 pays d'Europe et du Moyen-Orient). Les caractères de reproduction femelle n'ont jusqu'à maintenant que peu fait l'objet de sélection, si ce n'est historiquement le caractère de précocité sexuelle dans la saison. Toutefois, et comme en aviculture ou en production porcine ou bovine, ces caractères jouent un rôle important dans l'économie d'une entreprise de sélection en raison de la vente des œufs embryonnés destinés à produire de futurs alevins pour les grossisseurs. L'augmentation de la taille des œufs est généralement associée à une meilleure survie à l'éclosion et à un démarrage plus facile des alevins acceptant des aliments de grosse taille. L'augmentation du nombre d'œufs vendus par tonne de femelles élevées constitue un ratio économique ayant des répercussions importantes sur le coût de production des œufs embryonnés. De plus, la production de caviar d'œufs de truite est un marché en plein essor (111 T/an) (AGRESTE 2020). Une sélection sur la masse d'œufs (caractère composite associant nombre et taille des œufs) rapportée au poids des

femelles pourrait améliorer la productivité de la filière destinée à la production de caviar d'œufs de truite.

Depuis les années 1970, plusieurs études ont estimé des héritabilités généralement moyennes pour les caractères de reproduction femelle (date de ponte, poids de la ponte, nombre d'œufs, taille des œufs) chez la truite arc-en-ciel et chez d'autres espèces de salmonidés (Gall 1975; Gall and Gross 1978; Gjedrem et al. 1986; Gall and Huang 1988; Su et al. 1997, 2002; Gall and Neira 2004), la truite arc-en-ciel constituant l'espèce aquacole la plus étudiée pour ce type de caractères. Des sélections massales ont été mises en place de longue date aux USA sur des caractères de précocité de la date de ponte avec aujourd'hui des lignées à périodes de ponte différentes, dites estivales, automnales, hivernales ou printanières. A ce jour, et à l'échelle internationale aucun programme de sélection commerciale intégrant d'autres types de caractères de ponte n'est recensé, sauf en France, pays dans lequel l'amélioration de la productivité des écloseries s'est développée depuis quelques années. La combinaison de différents caractères de production, de découpe, de résistance à des pathologies et de performances de ponte implique l'usage d'une sélection multicaractère pouvant être abordée plus efficacement par la sélection généalogique et bien sûr aujourd'hui éventuellement par grâce à la SG. Au regard des enjeux économiques potentiels, il est opportun d'étudier l'architecture génétique des caractères de reproduction pour, d'une part évaluer si la sélection génomique pourrait être plus efficace que la sélection généalogique et d'autre part connaître les associations génétiques favorables et/ou défavorables entre ces caractères et les caractères de production déjà sélectionnés afin de mieux piloter la sélection en tenant compte de potentiels antagonismes génétiques.

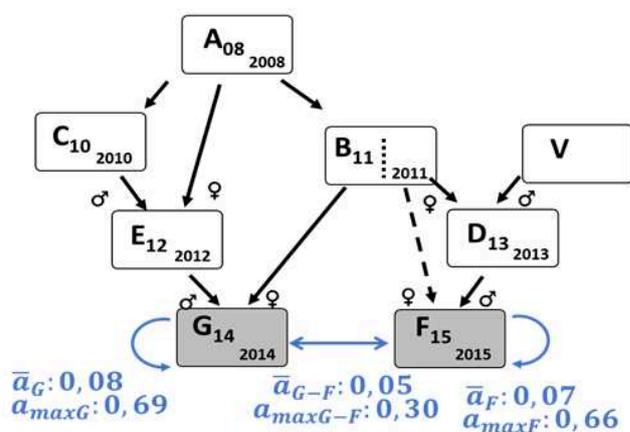
Les objectifs de cette étude pour les caractères de reproduction femelle sont : (i) de confirmer les estimations de paramètres génétiques des caractères de reproduction déjà publiées aux USA il y a une vingtaine d'années mais ici mesurés dans les conditions françaises d'élevage et sur la lignée d'un sélectionneur français (Viviers de Sarrance) avec une histoire de sélection récente sur la croissance ; (ii) identifier d'éventuelles régions du génome (QTL) pouvant avoir un effet sur ces caractères ; (iii) évaluer l'efficacité d'une sélection génomique *versus* sur pedigree; (iv) évaluer l'impact des facteurs de variation de la précision de la sélection génomique (densité de marqueurs, taille et structure de la population de référence).

## 2. Matériel et méthodes

### 2.1. Caractéristiques de la population étudiée

Les animaux proviennent de deux cohortes de l'entreprise de sélection Viviers de Sarrance basée dans les Pyrénées-Atlantique (64).

Les cohortes étudiées AG92 et FG01 renommées G<sub>14</sub> et F<sub>15</sub> respectivement proviennent de la 9<sup>ème</sup> génération d'une sélection massale de type PROSPER (Chevassus et al. 2004) portant essentiellement sur la croissance jusqu'à 18 mois et le rendement à l'aide de mesures échographiques (Haffray et al. 2013). Les mères des cohortes G<sub>14</sub> et F<sub>15</sub> proviennent de la même cohorte B<sub>11</sub> mais ne sont pas les mêmes entre les deux cohortes (Figure 14). Les mères de la cohorte F<sub>15</sub> ne sont pas phénotypées. Les pères proviennent de deux autres cohortes différentes, cohorte E<sub>12</sub> pour G<sub>14</sub> et cohorte D<sub>13</sub> pour F<sub>15</sub>. La cohorte E<sub>12</sub> a été étudiée pour des caractères de reproduction femelle dans un projet antérieur « Best'Ov » soutenu par le FEAMP.



**Figure 14 :** Schéma de parenté entre les cohocohortes.

Les années de naissance sont précisées. G<sub>14</sub> et F<sub>15</sub> (gris) sont les deux cohocohortes phénotypées et génotypées. En bleu, l'apparentement moyen ( $\bar{a}$ ) et maximum ( $a_{\max}$ ) inter et intra cohocohortes G<sub>14</sub> et F<sub>15</sub>.

L'apparentement moyen entre les individus génotypés de la cohorte G<sub>14</sub> et F<sub>15</sub> est estimé à partir du calcul des identités par état (IBS, Identity by state) des marqueurs de la puce (voir ci-dessous) avec le logiciel PLINK (option—genome). L'IBS entre deux individus varie entre 0 et 1, 0 si aucun allèle n'est commun aux deux individus et 1 si tous les allèles sont identiques. Les IBS moyens intra-cohorte entre les individus génotypés des cohortes G<sub>14</sub> et F<sub>15</sub> sont respectivement de 0,08 et 0,07. Les valeurs maximales d'apparentement intra-cohorte sont de 0,69 et 0,66 pour G<sub>14</sub> et F<sub>15</sub>, respectivement. L'apparentement moyen entre cohortes G<sub>14</sub> et F<sub>15</sub> est de 0,05 avec une valeur IBS maximale de 0,30. La structure familiale des deux cohortes est décrite plus loin.

## 2.2. Phénotypes bruts collectés

Les différents caractères de reproduction femelle ont été collectés pour 1518 femelles (755 femelles de G<sub>14</sub> et 763 femelles de F<sub>15</sub>). Les phénotypes ont été mesurés durant deux chantiers de phénotypage : de la semaine 49 de 2016 à la semaine 6 de 2017 pour G<sub>14</sub> et de la semaine 48 de 2017 à la semaine 4 de 2018 pour F<sub>15</sub>.

Les phénotypes recueillis sont le poids de la femelle pleine avant la ponte (PF) et après la ponte (PW), la longueur à la fourche de la femelle, la semaine de ponte, le poids total de la ponte (SW), le poids et le nombre d'œufs contenu dans une cuillère de 2,5 mL de volume, la longueur de 50 œufs alignés le long d'une règle de 20 cm, la présence ou non d'œufs surmatures dans la ponte. La semaine de ponte (SD), le poids de la ponte (SW), et le poids de la femelle après la ponte (PW) sont également disponibles pour l'étude pour certaines cohortes des 3 générations précédentes (321, 239, 287, 691 individus phénotypés dans les cohortes A<sub>08</sub>, C<sub>10</sub>, B<sub>11</sub> et E<sub>12</sub>, respectivement).

Les mesures dont les valeurs étaient inférieures ou supérieures à 4 écarts-types de la moyenne (intra-cohorte) ou les ratios (PW/PF, SW/PW) ont été éliminées du jeu de données analysé.

## 2.3. Phénotypes élaborés

À partir des phénotypes collectés sur la pisciculture, le phénotype brut « poids de la femelle pleine avant la ponte » (PF) et de nouveaux phénotypes élaborés feront l'objet d'analyses génétiques.

Le poids moyen d'un œuf (EW) est le rapport du poids du pool d'œufs sur le nombre d'œufs contenus dans une cuillère de 2,5 mL. Le nombre d'œufs dans la ponte (EN) est calculé comme le poids de la ponte divisé par le poids moyen d'un œuf. Le diamètre moyen d'un œuf (ED) est obtenu en divisant par 50 la longueur mesurée de 50 œufs alignés. La semaine de ponte (SD) est obtenue en classant coholes pontes de chaque cohorte des semaines les plus précoces aux semaines les plus tardives ; ainsi SD=1 correspond à la semaine 49 pour la cohorte G<sub>14</sub> et à la semaine 48 pour la cohorte F<sub>15</sub> et SD=5 à la 5<sup>ème</sup> semaine (et au-delà) de ponte pour les deux cohortes.

Les données dont les valeurs étaient extrêmes, inférieures ou supérieures à 4 écarts-types de la moyenne (intra-cohorte) ont été écartées du jeu de données analysé, éliminant de 1 à 17 données selon le caractère. Sur l'ensemble des caractères ([Tableau 3](#)), on observe une variabilité importante des performances, avec des coefficients de variation allant d'environ 5 à 38%.

Caractère	N	Moyenne	Ecart-type	Min	Max	CV (%)
Semaine de ponte (SD,#)	1517	2,52	1,54	1	5	-
Poids de la femelle avant la ponte (FW, g)	1516	2067	448	1010	3580	21,7
Poids de la femelle après la ponte (PW, g)	1515	1845	408	840	3116	22,1
Poids de la ponte (SW, g)	1517	187	71	12	398	38,0
Nombre d'oeuf dans la ponte (EN, #)	1507	4710	1743	577	10434	37,0
Poids moyen d'un oeuf (EW, mg)	1501	39,9	6,5	20,0	68,1	16,3
Diametre moyen d'un oeuf (ED, mm)	1504	4,03	0,19	3,32	4,72	4,7

**Tableau 3** : Statistiques descriptives des phénotypes élaborés mesurés dans les cohortes G<sub>14</sub> et F<sub>15</sub>.

CV : coefficient de variation

## 2.4. Génotypages

1536 échantillons de nageoires ont été collectés, mais seulement 1517 individus ont une quantité et une qualité suffisantes d'ADN pour permettre le génotypage avec la puce Axiom® Trout Genotyping array (57501 SNPs), ci-après nommée puce 57K. Les génotypages de 14 individus ont été éliminés, car moins de 95% des SNPs avaient pu être génotypés sur les échantillons correspondants.

Après ce premier contrôle qualité, 1503 individus génotypés sont retenus pour l'analyse: 1346 femelles des cohortes G<sub>14</sub> (726 individus) et F<sub>15</sub> (622 individus) et leurs parents (34 mères dont 1 mère sans descendant génotypé) et 32 pères pour G<sub>14</sub> ; 50 mères et 39 pères pour F<sub>15</sub>).

Le nombre final d'individus génotypés et phénotypés par cohorte passant les contrôles de qualité et la structuration familiale sont résumés dans le [Tableau 4](#).

Des contrôles qualité ont également été appliqués aux 38 350 SNPs exploitables de la puce 57K (voir chapitre 2) : call-rate > 97%, MAF > 5% et test d'écart à l'équilibre d'Hardy-Weinberg avec p-value < 0,0001 dans la population des 1503 individus génotypés. Cela conduit à conserver 29 799 SNPs pour l'analyse. Les contrôles de qualité ont été effectués avec Plink 1.9 (Purcell et al. 2007).

	Cohorte G <sub>14</sub>	Cohorte F <sub>15</sub>
<b>Descendants phénotypés</b>	726	620
<b>Pères</b>	32	39 ( 2 ind. avec père inconnu)
<b>Mères</b>	33 (45 ind. avec mère inconnu)	50
<b>Couples Père x Mère</b>	296	320
<b>Descendants par couple</b>	Moyenne : 4±2 min : 1 max : 11	Moyenne ± écart-type : 3± 2 min : 1 max : 7
<b>Descendants par père</b>	moyenne : 23±18 min : 5 max : 94	moyenne: 16± 8 min : 1 max : 36
<b>Descendants par mère</b>	moyenne : 20± 14 min : 4 max : 71	moyenne: 12± 7 min : 1 max : 28
<b>Nombre de demi-sœur par individus</b>	moyenne : 59± 31 min : 13 max : 145	moyenne: 30± 9 min : 9 max : 56

**Tableau 4 :** Nombre d'individus (génotypés et phénotypés passant les contrôles de qualité) dans le jeu de données final et structuration familiale des cohortes G<sub>14</sub> et F<sub>15</sub>.

± : écart-type

### 2.4.1. Puce moyenne densité 30 K

La puce MD (ou puce 30K) contenant les 29 799 marqueurs validés ci-dessus sera notre jeu de marqueurs de référence pour évaluer l'efficacité de la sélection génomique. Le génotype complet pour les 29 799 marqueurs est obtenu par imputation des génotypes aux marqueurs manquants pour tous les individus retenus pour l'analyse. L'imputation est réalisée à l'aide du logiciel FImpute (Sargolzaei et al. 2014) version 2.2. FImpute utilise une approche de fenêtre coulissante de SNPs le long du génome et l'information du pedigree et des haplotypes entre les individus aux génotypes à imputer et les individus aux génotypes connus pour imputer les SNPs manquants (Sargolzaei et al. 2014).

Le test des incompatibilités mendéliennes réalisé par FImpute sur les 29 799 marqueurs a également permis de contrôler la qualité du pedigree fourni quand les parents étaient génotypés. Dans ce cas, il a été vérifié que les descendants portaient l'un des allèles de leurs parents. Si le taux de marqueurs avec incompatibilité mendélienne entre un parent et un descendant était supérieur à 1%, le descendant a été considéré comme assigné à tort à ce parent. Pour les individus avec une mauvaise assignation (22 individus avec l'un des parents rejetés et 2 individus avec les deux parents du pedigree non validés), une recherche des bons parents a été

effectuée à l'aide des coefficients IBS calculés par Plink 1.9 (Purcell et al. 2007). Si un descendant mal assigné avait un IBS maximal avec les individus des cohortes parentales qui correspondait à une ressemblance génomique supérieure ou égale à 0,40 avec un seul des individus génotypés de la cohorte parentale, ce parent a alors été considéré comme son « vrai » parent. Pour 8 individus mal assignés, le parent a pu être identifié (IBS  $\approx$  0,50) ; pour les autres individus mal assignés, le parent a été déclaré inconnu dans le pedigree corrigé.

#### 2.4.2. Puce basse densité (LD)

Des génotypes LD ont été simulés à partir des génotypes de la puce MD. Différentes stratégies existent pour choisir les marqueurs soit aléatoirement, soit à partir de critères de MAF, de déséquilibre de liaison (DL), de répartition homogène (ou équi-distante) le long du génome ou de voisinage de QTLs (Moser et al. 2010; Ødegård et al. 2014; Tsai et al. 2015b; Vallejo et al. 2018)).

Dans notre étude, deux approches ont été choisies, toutes deux considérant un tri plus ou moins fort sur le DL car, dans le chapitre 2, nous avons montré que le DL se maintient à une longue distance ( $r^2 > 0,17$  à 1 Mb) dans la population de Viviers de Sarrance.

La première approche a donc été un simple tri sur le DL pour éliminer les marqueurs très redondants localement le long du génome en considérant une fenêtre coulissante de  $n$  kb. La seconde approche a été d'effectuer un choix préalable des SNPs basé sur une MAF minimum de 25% (au lieu du minimum initial de 5%) avant le tri sur le DL, afin de ne retenir que les marqueurs les plus informatifs.

Les génotypes LD ont été obtenus avec la commande (-- inder-parwise) de Plink 1.9 (Purcell et al. 2007) qui dans une fenêtre coulissante d'une longueur ( $f$ ) de  $n$  kb, retire les SNPs avec un DL (mesure  $r^2$ ) dépassant le seuil donné dans la fenêtre coulissante. Le choix des tailles  $f$  des fenêtres et des seuils de  $r^2$  s'est effectué de manière à atteindre les objectifs souhaités en nombre de SNPs à considérer dans des puces LD, soit environ 1 000 (1K), 3 000 (3K), 6 000 (6K) ou 10 000 (10K) SNPs. Le choix des paramètres des paramètres  $f$  et  $r^2$  et les statistiques de la distance entre les marqueurs des puces simulées sont résumés dans le [Tableau 5](#) ci-dessous.

Puce	N. SNP	distance entre les SNPs adjacents (en kb)					para. Plink	
		1 <sup>er</sup> quartile	médiane	moyenne	3 <sup>ème</sup> quartile	max	f	r <sup>2</sup>
<b>1K-05-VD</b>	1 009	314	944	1 808	2 383	31 114	1 800	0,140
<b>1K-25-VD</b>	1 006	394	1 053	1 775	2 165	31 070	1 800	0,195
<b>3K-05-VD</b>	3 042	126	326	608	749	15 043	600	0,275
<b>3K-25-VD</b>	3 043	138	341	600	719	12 850	600	0,395
<b>6K-05-VD</b>	5 996	68	169	309	367	11 782	300	0,403
<b>6K-25-VD</b>	6 016	69	172	305	367	12 850	300	0,575
<b>10K-05-VD</b>	10 021	39	104	185	227	11 176	180	0,542
<b>10K-25-VD</b>	10 017	40	104	184	221	12 850	180	0,800
<b>Puce 30K</b>	29 799	10	34	62	79	11 176	-	-

**Tableau 5** : Nombre de SNPs (N.SNP) et statistiques des distances (en kb) entre SNPs successifs pour chaque puce ainsi que les paramètres utilisés pour le choix des marqueurs (para. Plink).

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison ; f et r<sup>2</sup> sont les paramètres utilisés pour retenir le nombre souhaité de marqueurs à l'aide de Plink.

Par ailleurs, des génotypages LD de la population Viviers de Sarrance (VD) ont aussi été simulés à partir de choix de marqueurs effectués sur la population d'un autre sélectionneur (Aqualande) pour constituer des puces LD d'environ 3 000 (3K), 6 000 (6K) et 10 000 (10K) SNPs. Les mêmes approches que celles décrites ci-dessus ont été réalisées sur les génotypages de la population Aqualande (AQ), amenant à des choix de SNPs différents pour constituer les puces LD. Les caractéristiques de ces puces sont décrites dans le chapitre 4 ([Tableau 18](#)). Le nombre de marqueurs polymorphes avec une MAF > 5% dans la population VD pour ces puces établies à partir de la population AQ sont de :

- 2 591 et 2 720 respectivement pour les puces 3K-05-AQ et 3K-25-AQ
- 5 151 et 5 443 respectivement pour les puces 6K-05-AQ et 6K-25-AQ
- 8 584 et 8 974 respectivement pour les puces 10K-05-AQ et 10K-25-AQ

Les génotypages sur puces LD ont été sélectionnés à partir des génotypages MD avant imputation des génotypes manquants. Les génotypes des SNPs manquants sur puces LD ont alors été imputés comme décrit précédemment pour la puce MD et sans considérer une

population de référence génotypée en MD (ceci afin de simuler une situation où tous les génotypes disponibles ne seraient connus qu'en LD).

## 2.5. Modèles et méthodes d'analyse génétique

### 2.5.1. Modèle animal et évaluation BLUP sur pedigree

Pour prédire les valeurs génétiques (EBV) et estimer les paramètres génétiques, un modèle animal est considéré pour décrire les performances.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \quad (1)$$

où  $\mathbf{y}_i$  est le vecteur des performances pour le caractère  $i$ ,  $\mathbf{X}_i$  la matrice d'incidence pour le vecteur des effets fixes  $\boldsymbol{\beta}_i$ ,  $\mathbf{Z}$  la matrice d'incidence pour le vecteur  $\mathbf{u}_i$  des valeurs génétiques des animaux et  $\mathbf{e}_i$  le vecteur des résidus du modèle pour le caractère  $i$ .

L'effet cohorte est inclus dans la modélisation des effets fixes de tous les caractères. L'effet de la présence d'œufs surmatures (codé en 0/1) est significatif et donc inclus dans l'analyse des caractères FW, PW, SW, EN et ED. La semaine de ponte est considérée comme une covariable nichée dans la cohorte pour FW, PW, SW, EN, EW et ED.

Pour l'étude de l'efficacité de la sélection, les caractères SW et EN ont été ajustés intra-cohorte pour le poids FW, car l'objectif de sélection est d'améliorer les performances SW et EN à poids de femelle constant.

La résolution des équations de ce modèle est basée sur la méthodologie BLUP et l'information du pedigree permettant de construire la matrice de parenté A (cf. chapitre 1). Le pedigree est composé de 15 086 individus remontant les 9 générations de sélection de la lignée VD.

Pour estimer toutes les composantes de la variance, huit modèles multi-caractères ont été effectués en considérant simultanément quatre caractères avec soit FW ou PW représentant le poids de la femelle (BW) : [BW, SD, EW, SW], [BW, SD, EW, EN], [BW, SD, ED, EN] et [BW, SD, ED, SW]. Ces groupes ont été choisis pour éviter les variables trop corrélées empêchant des problèmes de convergence. Pour un BW donné, les résultats pour tous les paramètres génétiques ou phénotypes étaient les moyennes des 4 estimations. Cependant, les corrélations entre FW et PW, SW et EN, EW et ED ont été estimées dans le cadre de modèles

bi-caractères pour chaque paire en raison de problèmes de convergence liés aux corrélations très élevées entre les caractères.

Les EBV et les composantes de la variance sont obtenues à l'aide des packages BLUPf90 et AIREMLf90 (Misztal et al. 2002, 2018).

### 2.5.2. Modèle animal et évaluation GBLUP basée sur la parenté génomique

L'évaluation GBLUP est très similaire à une évaluation BLUP, la matrice de parenté sur pedigree (A) étant simplement remplacée par une matrice de parenté génomique (G) (chapitre 1). Les valeurs génomiques prédites (GEBV) sont obtenues à l'aide du logiciel BLUPf90 (Misztal et al. 2018).

Les GEBV permettent de calculer les effets estimés des SNPs grâce au package postGSF10 (Aguilar et al. 2014) à l'aide de la formule suivante (Wang et al. 2012) :

$$\hat{\mathbf{a}} = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{\mathbf{u}}_g \quad (2)$$

avec **D** la matrice diagonale des poids associés aux effets de SNPs (D=I en GBLUP), **Z** la matrice d'incidence qui relie les génotypes aux SNPs des individus génotypés au vecteur **â** des effets estimés des SNPs et **û<sub>g</sub>** le vecteur des GEBV des individus génotypés ( $\hat{\mathbf{u}}_g = \mathbf{Z}\hat{\mathbf{a}}$ ).

Ensuite, on peut estimer les GEBV d'individus génotypés non inclus dans l'évaluation à partir des effets estimés des SNPs comme la somme des effets de leurs génotypes aux SNPs.

Le modèle ssGBLUP est une évolution du modèle GBLUP qui permet de considérer des individus seulement phénotypés dans l'évaluation génomique pour espérer augmenter la précision des GEBV (chapitre 1). Celui-ci permet d'inclure via la matrice de parenté A l'information des phénotypes d'individus non génotypés, mais apparentés à la population de référence.

Pour ces deux modèles génomiques, les mêmes effets fixes que pour le modèle BLUP ont été considérés.

Un modèle tri-caractère a été aussi utilisé afin de quantifier les augmentations de précision des (G)EBV pour des caractères corrélés. Les combinaisons de caractères suivantes ont été testées : FW-SW-EW, FW-SW-ED, FW-EN-EW, FW-EN-ED. Aucun gain de précision n'a été observé en GBLUP et un très léger gain seulement en BLUP (résultats non présentés). Tous les individus génotypés ayant des phénotypes pour tous les caractères et les corrélations restant modérées entre caractères, cette absence de gain de précision était attendue.

Afin d'identifier des QTLs sur l'ensemble des caractères étudiés, une analyse d'association pangénomique basée sur le GBLUP a été réalisée, à l'aide du module postGSf90 (Aguilar et al. 2014) de BLUPf90. Une région du génome a été considérée comme un QTL quand un SNP de la région avait un effet significatif à un seuil de 1% à l'échelle du génome (après correction de Bonferroni), correspondant à un  $\log(p\_value) > 6,5$ . Le QTL a été considéré comme seulement suggestif pour un  $\log(p\_value) > 5$  (correspond à un seuil de significativité à 1% à l'échelle du chromosome après correction de Bonferroni). L'intervalle de confiance d'un QTL a été estimé par la méthode du drop-off (Li 2011) en considérant une limite inférieure du  $\log(p\_value)$  correspondant à une valeur réduite de -1,5 à partir de la valeur maximale atteinte au SNP pic du QTL. Tous les SNPs dans le voisinage du pic SNP (avec une fenêtre coulissante de 1 Mb) avec un  $\log(p\_value)$  au-dessus de la limite inférieure de drop-off sont inclus dans l'intervalle de confiance du QTL.

La variance expliquée par un SNP est calculée comme :

$$V_{snp_i} = 2p_iq_ia_i^2 \quad (3)$$

avec  $p_i$  et  $q_i$  les fréquences des allèles pour le SNP<sub>i</sub>,  $a_i$  l'effet additif du SNP<sub>i</sub>.

La part de variance expliquée par un QTL a été calculée comme la somme des parts de variance expliquée par chaque SNP compris dans l'intervalle de confiance.

### 2.5.3. Modèle BayesC $\pi$ et sélection de variables

Une approche bayésienne de sélection de variables a également été utilisée pour mieux estimer les effets des SNPs et leurs variances en considérant le problème statistique du nombre très élevé de paramètres (les effets de SNPs) à estimer par rapport au nombre de données disponibles (Habier et al. 2011). L'approche BayesC $\pi$  consiste à considérer un modèle de mélange avec une proportion  $\pi$  de SNPs ayant des effets non nuls (distribués selon une loi normale de

moyenne nulle et variance suivant une loi de Chi2-inverse) et une proportion  $(1-\pi)$  des SNPs sans effet à chaque cycle d'estimation des paramètres par méthode de Monte-Carlo par chaîne de Markov (avec algorithme de Metropolis-Hastings). Le modèle d'analyse des performances est le suivant :

$$y_i = \mu + \mathbf{X}\beta_i + \sum_{j=1}^n \delta_j z_{ij} a_j + e_i \quad (4),$$

avec  $y_i$  la performance de l'animal  $i$ ,  $\mu$  la moyenne du caractère,  $\mathbf{X}$  la matrice d'incidence pour les effets fixes  $\beta$  d'une performance,  $n$  le nombre de SNP,  $z_{ij}$  le génotype (codé 0 et 2 pour les homozygotes et 1 pour l'hétérozygote) au locus  $j$  pour animal  $i$ ,  $a_j$  l'effet du SNP  $j$ ,  $e_i$  le résidu,  $\delta_j$  est une indicatrice : à l'itération  $k$ ,  $\delta_j=1$  si le SNP  $j$  est inclus dans le modèle (probabilité  $\pi$ ) et  $\delta_j=0$  s'il n'est pas inclus (probabilité  $1-\pi$ ).

L'estimation du paramètre  $\pi$  se fait sous l'hypothèse d'une distribution bêta  $B(\alpha, \beta)$  de paramètres  $\alpha=300$  et  $\beta=29800$ , pour considérer en espérance que 300 SNPs sont retenus comme ayant un effet non-nul parmi les 29 800 SNPs inclus dans l'analyse. La valeur moyenne de  $\pi$  a ainsi été considérée proche de 1%. Pour établir ce choix, différents tests ont été effectués afin d'effectuer un compromis raisonnable entre le nombre limité de données disponibles (environ 1 300) et le fait que les caractères semblaient très polygéniques. Nous avons considéré qu'estimer les effets d'environ 300 SNPs simultanément nous permettait une analyse pertinente dans ce contexte.

Pour chaque analyse, 100 000 cycles de MCMC sont effectués, avec une chauffe de 5 000 cycles. Les résultats sont sauvegardés tous les 20 cycles pour estimer les paramètres. En particulier, la variance expliquée par chaque SNP est calculée à chaque cycle retenu selon la formule de l'équation 3 et le résultat final est la moyenne des résultats intermédiaires. Les analyses ont été effectuées avec le logiciel BESSiE (Boerner and Tier 2016).

Afin de confirmer les estimations obtenues, 4 chaînes différentes de 100 000 cycles ont été lancées pour tous les caractères. La convergence des estimations des variances génétique et résiduelle est confirmée par observation visuelle ainsi que par le calcul de corrélation entre les effets des SNPs ( $cor > 0,99$ ) entre différentes chaînes de MCMC.

L'importance de l'association entre un SNP et le phénotype est appréciée par le Bayes Factor (BF), calculé avec la formule suivante (Schurink et al. 2012) :

$$BF = \left( \frac{p_i}{1-p_i} \right) / \left( \frac{\pi}{1-\pi} \right), \quad (5)$$

avec  $p_i$  la probabilité d'inclusion du SNP  $i$  dans le modèle. Kass et Raftery (1995) propose de transformer le BF en  $\log BF = 2 * \ln(BF)$  pour avoir des valeurs dans l'échelle usuelle de valeurs observées dans les tests de déviance et ratio de vraisemblance. Selon les considérations issus des travaux de Michenet et al. (2016), un seuil de  $\log BF \geq 6$  est considéré pour définir un QTL à évidence forte et son intervalle de crédibilité est calculé autour du SNP au  $\log BF$  maximal en intégrant dans la région QTL tous les SNPs avec un  $\log BF \geq 3$  dans une fenêtre coulissante de 1 Mb à partir du pic SNP.

## 2.6. Comparaison de l'efficacité de la sélection selon différents scénarios

Pour estimer l'efficacité de la sélection sur pedigree (BLUP) ou génomique (GBLUP ou ssGBLUP), nous avons effectué une étude de validation croisée. Pour cela, nous avons divisé notre jeu de données en deux populations : une population d'apprentissage (PA) et une population de validation (PV). On utilise les génotypes et les phénotypes (ou seulement les phénotypes en BLUP) de PA pour établir les équations de prédiction. Ensuite, on estime les valeurs génétiques (EBV ou GEBV) des animaux de PV.

### 2.6.1. Les scénarios

Afin de tester l'importance de l'effet de la taille de la population d'apprentissage, deux scénarios ont été testés. On note PA- le scénario à petite population d'apprentissage (672 individus) et PA+ le scénario à plus grande population d'apprentissage (1077 individus). Dans la population de référence totale, 40 échantillons de 269 individus ont été tirés aléatoirement dans l'ensemble des cohortes  $G_{14}$  et  $F_{15}$  pour constituer les 40 populations de validation (PV) communes aux scénarios PA- et PA+ nécessaires à la validation croisée effectuée sur 40 simulations. Pour chaque simulation, les 1077 individus restants de la population de référence ont constitué la grande population d'apprentissage (PA+) et 672 individus ont été tirés aléatoirement parmi cette dernière pour constituer la petite population d'apprentissage (PA-). Pour étudier l'impact du degré d'apparentement entre PA et PV sur l'efficacité de la sélection, deux autres scénarios ont été testés avec l'une des deux cohortes considérées comme PA et

l'autre comme PV (et *vice-versa*). On note PAg le scénario où la cohorte G14 est considérée comme PA et la cohorte F15 comme PV et PAF le scénario où la cohorte F15 devient PA et la cohorte G14 PV. Les scénarios sont récapitulés dans le tableau ci-dessous (Tableau 6).

Effet Scenario	Taille de la pop d'apprentissage PA		Apparement entre PA et PV	
	PA+	PA-	PAg	PAf
Population de validation (PV)	269 poissons choisis aléatoirement dans G <sub>14</sub> et F <sub>15</sub>	Mêmes 269 poissons que le scénario PA+	620 poissons de la cohorte F <sub>15</sub>	726 poissons de la cohorte G <sub>14</sub>
Population d'apprentissage (PA)	1077 poissons restants de G <sub>14</sub> et F <sub>15</sub>	672 poissons choisis aléatoirement dans les 1077 poissons restants de G <sub>14</sub> et F <sub>15</sub>	726 poissons de la cohorte G <sub>14</sub>	620 poissons de la cohorte F <sub>15</sub>

**Tableau 6:** Récapitulatif du choix et du nombre d'individus dans la population d'apprentissage et validation pour chaque scénario.

## 2.7. Validation croisée

La précision de la sélection est estimée à partir de la corrélation entre les (G)EBV des individus de PV et leurs phénotypes corrigés ( $y^*$ ) pour les effets fixes d'environnement, qui sont alors considérés comme un « proxy » des valeurs génétiques vraies en standardisant par la racine carrée de l'héritabilité  $h^2$  du caractère. La précision de la sélection ( $r$ ) est ainsi approchée par la formule :

$$r = \frac{\text{cor}(y^*, (G)EBV)}{\sqrt{h^2}} \quad (6)$$

La pente ( $b$ ) de la droite de régression des phénotypes  $y^*$  sur les (G)EBV pour savoir si les (G)EBV surestiment ou sous-estiment la variabilité des valeurs génétiques vraies est calculée comme :

$$b = \text{cor}(y^*, (G)EBV) * \left( \frac{\sigma_{y^*}}{\sigma_{(G)EBV}} \right) \quad (7)$$

Ce coefficient  $b$  est aussi appelé coefficient de dispersion ou inflation. Pour qu'une évaluation soit sans biais,  $b$  doit être égal à 1. Si  $b < 1$ , la variabilité des (G)EBV est surestimée (surdispersion ou inflation) et si  $b > 1$ , la variabilité est sous-estimée (sous-dispersion ou déflation).

Dans le cas des scénarios étudiés en échantillonnant au hasard des individus de la population de référence, la moyenne et l'écart type d'échantillonnage des coefficients  $r$  et  $b$  ont été calculés à partir des 40 simulations effectuées par scénario.

### 3. Résultats & discussion

#### 3.1. Paramètres génétiques

##### 3.1.1. Héritabilité des caractères

L'héritabilité ( $h^2$ ) des caractères (Tableau 7) varie entre 0,27 (EW) et 0,46 (ED), ce qui correspond donc à des valeurs d'héritabilité moyennes permettant une sélection de chacun des caractères considérés.

Trait	SD	FW	PW	SW*	EN*	EW	ED
SD	<b>0,27 (0,05)</b>	0,12 (0,15)	0,03 (0,15)	0,34 (0,15)	0,22 (0,16)	0,46 (0,15)	0,51 (0,12)
FW	0,04 (0,04)	<b>0,30 (0,06)</b>	0,99 (0,01)	0,12 (0,15)	0,13 (0,16)	-0,01 (0,16)	-0,01 (0,15)
PW	0,04 (0,04)	0,96 (0,01)	<b>0,33 (0,06)</b>	-0,06 (0,15)	-0,02 (0,16)	-0,10 (0,16)	-0,12 (0,14)
SW*	0,05 (0,04)	0,34 (0,03)	0,17 (0,03)	<b>0,32 (0,06)</b>	0,86 (0,05)	0,49 (0,13)	0,59 (0,10)
EN*	0,10 (0,04)	0,31 (0,03)	0,17 (0,03)	0,83 (0,01)	<b>0,24 (0,05)</b>	0,01 (0,17)	0,10 (0,15)
EW	0,06 (0,04)	0,09 (0,03)	0,03 (0,03)	0,33 (0,03)	-0,18 (0,03)	<b>0,27 (0,06)</b>	0,99 (0,02)
ED	0,13 (0,04)	0,12 (0,03)	0,05 (0,03)	0,38 (0,03)	0,01 (0,03)	0,70 (0,01)	<b>0,44 (0,07)</b>

**Tableau 7 :** héritabilité des caractères sur la diagonale (en gras), corrélations génétiques entre les caractères au-dessus de la diagonale, corrélations phénotypiques en dessous de la diagonale. Les erreurs standard sont indiquées entre parenthèses. SD : semaine de ponte, FW : poids de la femelle pleine, PW : poids de la femelle après la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf. \* si SW et EN sont ajustés pour FW, leurs corrélations génétiques estimées avec FW deviennent -0,26 (0,14) et -0,17 (0,16) respectivement et leurs corrélations phénotypiques avec FW deviennent -0,05 (0,04) et 0,06 (0,04) respectivement.

Il est à noter que si les estimations présentées dans le Tableau 7 concernent les seules données collectées sur les cohortes  $G_{14}$  et  $F_{15}$ , certains phénotypes (FW, PW, SW, ED, SD) étaient aussi disponibles pour la cohorte parentale  $E_{12}$ . Le calcul des paramètres génétiques incluant aussi les données de cette cohorte parentale a abouti à des estimations très similaires de paramètres

génétiques (résultats non présentés), seul SW ayant alors une héritabilité estimée un peu plus élevée (0,40 au lieu de 0,32).

Nous avons comparé nos estimations à celles décrites pour les salmonidés dans la littérature depuis la fin des années 1990 et qui sont basées sur un modèle BLUP animal et une estimation par REML des composantes de variance. En effet, les publications antérieures concernent des analyses de variance dans des familles de pleins ou demi-germains où les estimations peuvent être potentiellement très biaisées. Par ailleurs, pour la truite arc-en-ciel, seules des estimations obtenues sur les mêmes lignées expérimentales de l'Université de Californie sont disponibles à ce jour pour les caractères de reproduction femelle (Gall 1975; Gall and Gross 1978; Gall and Huang 1988; Su et al. 1997, 2002). Nos résultats concernant les caractères de reproduction seront donc seulement comparés aux travaux de Su et al. (1997) portant sur les 3 lignées de truite arc-en-ciel qui proviennent de la même population de l'université de Californie ainsi qu'aux résultats obtenus sur le saumon coho par Gall et Neira (2004).

Le **Tableau 8** fait la synthèse des héritabilités estimées dans les deux études antérieures à la nôtre (Su et al. 1997; Gall and Neira 2004).

Traits	Truite arc-en-ciel Su et al (1997)				Saumon Coho Gall et Neira (2004)	Notre étude
	Lignée contrôle	Lignée sélectionnée sur EN	Lignée sélectionnée sur le poids à 1 an	Pool des 3 lignées		
SD	0,49	0,84	0,87	0,65	0,24	0,27
PW	0,09	0,26	0,14	0,14	0,29	0,33
SW	0,45*	0,60*	0,42*	0,52*	0,39	0,32
EN	0,32	0,57	0,85	0,55	0,42	0,24
EW	0,70*	0,58*	0,58*	0,60*	0,32	0,27

**Tableau 8 :** héritabilités estimées pour les caractères de reproduction femelle de 3 lignées de truite arc-en-ciel issues de la même population de base (Su et al. 1997) et d'une population de saumon coho (Gall and Neira 2004).

SD : semaine de ponte, PW : poids de la femelle après la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf. \* les traits SW (mL) et EN ( $\mu$ L) ont été mesurés en volume dans l'étude de Su et al. (1997).

En ce qui concerne le poids de la femelle reproductrice, les héritabilités estimées pour FW (0,30) et PW (0,33) sont dans l'intervalle de valeurs 0,27 à 0,43 mentionné dans la littérature récente pour des poids de femelle mesurées entre 13 et 25 mois (Gonzalez-Pena et al. 2016; Leeds et al. 2016). Dans l'étude de Su et al. (1997), les héritabilités estimées pour PW sont plus faibles variant de 0,09 à 0,26 selon la lignée considérée de truite arc-en-ciel (**Tableau 8**). Ils estiment un effet significatif des familles de pleines sœurs, expliquant selon la lignée de 10% à

22% de la variance phénotypique. Dans notre étude, aucun effet maternel significatif n'est détecté malgré le mélange au stade oeillé des œufs de toutes les familles et de leur élevage commun en bassin. Ce résultat va à l'encontre de l'effet rapporté à 60 g dans une autre lignée française (Haffray et al. 2012b) et interroge sur les préconisations de maîtrise de cet effet tel que proposé antérieurement (Chevassus et al. 2004).

Concernant la semaine de la ponte, notre héritabilité modérée se rapproche de l'héritabilité estimée chez le saumon coho (Gall and Neira 2004), mais est très inférieure aux héritabilités estimées chez la truite arc-en-ciel par Su et al. (1997) (Tableau 8). Une explication de notre faible héritabilité pourrait être celle aussi émise par Gall et Neira (2004), à savoir une sous-estimation liée au fait que les données ne sont collectées que pour le groupe des femelles pondant le plus précocement, les pontes tardives n'étant pas recueillies. D'autres hypothèses pourraient aussi être envisagées comme celle d'une interaction avec l'environnement d'élevage, ici sur eau de source à température constante et très minéralisée.

De manière générale, toutes nos héritabilités de caractères de reproduction sont bien plus proches de celles estimées pour le saumon coho (Gall et Neira 2004) que pour la truite arc-en-ciel (Su et al. 1997). Seule l'héritabilité de EN apparaît faible dans notre étude aux regards de toutes les études antérieures.

### 3.1.2. Corrélations génétiques et phénotypiques

Nos estimations des corrélations pour les deux mesures du poids corporel des femelles avec (FW) ou sans (PW) prise en compte du poids de la ponte et du liquide coelomique ont montré que les deux mesures décrivaient essentiellement le poids propre des femelles puisque nous avons estimé une corrélation phénotypique entre FW et PW de 0,96 avec une corrélation génétique proche de un.

En ce qui concerne les corrélations entre les traits de reproduction, il faut d'abord mentionner que génétiquement parlant, EN et SW d'une part, et EW et ED d'autre part, correspondaient à peu près aux mêmes traits avec des corrélations génétiques ( $r_g$ ) estimées à 0,86 ( $\pm 0,05$ ) et 0,99 ( $\pm 0,02$ ), respectivement. Toutes ces corrélations élevées sont en partie dues aux auto-corrélations provenant des méthodes de dérivation de ces caractères à partir de mesures

communes de composants biologiques, comme le montrent les corrélations environnementales élevées observées entre ces paires de caractères (0,82 entre EN et SW et 0,55 entre EW et ED dans notre étude). Une forte corrélation entre EN et SW est aussi observée chez la truite arc-en-ciel et chez le saumon coho (Tableau 9).

Etudes		Caractères			
		SD	EW	EN	SW
Gall et Neira (2004)	PW	0,20	0,37	0,32	0,56
	Saumon coho	SW	0,01	0,15	0,79
	EN	0,19	-0,63	-	0,79
<i>Su et al (1997)*</i>	PW	0,73	0,51	0,47	0,67
	<i>Truite arc-en-ciel</i>	SW	0,54	0,48	0,81
	EN	0,25	-0,10	-	0,81

**Tableau 9 :** Corrélations génétiques estimées pour les caractères de reproduction femelle de 3 lignées de truite arc-en-ciel issues de la même population de base (Su et al. 1997) et d'une population de saumon coho (Gall and Neira 2004).

SD : semaine de ponte, PW : poids de la femelle après la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf. \* les traits SW (mL) et EN ( $\mu$ L) ont été mesurés en volume dans l'étude de Su et al. (1997).

Bien que EN et SW soient fortement corrélés dans notre étude, ces deux traits ne se sont pas comportés de manière similaire avec ED et EW. Au contraire, ED et EW pouvaient être considérés comme deux mesures différentes d'un trait biologique unique de la taille des œufs, ces deux traits étant associés de manière très similaire à tous les autres traits de l'analyse. SW était significativement et positivement associé aux caractéristiques de la taille de l'œuf (rg entre 0,5-0,6), tandis que EN était génétiquement non corrélé aux caractéristiques de la taille de l'œuf (rg entre 0,01-0,10) dans notre étude. Cependant Gall et al. (2004) ont estimé une corrélation génétique négative importante (-0,63) entre EN et EW pour le saumon coho et Su et al. (1997) ont estimé une corrélation génétique de -0,11 chez la truite arc-en-ciel. Par conséquent, la question est toujours en suspens de savoir si la sélection en fonction de la taille des œufs peut ou non entraîner une diminution du nombre d'œufs chez les salmonidés. Là encore une interaction entre déterminisme des caractères biologiques et environnement final de maturation ne peut aussi être exclue, amenant donc à considérer ces résultats dans les conditions de cet élevage.

Les caractères EN et SW étaient tous deux faiblement corrélés génétiquement à SD (0,22 et 0,34 respectivement). Ceci a également été rapportée par Su et al (1997) et Gall et Neira (2004)

entre EN et SD ([Tableau 9](#)). Cependant, chez le saumon Coho, la corrélation génétique estimée entre SW et SD était nulle alors qu'elle est de 0,54 dans l'étude de Su et al. (1997) ([Tableau 9](#)). Des corrélations génétiques positives intermédiaires ont également été estimées entre les traits de taille d'œuf et SD ( $\approx 0,5$ ). Nos estimations étaient cohérentes avec l'estimation (0,51) de l'étude de Su et al. (1997), tandis que Gall et Neira (2004) ont signalé une corrélation génétique plus faible (0,16) chez le saumon coho.

En ce qui concerne les corrélations entre SD et PW (ou FW), nous n'avons observé aucune association phénotypique ou génétique significative dans notre étude. Nos résultats sont en contradiction avec la corrélation génétique positive significative rapportée par Su et al. (1997) mais très cohérente avec celle estimée chez le saumon coho ([Tableau 9](#)). Cela peut être dû au fait que seules les femelles qui ont pondu le plus tôt ont été incluses dans notre étude, comme dans celle de Gall et Neira (2004).

D'un point de vue génétique, les caractéristiques de la taille des œufs (EW et ED) et de la quantité d'œufs (EN et SW) n'ont pas non plus été corrélées de manière significative avec le poids corporel des femelles. Le fait de considérer FW comme une covariable dans la modélisation de EN et SW a conduit à des corrélations génétiques négatives entre FW et les EN\* et SW\* ajustés, tout en faisant régresser vers 0 les corrélations phénotypiques correspondantes.

Bien que le poids des femelles ne soit pas significativement corrélé aux caractéristiques de quantité (EN, SW) ou de taille des œufs (EW, ED) dans notre étude, des corrélations génétiques positives significatives ont été trouvées pour PW avec EN, la taille et le volume des œufs dans des lignées américaines de truites arc-en-ciel et des corrélations plus modérées chez le saumon coho ([Tableau 9](#)). Nos résultats ne sont donc pas cohérents avec les deux études précédentes sur les espèces de salmonidés concernant les associations génétiques entre le poids corporel des femelles et les caractéristiques de production d'œufs. Ce manque de similarité entre les résultats interroge et conduit à ne pouvoir généraliser nos conclusions.

Des études antérieures ont indiqué que la sélection pour le poids corporel à tout âge (278 à 740 jours) entraînerait une réponse corrélée pour le poids corporel à d'autres âges, y compris l'âge à l'abattage pour les femelles (Crandell and Gall 1993) et que les corrélations génétiques entre le poids des juvéniles, le poids des femelles après le frai et la taille des œufs, le nombre d'œufs et le volume des œufs étaient modérées, mais positives chez la truite arc-en-ciel ( $\approx 0,30$ ) (Su et

al. 2002). Ces résultats ont conduit à l'hypothèse que les programmes de sélection conçus pour augmenter la taille du corps entraîneraient une amélioration des caractéristiques de production d'œufs (Su et al. 2002). Cependant, nos résultats ne valident pas cette hypothèse dans cette population puisque le poids corporel des femelles ne semble pas être génétiquement corrélé à l'un des traits de reproduction (SD, SW, EN, EW ou ED) dans la lignée et l'environnement étudiés. Ces résultats sont acquis pour cette population dans cet environnement d'élevage. Au regard des écarts avec les autres travaux, il pourrait être judicieux d'étendre la caractérisation des corrélations entre caractères à d'autres lignées. D'une façon plus appliquée, ces résultats soulignent l'importance d'envisager la sélection directement sur les caractères de production d'œufs afin de les améliorer dans les programmes d'élevage de truites arc-en-ciel.

### 3.2. Identification de QTLs et gènes candidats

L'architecture génétique des caractères de reproduction femelle s'avère très polygénique (Figure 15) avec seulement 1 à 4 QTL expliquant entre 1% à 3% de la variance génétique (Tableau 10), le reste de la variance génétique étant due à de nombreux effets génétiques non détectables dans le dispositif étudié. Un seul QTL explique plus de 3% de la variance génétique pour le caractère SD (Tableau 10). Les deux types d'analyses GWAS conduites amènent à des résultats très cohérents. En effet, 10 QTLs sur les 12 ayant une p-value en GBLUP significative à 1% au niveau du chromosome correspondent à des QTLs à évidence forte indiquée par le Bayes Factor. Dans la majorité des cas, les pics SNPs sont les mêmes dans les deux analyses. Une moindre cohérence entre les résultats est observée pour les caractères EW et ED, en effet sur 9 QTLs, 2 seulement sont commun entre les deux types d'analyses GWAS.

L'approche bayésienne a permis de détecter 17 QTLs expliquant chacun plus de 1% de la variance génétique mais seulement 9 de ces QTLs ont une évidence très forte ( $\log\text{BF} > 8$ ).

Les intervalles de confiance définis pour ces QTLs sont relativement larges, ce qui ne permet pas une identification certaine des gènes causatifs des variations de performance observées. Cependant, nous pouvons proposer 5 gènes candidats (3 pour SD, 1 pour ED et 1 partagé pour EN et SW) sur la base de l'information fonctionnelle fournie dans la base de gènes humains GeneCard® (Rebhan et al. 1997, 1998) et de la description associée de phénotypes de souris ou

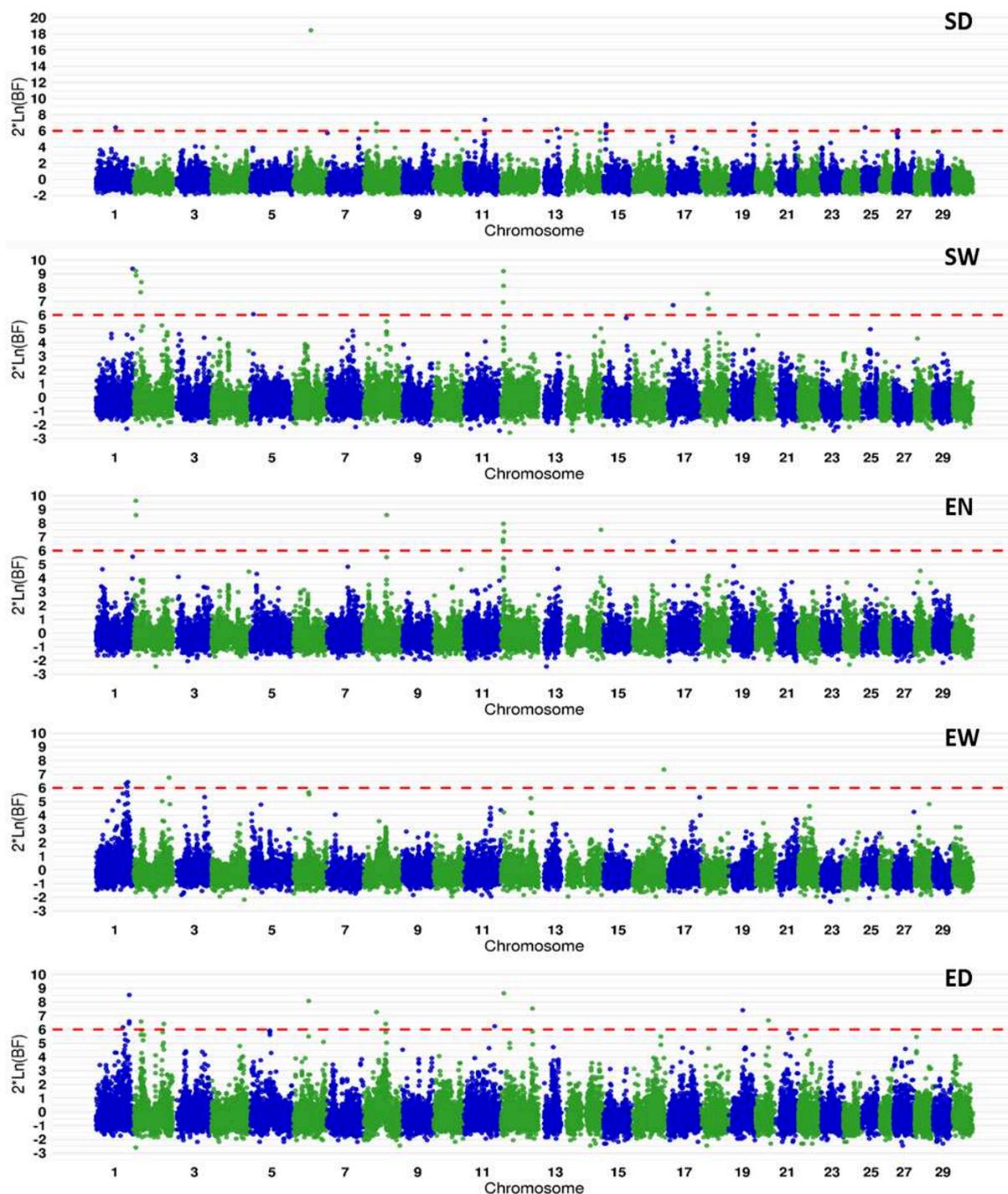
vers mutants. Aucun phénotype n'est décrit pour ces gènes candidats dans la base de données zebrafish ZFIN (Sprague et al. 2008).

Trait	SNP peak			GBLUP		BayesC $\pi$		
	Omy	Name SNP	Position (pb)	Log(P-value)	Intervalle de confiance (Mb)	LogBF	Intervalle de crédibilité (Mb)	Var* (%)
SD	6	Affx-88920061	38 611 684	11,5	38,61 - 38,61	18,4	38,61 - 38,61	7,2
SD	11	Affx-88936987	44 670 771	-	-	7,4	43,17 - 45,32	1,9
SD	15	Affx-88942247	5 640 348	-	-	6,8	4,86 - 5,64	1,2
SD	27	Affx-88907414	8 854 196	-	-	6,1	7,04 - 8,95	1,7
SW	1	Affx-88918810	79 053 674	-	-	9,4	77,68 - 79,05	1,3
SW	2	Affx-88951720	5 801 797	6,3	5,80 - 6,18	9,2	5,80 - 6,18	2,9
SW	2	Affx-88927743	17 660 020	6,1	16,29 - 17,66	8,4	16,29 - 17,66	1,9
SW	12	Affx-88950456	5 349 425	6,9	4,56 - 6,22	9,2	4,56 - 6,80	2,5
EN	2	Affx-88951720	5 801 797	6,5	5,80 - 6,18	9,6	5,80 - 6,18	3,2
EN	8	Affx-88953704	50 018 409	5,9	49,64 - 50,02	8,6	49,64 - 50,02	1,4
EN	12	Affx-88950456	5 349 425	6,2	4,23 - 6,80	8,0	4,56 - 6,80	2,7
EW	1	Affx-88932157	48 339 143	5,4	48,34 - 48,60	-	-	-
EW	1	Affx-88950822	64 632 546	7,0	60,50 - 64,63	6,3	61,59 - 64,63	1,2
EW	1	Affx-88915911	67 153 692	7,2	66,61 - 68,88	-	-	-
EW	1	Affx-88957820	68 664 516	-	-	6,4	66,54 - 68,88	2,0
ED	1	Affx-88941218	68 345 854	6,0	68,35 - 68,66	-	-	-
ED	1	Affx-88942198	71 813 379	7,1	70,85 - 71,81	8,5	70,85 - 71,81	1,9
ED	2	Affx-88930958	17 234 307	-	-	6,6	16,14 - 18,25	1,5
ED	8	Affx-88908356	47 852 976	-	-	6,4	47,84 - 49,70	1,6
ED	12	Affx-88930328	68 199 037	-	-	7,5	67,33 - 68,79	1,1

**Tableau 10 :** Statistiques descriptives des QTL détectés pour les caractères de reproduction femelle en GWAS par approches GBLUP et BayesC $\pi$ .

SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf. Var\* : proportion de la variance génétique expliquée par l'ensemble des SNPs inclus dans l'intervalle de crédibilité du QTL ; SNP speak : SNP avec le Log(p-value) ou LogBF le plus élevés. – zone non identifié avec le modèle ou SNP peak différent

La plupart des études récentes sur les QTLs chez la truite arc-en-ciel se sont concentrées sur les caractères de croissance et de découpe (Gonzalez-Pena et al. 2016; Neto et al. 2019). Le seul SNP significatif que nous avons détecté pour le poids de la femelle ne se situe à proximité d'aucune région QTL signalée antérieurement pour le poids corporel de la truite. Ce SNP n'est pas mentionné dans le [Tableau 10](#), car l'association détectée a été considérée comme suspecte. En effet, l'association n'est pas détectée par GBLUP et l'analyse bayésienne ne pointait qu'un unique SNP (logBF > 6) dans une zone intergénique entre les gènes CROCC2 et SLCO1F1, deux gènes qui ne se sont pas annotés comme jouant un rôle dans la fonction de la croissance.



**Figure 15 :** Manhattan plot des QTLs détectés en analyse BayesC $\pi$  pour les caractères de reproduction femelle.

La ligne pointillée rouge correspond au seuil  $\log_{10}BF > 6$  pour définir l'évidence d'un QTL. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

Dans notre étude, le SNP le plus significatif ( $p$ -valeur  $< 10^{-6}$  au niveau du génome) a été détecté pour SD sur Omy6. Il correspond au seul QTL significatif pour SD en analyse GBLUP. Ce seul SNP Affx-88920061 explique plus de 7% de la variance génétique en analyse BayesC $\pi$ . Il est

situé dans le gène importin-11 (alias Ran-Binding Protein 11) qui joue un rôle de récepteur dans l'importation de protéines nucléaires. Les phénotypes observés dans une lignée de souris MGI (ID MGI: 5617259) mutante pour ce gène peuvent éventuellement aider à comprendre les effets qu'un allèle variant pourrait avoir sur la date de ponte de la truite arc-en-ciel. L'homozygote double mutant correspond à un phénotype létal avant le sevrage du souriceau et l'hétérozygote présente des niveaux faibles de fer et de glucose dans le sang. De plus, le mâle hétérozygote a une morphologie anormale de l'œil (cristallin) qui pourrait expliquer des variations de réception de la lumière et donc de date de ponte si ce phénotype était observé chez la truite arc-en-ciel. Cependant, cette hypothèse est incertaine et nous pouvons tout aussi bien nous demander si l'association n'est pas détectée en raison d'erreurs de génotypage. En effet, un taux très élevé (13,5%) d'erreurs mendéliennes est observé pour ce SNP lorsque l'on compare les génotypes des descendants et de leurs parents alors que la très grande majorité des SNPs n'ont aucune erreur et que parmi les 4067 SNPs pour lesquels des erreurs mendéliennes sont observées, la médiane de ce taux est inférieure à 3‰.

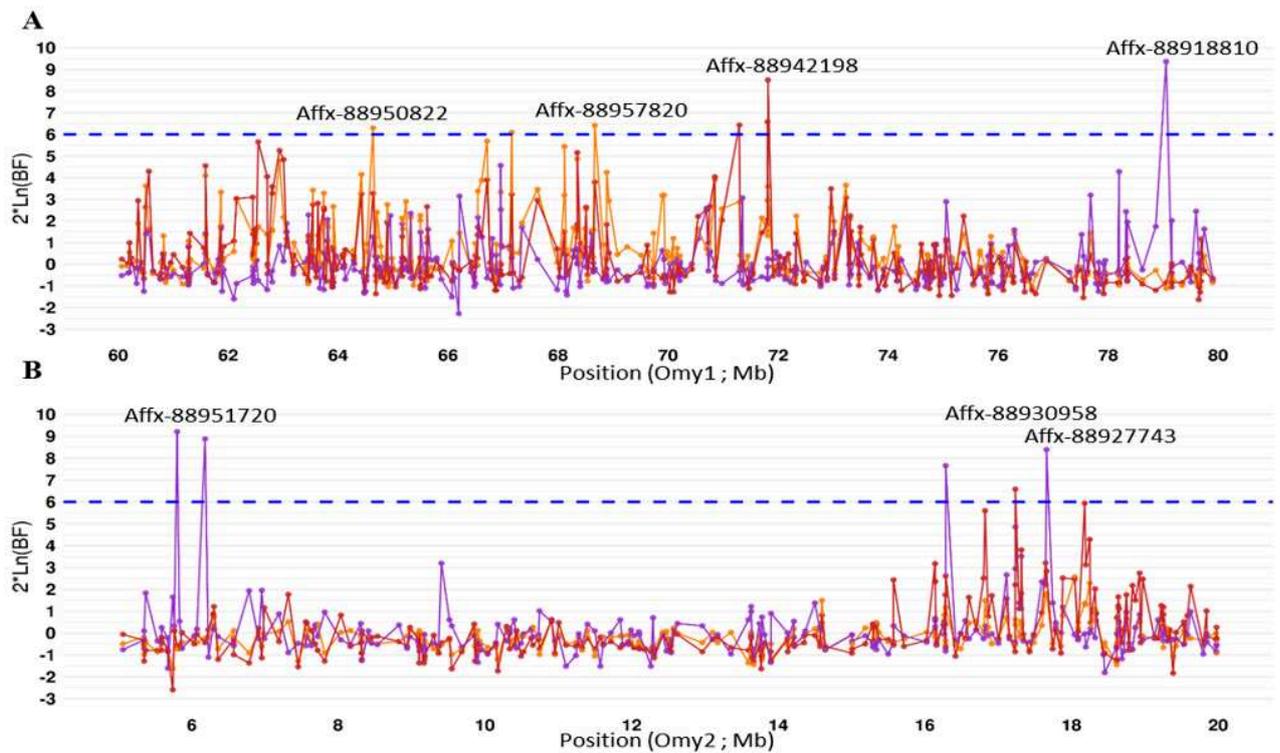
Bien que l'architecture génétique sous-jacente à la date de ponte soit encore largement inconnue, divers QTLs ont été identifiés pour la date de la ponte ou pour l'âge de la maturation sexuelle chez les salmonidés (Sakamoto et al. 1999; O'Malley et al. 2003; Leder et al. 2006; Moghadam et al. 2007; Haidle et al. 2008; Küttner et al. 2011; Gutierrez et al. 2014). En raison de la faible densité de marqueurs de ces études antérieures et de la non-standardisation des noms des groupes de liaison dans les études les plus anciennes, il est difficile d'indiquer si les QTLs détectés dans notre étude l'ont été précédemment. Néanmoins, à notre connaissance, aucun QTL pour SD n'a été signalé dans le voisinage du SNP Affx-88920061 très significatif que nous observons sur Omy6. Dans la GWAS par BayesC $\pi$ , 3 autres QTLs pour SD sont détectés sur Omy11, Omy15 et Omy27 expliquant chacun de 1,2 à 1,9% de la variance génétique de SD. Il n'y a pas de gène candidat évident dans la grande région QTL définie sur Omy11. En revanche, on peut noter que le SNP le plus significatif sur Omy15 est localisé dans le gène ARHGEF4 (Rho Guanine Nucléotide Exchange Factor 4) qui est un facteur d'échange des nucléotides guanine pour les GTPases RHOA, RAC1 et CDC42. Les phénotypes mutants dans MGI pour ARHGEF4 concernent en particulier le système immunitaire et le métabolisme avec une diminution de la teneur en hémoglobine, en triglycérides circulants, des taux d'IgE et IgG1 et une diminution du nombre de lymphocytes T, mais une augmentation du nombre de cellules B matures, du taux de phosphatase alcaline circulante et du niveau total de protéines circulantes. Malgré l'importance de l'intervalle de crédibilité du dernier QTL détecté pour SD sur Omy27,

nous pouvons suggérer que le gène NR2E1 (Estrogen-related receptor gamma-like) est un gène candidat convainquant, car d'une part le SNP le plus significatif est situé juste à l'extrémité de ce gène. D'autre part, ce gène est un récepteur nucléaire, facteur de transcription de type gamma qui est en particulier impliqué dans la régulation des récepteurs de l'acide rétinoïque dans les cellules rétinienne, ce qui est essentiel pour la vision. Les phénotypes mutants de souris MGI pour NR2E1 ont une morphologie anormale du nerf optique et de la rétine, une morphologie cérébrale anormale, une taille corporelle réduite, une quantité de graisse corporelle réduite ainsi qu'une fertilité femelle réduite.

Nous pouvons émettre l'hypothèse que deux des quatre QTLs détectés pour SD sont peut-être associés à des morphologies oculaires anormales et à des défauts de vision qui pourraient rendre la truite moins sensible aux stimuli de la photopériode.

Concernant les caractères de fécondité, 2 QTL avec les mêmes pics SNP sont détectés pour SW et EN, l'un sur Omy2 (Affx-88951720), l'autre sur Omy12 (Affx-88950456), ce qui confirme que les 2 caractères sont biologiquement très proches. Le QTL sur Omy2 explique 2,9% et 3,2% de la variance génétique pour SW et EN, respectivement, et le QTL sur Omy12 explique 2,5% et 2,7% de la variance génétique de SW et EN, respectivement. Pour EN, un troisième QTL est détecté sur Omy8 qui explique 1,4% de la variance génétique, sans qu'aucun gène candidat ne puisse être proposé pour ce QTL. Pour SW, deux autres QTLs sont détectés sur Omy2 et Omy1 expliquant respectivement 1,9% et 1,3% de la variance génétique. L'intervalle de crédibilité est important (> 1Mb) pour ces deux QTLs et nous n'avons pas de gène candidat évident à proposer. En revanche, concernant le QTL commun à SW et EN sur Omy2, le pic SNP est localisé entre le gène MPRD (mannose-6phosphate receptor-like, cation-dependent) et le gène PHC1 (polyhomeotic-like protein 1). Le gène PHC1 est homologue à un gène polyhoméotique de la drosophile appartenant au groupe des gènes Polycomb. C'est un composant d'un complexe multiprotéique de type PRC1 du groupe Polycomb (PcG), qui est nécessaire pour maintenir l'état de répression transcriptionnelle de nombreux gènes (incluant les gènes Hox) pendant le développement. Les phénotypes homozygotes mutants de souris MGI pour le gène PHC1 présentent une létalité périnatale, des malformations squelettiques postérieures, une fente palatine, une hypoplasie parathyroïde et thymique ainsi que des anomalies oculaires et cardiaques. Nous émettons donc l'hypothèse que ce gène pourrait avoir un rôle dans la fécondité de la truite arc-en-ciel.

Pour les caractères de taille de l'œuf EW et ED, nous allons nous concentrer sur les 3 QTLs détectés sur Omy1 dans les deux types d'analyse GWAS des deux caractères (Tableau 10). Une première région QTL, commune à EW et ED, est détectée sur Omy1 entre 66,539 Mb et 68,881 Mb, bien que les pics SNPs ne soient pas les mêmes (Figure 16). Il n'y a pas de gène candidat évident dans cette large zone QTL expliquant 2% de la variance génétique pour EW. Pour EW, on distingue un second QTL sur Omy1 dans la région entre 60,498 Mb et 64,633Mb expliquant 1,2 % de la variance génétique (Figure 16). Bien qu'il n'y ait pas de gène candidat pouvant être proposé dans cette région QTL, on note la présence du gène PRKG2 (situé entre 64,660 et 64,676 Mb) à proximité immédiate et dont le rôle important est connu dans la maturation ovocytaire des mammifères et du poisson-zèbre (Li et al. 2018). Ce gène est donc un gène candidat potentiel.



**Figure 16 :** Manhattan plot représentant l'association entre les SNP par une approche BayesC $\pi$  et les caractères SW (violet), EW (jaune), ED (rouge) avec le SNP peak pour chaque QTL évident sur le chromosome 1 entre 60 et 80 Mn (A) et le chromosome 2 entre 5Mb et 20Mb (B).

La ligne bleue correspond au seuil  $\log\text{BF} > 6$  pour définir l'évidence d'un QTL. SW : poids de la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

Pour le caractère ED, on distingue également un autre QTL entre 70,848 Mb et 71,813 Mb sur Omy1, qui explique environ 2% de la variance génétique (Figure 16). Le pic SNP est situé à

proximité du gène WAPAL (wings apart-like protein homologs, positionné entre 71,675 et 71,699 Mb) qui est un gène candidat sérieux pour ED. En effet, ce gène est un régulateur de la structure et de la fonction des chromosomes méiotiques, jouant un rôle dans la cohésion des chromatides sœurs, la cohésine associée aux deux chromatines, la réparation de cassure des brins doubles d'ADN ainsi que dans le positionnement du globule polaire après les divisions méiotiques durant l'oogenèse (Stanvitch et Moore 2008; Crawley et al. 2016). Les mutants à ce gène du ver *C. elegans* présentent un défaut de la ponte, une réduction de la taille de la portée, une mortalité embryonnaire de 21% et une mortalité au stade larvaire de 28% (Crawley et al. 2016).

### 3.3. Efficacité de la sélection

#### 3.3.1. Impacts liés aux caractéristiques de la population d'apprentissage

##### **Effet de la taille de la population d'apprentissage**

Pour étudier l'impact de la taille de la population d'apprentissage sur l'évaluation génétique ou génomique, nous comparons la précision  $r$  et le biais  $b$  entre les scénarios PA+ et PA- pour des évaluations BLUP et GBLUP.

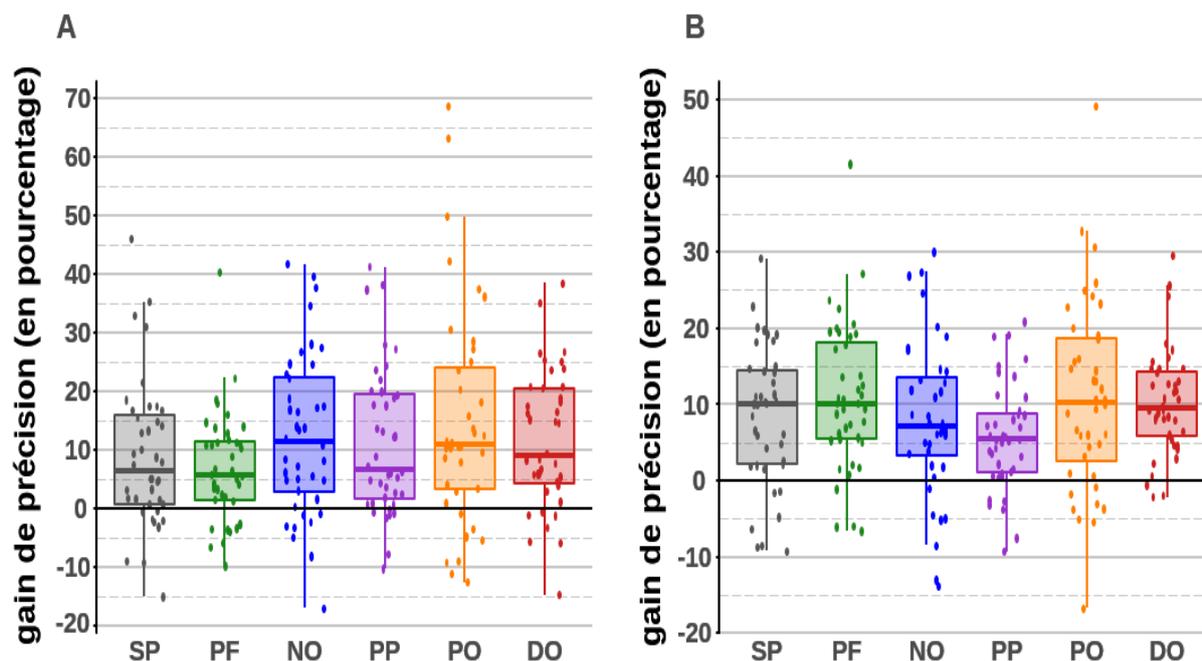
Pour le BLUP, la précision moyenne  $r$  de la sélection varie de 0,41 à 0,48 selon les caractères quand la population d'apprentissage est constituée de 672 individus (scénario PA-) et de 0,45 à 0,51 pour une population d'apprentissage de 1077 individus (Tableau 11). Les gains de précision sont donc en moyenne de 7% (FW) à 15% (EW) entre les scénarios PA- et PA+. Dans 80% des 40 simulations effectuées, la précision est meilleure avec une plus grande population d'apprentissage (Figure 17-A). Le gain de précision est variable d'un échantillonnage à un autre de la population et aussi d'un caractère à un autre. Pour une simulation donnée, il est possible d'avoir un gain de précision de 70% pour un caractère et aucun gain de précision pour un autre caractère, voire même, observer une baisse de précision (Figure 17-A). En ce qui concerne les coefficients de dispersion  $b$ , leurs valeurs moyennes sont proches de 1 et varient entre 0,93 à 1,05 sans qu'il y ait de différence notable entre les scénarios PA- et PA+.

Trait	BLUP				GBLUP			
	PA-		PA+		PA-		PA+	
	r	b	r	b	r	b	r	b
SD	0,48 (0,12)	1,01 (0,25)	0,51 (0,10)	0,97 (0,20)	0,57 (0,12)	0,92 (0,21)	0,61 (0,10)	0,92 (0,17)
FW	0,48 (0,11)	1,05 (0,29)	0,51 (0,10)	1,01 (0,26)	0,53 (0,10)	0,96 (0,21)	0,59 (0,10)	0,95 (0,20)
SW	0,43 (0,09)	0,96 (0,21)	0,47 (0,08)	0,97 (0,19)	0,58 (0,06)	0,94 (0,12)	0,62 (0,06)	0,93 (0,12)
EN	0,41 (0,10)	0,97 (0,27)	0,45 (0,09)	0,97 (0,23)	0,55 (0,08)	0,96 (0,14)	0,59 (0,08)	0,97 (0,15)
EW	0,41 (0,09)	0,93 (0,20)	0,46 (0,08)	0,95 (0,19)	0,50 (0,10)	0,88 (0,17)	0,55 (0,10)	0,86 (0,15)
ED	0,47 (0,08)	0,96 (0,19)	0,51 (0,07)	0,98 (0,15)	0,60 (0,07)	0,95 (0,14)	0,66 (0,07)	0,96 (0,12)

**Tableau 11 :** Précision moyenne (r) et coefficient moyen de dispersion (b) des 40 simulations d'évaluation BLUP et GBLUP sur les caractères de reproduction avec une population d'apprentissage de 1077 individus (PA+) ou seulement de 672 individus (PA-). Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

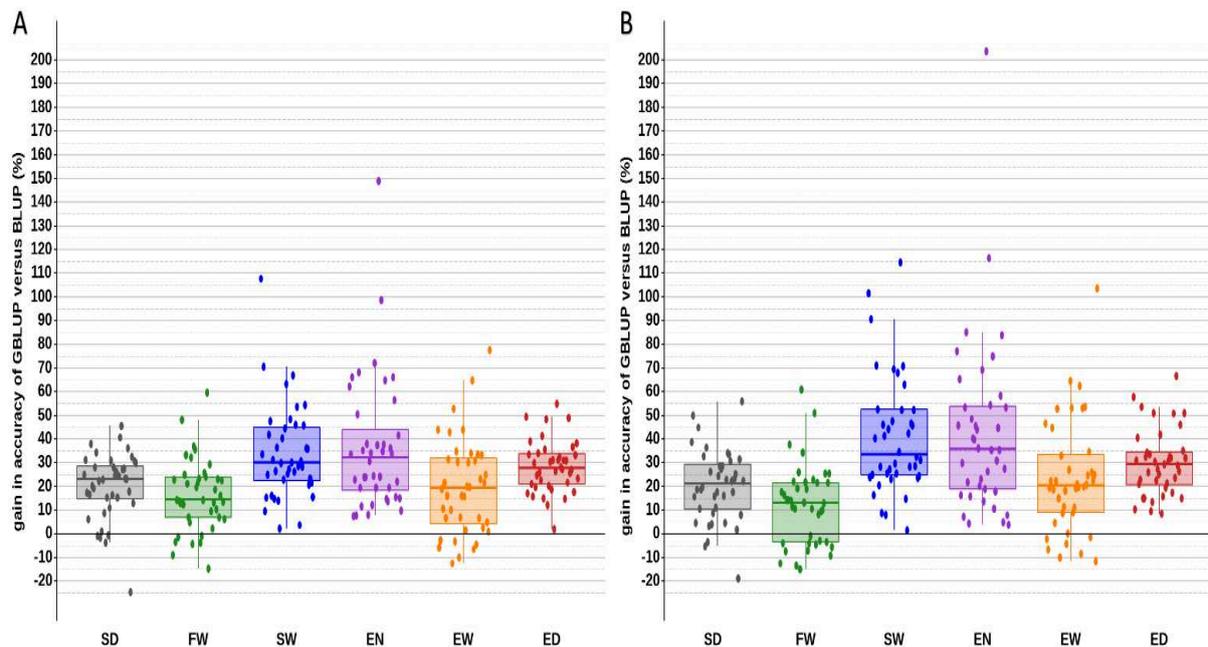
Pour le GBLUP, la précision de la sélection varie de 0,50 et 0,60 pour PA- et de 0,55 à 0,66 pour PA+, soit un gain moyen de précision de 6% à 11% (Tableau 11). Dans 85% des simulations, la précision augmente avec la taille de la population d'apprentissage (Figure 17-B). Tout comme en BLUP, le gain de précision avec la taille de la population d'apprentissage est variable d'une simulation à une autre ou d'un caractère à un autre (Figure 17-B); par exemple la simulation n°37 montre une perte de précision pour 3 caractères et un gain pour les 3 autres caractères ; la simulation n°36 montre, quant à elle, un gain de précision sur 5 caractères, dont deux, avec un gain supérieur à 20%, mais une perte de précision supérieure à 10% sur le dernier caractère (Figure 17-D). En GBLUP, tous les coefficients de dispersion b sont inférieurs à 1 (variant en moyenne entre 0,86 et 0,97), indiquant une faible sur-estimation de la variance des GEBV sans différence notable entre les scénarios PA- et PA+.



**Figure 17 :** Boîtes à moustaches du gain de précision (%) pour le BLUP (A) et le GBLUP (B) pour chaque répétition et chaque caractère entre une population d'apprentissage de 1077 individus (PA+) et une population d'apprentissage de 672 individus (PA-).

SD : semaine de ponte, FW : poids de la femelle avant la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids d'un œuf, ED : diamètre d'un œuf.

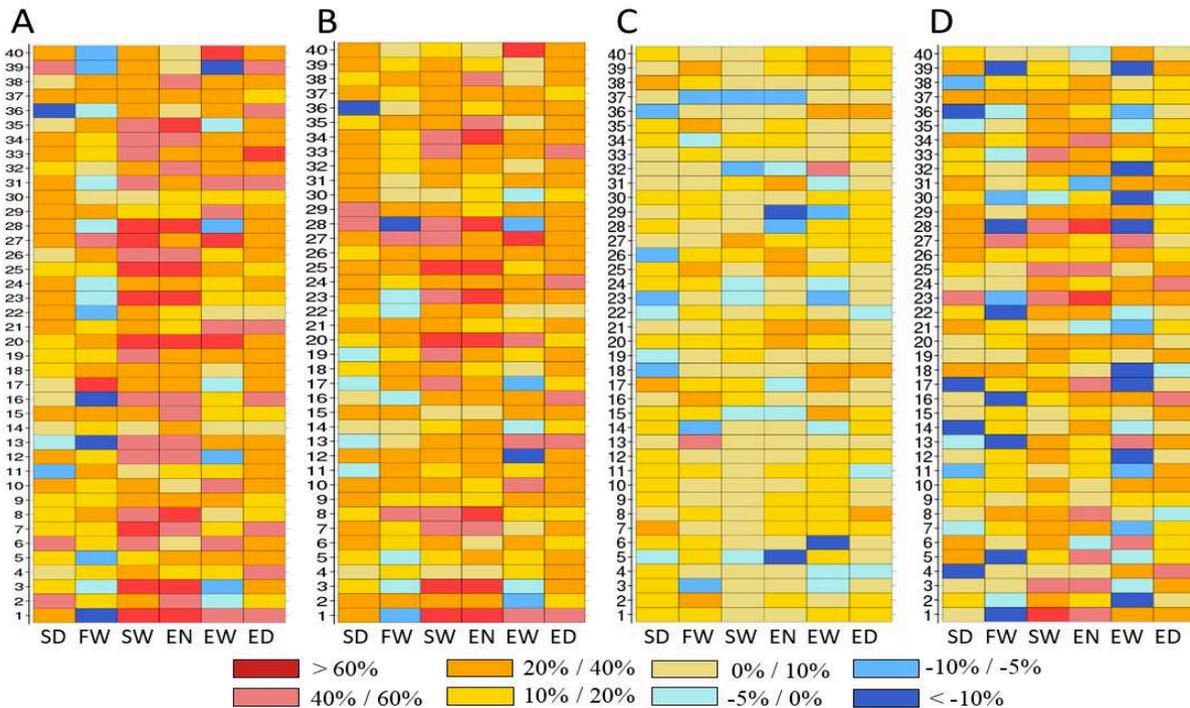
Le gain moyen de précision moyenne entre GBLUP et BLUP est de 10% à 39% pour le scénario PA- (Figure 18-A). Ce gain de précision est retrouvé pour le scénario PA+, avec un gain de précision moyen de 15 à 32% (Figure 18-B). Pour le scénario PA+, 48% des 40 simulations ont un gain de précision supérieur à 5% pour l'ensemble des caractères et 33% ont même un gain de précision supérieur à 10% sur tous les caractères (Figure 19-B). Dans 92% des simulations, le GBLUP permet un gain de précision par rapport au BLUP (Figure 19-A et Figure 19-B). À l'inverse, dans 15% des 40 simulations, on note une perte de précision de 0 à 5% entre GBLUP et BLUP pour au moins 1 caractère, dont 8% des simulations avec une perte de précision supérieure à 10% (Figure 19-A et Figure 19-B).



**Figure 18 :** Boîtes à moustaches du gain de précision (%) entre GBLUP et BLUP pour chaque simulation et chaque caractère avec une population d'apprentissage de (A) 1077 individus (PA+) ou (B) 672 individus (PA-).

SD : semaine de ponte, FW : poids de la femelle avant la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

Nous comparons maintenant l'efficacité de la sélection sur pedigree (BLUP) dans le scénario PA+ à une sélection génomique (GBLUP) dans le scénario PA- afin d'étudier la possibilité d'être aussi précis en GBLUP qu'en BLUP à moindre nombre d'individus phénotypés. L'intérêt de cette comparaison est d'envisager une SG pouvant être économiquement compétitive par rapport à une sélection sur pedigree, car le coût des puces à ADN est un frein économique à sa mise en œuvre dans les schémas de sélection. Le gain de précision moyen entre les scénarios GBLUP PA- et BLUP PA+ varie de 5% (FW) à 25% (SW) selon les caractères (Figure 19-D). Le coefficient de dispersion  $b$  moyen est similaire entre ces deux scénarios. Une évaluation génomique avec une population de référence de moins de 700 individus peut donc être plus précise qu'une évaluation génétique avec une population de référence de près de 1 100 individus.



**Figure 19 :** Heatmap du gain de précision pour chaque simulation et chaque caractère entre les scénarios (A) GBLUP PA- versus BLUP PA-, (B) GBLUP PA+ versus BLUP PA+, (C) GBLUP PA+ versus GBLUP PA- et (D) GBLUP PA- versus BLUP PA+.

PA+ : scénario avec une population d'apprentissage de 1077 individus ; PA- : population d'apprentissage de 672 individus

En conclusion, la sélection génomique est presque toujours plus efficace que la sélection sur pedigree à taille de population de référence constante. La taille de la population de référence a un impact clair, mais modéré sur la précision de l'évaluation. Dans nos scénarios, une augmentation de 60% de la taille de la population d'apprentissage permet un gain de précision de 6% à 15% selon le caractère .

### **Effet du niveau d'apparentement entre populations d'apprentissage et de validation**

Pour mesurer l'effet de l'apparentement entre la population d'apprentissage et la population de validation, nous comparons le scénario PA- aux scénarios PAg et PAf dans lesquels l'une des deux cohortes phénotypées est utilisée comme population d'apprentissage pour prédire les valeurs génétiques des individus de la seconde cohorte. L'apparentement moyen entre populations d'apprentissage (PA) et de validation (PV) dans ces scénarios PAg et PAf est de 0,051 avec une valeur maximale d'apparentement entre individus des deux cohortes de 0,30. Le scénario contrôle est le scénario PA- (individus de PA et PV tirés aléatoirement dans les 2

cohortes) dans lequel l'apparement moyen entre PA et PV  $\approx 0,065$  et où des pleines-sœurs se retrouvent réparties entre PA et PV.

#### Evaluation BLUP

Trait	Pag		Paf		PA-	
	r	b	r	b	r	b
SD	0,09	0,64	0,23	2,64	0,48 (0,12)	1,01 (0,25)
FW	0,17	0,88	0,23	2,26	0,48 (0,11)	1,05 (0,29)
SW	0,08	0,74	0,01	0,09	0,41 (0,10)	0,97 (0,27)
EN	0,13	1,03	0,10	0,86	0,43 (0,09)	0,96 (0,21)
EW	0,09	1,06	0,08	0,54	0,41 (0,09)	0,93 (0,20)
ED	0,15	1,50	0,19	1,36	0,47 (0,08)	0,96 (0,19)

#### Evaluation GBLUP

Trait	Pag		Paf		PA-	
	r	b	r	b	r	b
SD	0,30	0,58	0,39	1,29	0,57 (0,12)	0,92 (0,21)
FW	0,24	0,92	0,30	1,25	0,53 (0,10)	0,96 (0,21)
SW	0,39	2,26	0,29	0,69	0,55 (0,08)	0,96 (0,14)
EN	0,38	0,88	0,33	0,88	0,58 (0,06)	0,94 (0,12)
EW	0,30	1,05	0,33	0,72	0,50 (0,10)	0,88 (0,17)
ED	0,26	1,25	0,29	0,67	0,60 (0,07)	0,95 (0,14)

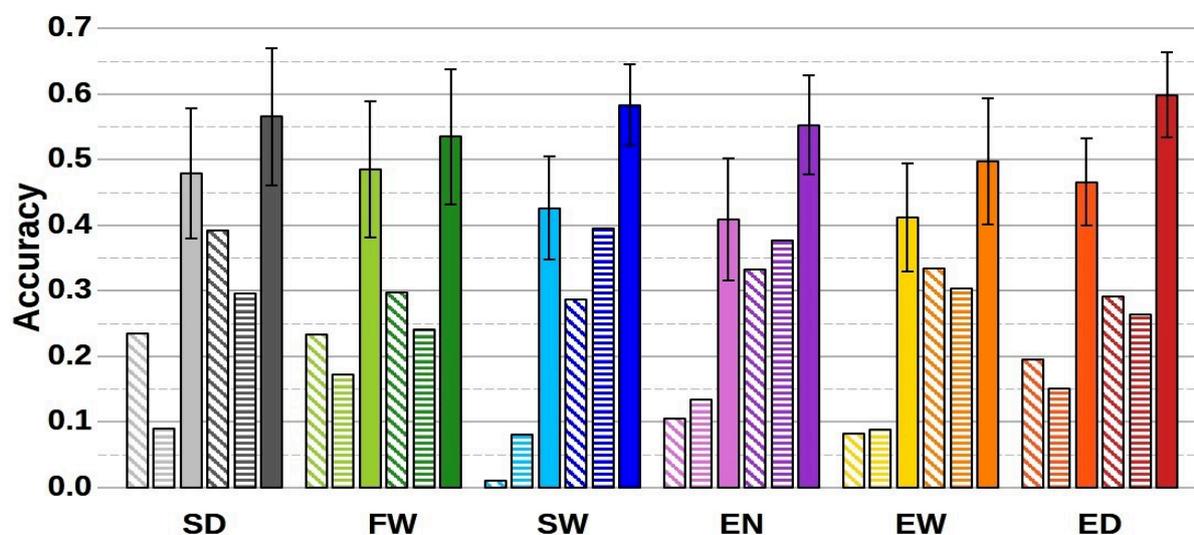
**Tableau 12 :** Précision (r) et coefficient de dispersion (b) pour les évaluations BLUP et GBLUP selon les scénarios PAG, PAg et PA- (moyennes et écarts-types (entre parenthèses) calculés sur 40 simulations)

PAG : population d'apprentissage de 720 individus de la cohorte G<sub>14</sub> et population de validation de 620 individus de la cohorte F<sub>15</sub> ; Paf population d'apprentissage de 620 individus de la cohorte F<sub>15</sub> et population de validation de 720 individus de la cohorte G<sub>14</sub>, PA- : population d'apprentissage de 672 individus. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte , EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

La précision r d'une sélection sur BLUP varie de 0,01 à 0,23 pour les scénarios PAG et Paf ce qui correspond à une chute de précision variant selon les caractères de -58% à -89% en moyenne par rapport au scénario PA- (Tableau 12). Par ailleurs, on note une forte variabilité du coefficient de dispersion selon les caractères (b varie de 0,08 à 2,64) dans les scénarios PAG et Paf alors que b reste très proche de 1 (estimations entre 0,93 à 1,05) pour tous les caractères dans le scénario PA-. La précision d'une sélection sur GBLUP varie de 0,24 à 0,39 pour les scénarios Paf et Pag, ce qui correspond à une baisse de précision variant selon les caractères de -36% à -53% par rapport au scénario PA- (Tableau 12). Si la variation du coefficient de dispersion en GBLUP est accrue dans les scénarios PAG et Paf (b varie de 0,42 à 1,29 selon le caractère) par rapport au scénario PA-, l'amplitude de cette variation est moindre que dans le cas de l'évaluation BLUP. Notons que la précision d'une évaluation BLUP dans le scénario

PA- est meilleure que celle d'une évaluation GBLUP dans les scénarios PAg et PAF où la population de validation est moins apparentée à la population d'apprentissage (Figure 20).

Ces résultats montrent qu'une évaluation génétique ou génomique efficace nécessite que les candidats à la sélection soient le plus apparentés possible à la population de référence. Même si la perte de précision est moindre en évaluation GBLUP (43% en moyenne dans notre étude) qu'en BLUP (71% en moyenne dans notre étude), un bon apparentement entre candidats à la sélection et population de référence reste un facteur très important pour l'efficacité de la sélection génomique (Figure 20).



**Figure 20 :** Histogramme de la précision de l'évaluation génétique pour 6 caractères de reproduction femelle en BLUP (couleurs claires) et GBLUP (couleurs foncées) avec la puce 30K pour 3 scénarios de population d'apprentissage (barres hachurées diagonalement : scénario PAg, barres hachurées horizontalement : scénario PAF et barres pleines : scénario PA-).

PAg : population d'apprentissage de 720 individus de la cohorte G<sub>14</sub> et population de validation de 620 individus de la cohorte F<sub>15</sub> ; PAF population d'apprentissage de 620 individus de la cohorte F<sub>15</sub> et population de validation de 720 individus de la cohorte G<sub>14</sub>, PA- : population d'apprentissage de 672 individus. SD : semaine de ponte, FW : poids de la femelle avant la ponte, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids d'un œuf, ED : diamètre d'un œuf.

Notre résultat est confirmé par une étude de l'efficacité de la sélection génomique pour la résistance au pou de mer du saumon Atlantique (Tsai et al. 2016). Dans cette étude, deux cohortes ont été phénotypées et génotypées sur 531 et 588 individus respectivement. Des précisions très faibles (0,05 – 0,11) sont observées quand l'une des deux cohortes sert de PA et l'autre sert de PV par rapport aux précisions obtenues (0,45 – 0,65) quand les deux populations sont constituées d'individus provenant des mêmes familles. L'étude de Palaiokostas et al. (2019) chez la carpe koï pour la résistance à l'herpèsvirus montre également une perte de précision de la sélection génomique quand il n'y a pas de familles de germains ou demi-

germaines réparties entre PA et PV et une augmentation de la précision quand des familles de pleins frères-sœurs sont réparties entre PA et PV.

Nos résultats ainsi que ceux de Tsai et al. (2016) et Palaiokostas et al. (2019) chez les poissons ne font que confirmer un phénomène très bien connu pour toutes les espèces d'élevage (Habier et al. 2010) : la sélection génomique ne peut être efficace que si les candidats à la sélection sont bien apparentés à la population de référence.

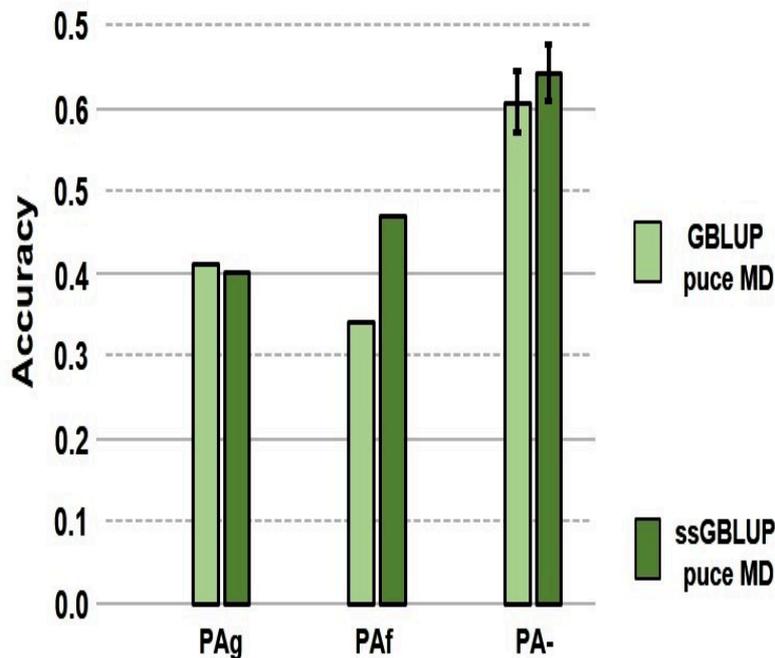
### **Effet de la prise en compte d'individus phénotypés mais non génotypés dans une évaluation ssGBLUP**

Le modèle ssGBLUP permet d'inclure des individus seulement phénotypés dans l'évaluation génomique pour espérer d'augmenter la précision des GEBV. Nous avons étudié deux caractères pour le ssGBLUP en fonction des données disponibles : SW (non ajusté pour FW) et ED. Pour les caractères SW et ED, nous avons pu inclure en plus dans les populations d'apprentissage : 321, 239, 287, 691, 29 et 135 individus phénotypés dans respectivement les cohortes A<sub>08</sub>, C<sub>10</sub>, B<sub>11</sub>, E<sub>12</sub>, G<sub>14</sub>, et F<sub>15</sub>. Pour la cohorte G<sub>14</sub>, les phénotypes des mères sont inclus dans les populations d'apprentissage alors que pour la cohorte F<sub>15</sub>, les phénotypes des mères ne sont pas disponibles.

Pour le caractère ED différentes méthodes de phénotypes ont été réalisées en fonction de la cohorte. Pour les cohortes E<sub>12</sub>, G<sub>14</sub> et F<sub>15</sub>, ED est obtenu en divisant la longueur de 50 œufs alignés le long d'une règle (après hydratation). Pour les cohortes A<sub>08</sub> et C<sub>10</sub>, ED est obtenue en divisant le nombre d'œufs sur 20 cm avant gonflement et pour la cohorte B<sub>11</sub>, ED est obtenue en divisant le nombre d'œufs sur 20 cm après hydratation.

Pour SW, la précision d'une évaluation ssGBLUP est de 0,64, 0,46, 0,40 respectivement pour les scénarios PA-, PA<sub>g</sub> et PA<sub>f</sub> (Figure 21). Pour le scénario PA-, l'augmentation de la précision est de 5,8% entre le ssGBLUP et GBLUP (Figure 21). On remarque un gain important de 37% de la précision pour le scénario PA<sub>f</sub> comparé au scénario PA<sub>g</sub> dans lequel, au contraire l'évaluation ssGBLUP apparaît même légèrement moins précise (0,40) que l'évaluation GBLUP (0,41). Cela s'explique par la prise en compte des phénotypes des mères (aussi génotypées) de la cohorte G<sub>14</sub> dans la population d'apprentissage, donc des phénotypes très apparentés à la population de validation (cohorte G<sub>14</sub> dans le scénario PA<sub>f</sub>). Pour le scénario PA<sub>g</sub>, l'ajout de 1 704 phénotypes dont 135 phénotypes de la cohorte F<sub>15</sub> (même cohorte des

individus de la population de validation) dans la population d'apprentissage, n'a pas permis un gain de précision pour ce scénario.



**Figure 21 :** Histogramme de la précision des évaluations GBLUP et ssGBLUP pour les scénarios PAg, PAF et PA- pour le caractère SW (non ajusté pour FW).

PAg : population d'apprentissage de 720 individus de la cohorte G<sub>14</sub> et population de validation de 620 individus de la cohorte F<sub>15</sub> ; PAF population d'apprentissage de 620 individus de la cohorte F<sub>15</sub> et population de validation de 720 individus de la cohorte G<sub>14</sub>, PA- : population d'apprentissage de 672 individus.

Pour ED, nous avons remarqué une augmentation de la précision de l'évaluation lors de l'ajout des individus phénotypés de la cohorte E<sub>12</sub> (même mesure que G<sub>14</sub> et F<sub>15</sub>), mais une perte de précision lors de l'ajout des individus des cohortes A<sub>08</sub>, C<sub>10</sub> et B<sub>11</sub> dont le phénotype était mesuré différemment pour apprécier ED (résultats non présentés). Rappelons donc que l'évaluation génétique ou génomique d'un caractère nécessite que les mêmes mesures soient effectuées au cours du temps, faute de quoi biais et perte de précision sont observés dans l'estimation des valeurs génétiques.

Le ssGBLUP permet d'inclure dans l'évaluation des individus phénotypés, mais non genotypés ce qui peut accroître la précision de la sélection si les individus rajoutés sont bien apparentés aux candidats à la sélection.

### 3.4. Impact de la densité en marqueurs de la puce à ADN

L'un des freins essentiels à la mise en œuvre de la sélection génomique est le coût du génotypage de la population de référence et des candidats à la sélection. Pour réduire le coût du

génotypage, il est nécessaire de réduire très significativement le nombre de marqueurs par rapport à la puce MD, mais une forte réduction du nombre de marqueurs s'accompagne d'une baisse de la précision de l'évaluation génomique. Nous cherchons donc ici à quantifier cette perte de précision si le set de marqueurs considéré varie de 1 000 à 10 000 SNPs au lieu des 30K SNPs utiles de la puce MD.

Des puces 1K, 3K, 6K et 10K ont été simulées avec un degré d'exigence sur les MAF plus (25% minimum) ou moins fort (5% minimum) soit sur la population de Viviers de Sarrance, soit sur la population d'Aqualande (AQ). Rappelons que la stratégie de constitution de ces puces basse-densité (LD) et leurs noms sont donnés en paragraphe : 2.4.2. .

Les évaluations ont été effectuées pour les scénarios PA- et PA+ (cf [Tableau 6](#)) afin d'évaluer les effets conjoints de la réduction du nombre de marqueurs et de la taille de la population d'apprentissage. Comme attendu, la précision de l'évaluation génomique diminue avec la densité en marqueurs pour l'ensemble des caractères ([Tableau 13](#) (scénario PA+), [Annexe - tableau 3](#) (scénario PA-)), quel que soit le scénario considéré. En moyenne sur les 40 simulations de chaque cas (scénario x puce LD x caractère), la précision de l'évaluation GBLUP est meilleure que celle du BLUP quels que soient le caractère et la puce LD considérée, sauf dans le cas du caractère SP et la puce 1K-25 ([Annexe - figure 1](#)).

Le biais change très peu avec la densité des marqueurs sauf pour la puce 1K et la puce 3K (conçue sur la population AQ).

Dans le scénario PA+, la perte de précision d'une évaluation GBLUP en LD par rapport à l'évaluation GBLUP en MD, moyennée sur l'ensemble des caractères, est de -11,2%, -5,3%, -2,5% et -1,7% pour les puces LD 1K, 3K, 6K, 10K respectivement. On note des pertes de précision plus importantes dans le scénario PA- que dans le scénario PA+ ([Tableau 14](#)). Il est donc important de définir le nombre minimal de marqueurs à considérer en fonction de la taille de la population de référence.

Si les pertes de précision associées à une réduction du nombre de marqueurs dépendent avant tout de la densité en marqueurs considérée, elles sont aussi fonction des MAF des marqueurs retenus ([Tableau 13](#)). Pour les puces 10K et 6K, il n'y a pas d'avantage à faire une présélection des SNPs sur des  $MAF \geq 25\%$ , voire celle-ci peut avoir un impact négatif sur la précision de l'évaluation de certains caractères (cf . FW et EW, [Tableau 13](#)), vraisemblablement quand des régions QTLs deviennent insuffisamment balisées par les marqueurs. Pour des puces à plus

faible densité en marqueurs (3K et 1K), la présélection des SNPs sur des  $MAF \geq 25\%$  s'avère souvent intéressante ou, à défaut, sans impact majeur (Tableau 13).

Pour les puces AQ, la perte de précision est plus importante par rapport aux puces conçues à partir des génotypages de VD (Tableau 14). Cette perte de précision est moins importante avec l'augmentation de la densité de la puce.

Puce LD	SD		FW		SW		EN		EW		ED	
	r	b	r	b	r	b	r	b	r	b	r	b
BLUP	0,51 (0,10)	0,97 (0,20)	0,51 (0,10)	1,01 (0,26)	0,47 (0,08)	0,97 (0,19)	0,45 (0,09)	0,97 (0,23)	0,46 (0,08)	0,95 (0,19)	0,51 (0,07)	0,98 (0,15)
1K-05-VD	0,54 (0,11)	0,85 (0,18)	0,55 (0,11)	0,88 (0,21)	0,54 (0,08)	0,88 (0,14)	0,49 (0,09)	0,86 (0,17)	0,52 (0,09)	0,85 (0,16)	0,57 (0,06)	0,84 (0,12)
1K-25-VD	0,52 (0,11)	0,79 (0,18)	0,58 (0,11)	0,92 (0,21)	0,54 (0,08)	0,88 (0,15)	0,49 (0,09)	0,86 (0,18)	0,50 (0,09)	0,79 (0,15)	0,58 (0,06)	0,84 (0,11)
3K-05-VD	0,56 (0,11)	0,91 (0,19)	0,61 (0,11)	1,00 (0,22)	0,56 (0,07)	0,95 (0,14)	0,50 (0,09)	0,94 (0,2)	0,52 (0,09)	0,86 (0,15)	0,61 (0,07)	0,93 (0,13)
3K-25-VD	0,58 (0,11)	0,94 (0,19)	0,58 (0,11)	0,95 (0,21)	0,57 (0,07)	0,96 (0,14)	0,54 (0,09)	0,99 (0,19)	0,56 (0,09)	0,90 (0,14)	0,64 (0,06)	0,94 (0,12)
3K-05-AQ	0,53 (0,10)	0,86 (0,18)	0,60 (0,10)	0,96 (0,20)	0,54 (0,07)	0,91 (0,14)	0,52 (0,09)	0,95 (0,19)	0,52 (0,09)	0,86 (0,16)	0,6 (0,06)	0,92 (0,13)
3K-25-AQ	0,55 (0,11)	0,87 (0,18)	0,58 (0,10)	0,92 (0,19)	0,53 (0,07)	0,87 (0,14)	0,50 (0,08)	0,91 (0,17)	0,52 (0,10)	0,82 (0,16)	0,6 (0,07)	0,91 (0,13)
6K-05-VD	0,59 (0,11)	0,98 (0,18)	0,61 (0,10)	1,00 (0,21)	0,58 (0,07)	0,99 (0,14)	0,55 (0,09)	1,03 (0,20)	0,56 (0,09)	0,92 (0,14)	0,65 (0,06)	0,98 (0,13)
6K-25-VD	0,57 (0,11)	0,93 (0,18)	0,57 (0,10)	0,94 (0,21)	0,59 (0,06)	0,98 (0,13)	0,55 (0,08)	1,01 (0,18)	0,55 (0,09)	0,88 (0,15)	0,65 (0,06)	0,95 (0,12)
6K-05-AQ	0,58 (0,10)	0,92 (0,17)	0,59 (0,10)	0,96 (0,20)	0,55 (0,07)	0,93 (0,14)	0,52 (0,08)	0,96 (0,18)	0,54 (0,10)	0,87 (0,16)	0,62 (0,07)	0,94 (0,13)
6K-25-AQ	0,58 (0,11)	0,92 (0,18)	0,59 (0,10)	0,95 (0,20)	0,58 (0,07)	0,95 (0,13)	0,54 (0,08)	0,98 (0,18)	0,53 (0,10)	0,86 (0,16)	0,62 (0,07)	0,94 (0,13)
10K-05-VD	0,59 (0,10)	0,96 (0,18)	0,61 (0,10)	1,00 (0,21)	0,60 (0,07)	1,01 (0,14)	0,56 (0,08)	1,04 (0,19)	0,57 (0,09)	0,91 (0,15)	0,66 (0,06)	0,98 (0,13)
10K-25-VD	0,59 (0,11)	0,93 (0,18)	0,57 (0,11)	0,93 (0,21)	0,60 (0,06)	0,96 (0,12)	0,56 (0,08)	0,99 (0,17)	0,54 (0,09)	0,86 (0,15)	0,64 (0,07)	0,94 (0,12)
10K-05-AQ	0,59 (0,11)	0,94 (0,18)	0,60 (0,10)	0,96 (0,21)	0,58 (0,06)	0,97 (0,13)	0,55 (0,08)	1,00 (0,18)	0,54 (0,10)	0,87 (0,16)	0,64 (0,07)	0,95 (0,13)
10K-25-AQ	0,58 (0,11)	0,92 (0,18)	0,60 (0,10)	0,95 (0,20)	0,58 (0,06)	0,94 (0,13)	0,56 (0,08)	0,99 (0,17)	0,54 (0,10)	0,87 (0,16)	0,64 (0,07)	0,95 (0,13)
MD 30K	0,61 (0,10)	0,92 (0,17)	0,59 (0,10)	0,95 (0,20)	0,62 (0,06)	0,93 (0,12)	0,59 (0,08)	0,97 (0,15)	0,55 (0,10)	0,86 (0,15)	0,66 (0,07)	0,96 (0,12)

**Tableau 13 :** Précision moyenne (r) et coefficient de dispersion moyen (b) d'une évaluation GBLUP pour les caractères de reproduction calculés sur 40 simulations du scénario PA+ et différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

Dans le nom des puces LD, K indique le nombre de marqueurs en milliers et 05 et 25 indiquent la MAF considérée pour le choix des SNPs et k. La mention AQ signifie que le choix des marqueurs a été effectué sur les génotypes de la population d'Aqualande. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

Nos résultats sont cohérents avec les études antérieures menées pour quantifier les baisses de précision associées à une moindre densité en marqueurs que cela soit chez la truite arc-en-ciel et la sélection pour la résistance à la flavobactériose (Vallejo et al. 2018) ou chez le saumon Atlantique et la sélection pour la résistance au pou de mer (Ødegård et al. 2014; Tsai et al. 2016; Tsairidou et al. 2020) ou pour des caractères de production (Tsai et al. 2015b). Toutes ces études conduites chez les salmonidés montrent qu'il n'y a aucun gain significatif de précision de la sélection génomique au-delà de 10K SNPs même en comparant avec des puces HD de 110K SNPs (Tsai et al. 2015b) ou 220K SNPs (Ødegård et al. 2014). Elles montrent également que les pertes de précision entre 5K et 10K SNPs restent faibles, mais qu'en revanche elles deviennent conséquentes (>10%) en deçà de 5K SNPs à très conséquentes en deçà de 1K (Ødegård et al. 2014; Tsai et al. 2015b; Vallejo et al. 2018; Tsairidou et al. 2020).

Puces LD	Scénarios	
	PA-	PA+
1K-05-VD	-13,49 (10,22)	-10,99 (8,94)
1K-25-VD	-12,89 (9,99)	-11,38 (8,22)
3K-05-VD	-8,92 (7,92)	-6,82 (6,66)
3K-05-AQ	-	-7,93 (6,64)
3K-25-VD	-6,01 (7,13)	-3,78 (5,95)
3K-25-AQ	-	-9,10 (6,23)
6K-05-VD	-4,00 (6,73)	-1,56 (5,45)
6K-05-AQ	-	-5,37 (5,13)
6K-25-VD	-5,22 (5,5)	-3,35 (4,18)
6K-25-AQ	-	-4,41 (4,96)
10K-05-VD	-2,57 (5,44)	-0,51 (4,15)
10K-05-AQ	-	-3,16 (4,24)
10K-25-VD	-4,23 (4,18)	-2,89 (3,12)
10K-25-AQ	-	-2,98 (3,56)

**Tableau 14 :** Perte de précision moyenne sur les six caractères de reproduction et écart-type d'échantillonnage (entre parenthèses) dans les scénarios PA- et PA+ des évaluations GBLUP sur puce LD par rapport à la précision des évaluations GBLUP sur puce MD.

Dans le nom des puces LD, K indique le nombre de marqueurs en milliers et 05 et 25 indiquent la MAF considérée pour le choix des SNPs et k. La mention AQ signifie que le choix des marqueurs a été effectué sur les génotypes de la population d'Aqualande. PA+ : scénario avec une population d'apprentissage de 1077 individus ; PA- : population d'apprentissage de 672 individus

Dans notre étude, comme dans la plupart des études pré-citées, nous avons évalué la précision de la sélection génomique pour une unique densité en marqueurs choisie pour génotyper à la fois la population de référence et les candidats à la sélection. Une autre stratégie envisageable

consiste à génotyper les candidats à la sélection sur une puce LD et à avoir une population de référence et les reproducteurs génotypés à plus forte densité (généralement en MD avec ~50 K SNPs). Cette stratégie est par exemple celle qui est conduite dans les populations porcines (Huang et al. 2012; Wellmann et al. 2013). Les génotypes des candidats à la sélection sont alors imputés en génotypages MD et l'évaluation GBLUP est effectuée sur une population d'individus aux génotypes vrais ou imputés en MD. Dans cette approche, les candidats retenus comme reproducteurs sont regénotypés en MD pour assurer l'efficacité de l'imputation des génotypes de la génération suivante de candidats à la sélection et permettre le renouvellement de la population de référence. Tsairidou et al. (2019) ont envisagé une telle stratégie chez le saumon Atlantique avec une population de référence et des reproducteurs génotypés en 5K et des génotypages de candidats réalisés sur seulement 200 marqueurs puis imputés en 5K. Ils ont montré une très bonne précision avec 5K marqueurs par rapport à la puce commerciale 132K (78K marqueurs valides). Ensuite, ils montrent une efficacité très proche avec seulement 200 marqueurs imputés en 5K. Une étude complémentaire sur l'intérêt et l'avantage de l'imputation par rapport à une puce LD peut être envisagée.

### 3.5. Progrès génétique et coût d'une évaluation génétique ou génomique de la production d'œufs

#### 3.5.1. Progrès génétique

La mise en place d'une sélection permet de réaliser un progrès génétique au sein d'une population et donc une amélioration (augmentation ou diminution) des performances pour les caractères sélectionnés. Le progrès génétique est conditionné par la variabilité génétique du caractère, l'intensité de sélection et la précision de l'évaluation des candidats à la sélection.

Dans le cas simple d'une sélection sur un seul caractère, on note des gains espérés importants pour un taux de sélection de 10% (Tableau 15). Une sélection BLUP permettrait ainsi une augmentation du nombre d'œufs de la ponte de 555 œufs par femelle et par génération, soit un gain de 12% par rapport à la moyenne de la génération précédente. Une sélection BLUP sur le poids de la ponte permettrait quant à elle une augmentation de 26 g de ce poids, soit un gain de 14% par rapport à la moyenne génération précédente. La SG permettrait des gains plus élevés,

de 732 œufs par femelle et génération pour une sélection sur EN et de 34g de poids de ponte pour une sélection sur SW, soit des augmentations respectives des performances de 15% et 18% par rapport à la moyennes de la génération précédente.

	BLUP				GBLUP			
	PA-		PA+		PA-		PA+	
	$\Delta G$	$\% \Delta G / \bar{y}$	$\Delta G$	$\% \Delta G / \bar{y}$	$\Delta G$	$\% \Delta G / \bar{y}$	$\Delta G$	$\% \Delta G / \bar{y}$
SD	0,64	25,0	0,69	26,8	0,76	29,6	0,82	31,8
FW	163,37	7,8	173,17	8,3	180,18	8,6	198,43	9,5
SW	23,74	12,5	26,06	13,7	32,49	17,1	34,28	18,1
EN	504,09	10,50	555,47	11,6	682,13	14,2	732,93	15,3
EW	2,26	5,7	2,54	6,4	2,73	6,9	3,00	7,6
ED	0,10	2,5	0,11	2,7	0,13	3,2	0,14	3,5

**Tableau 15 :** Progrès génétique attendu ( $\Delta G$ ) par génération pour un taux de sélection de 10% et une sélection mono-caractère et expression en % par rapport à la moyenne du caractère ( $\Delta G / \bar{y}$ ).

Scénario PA- : 40 simulations avec une population d'apprentissage de 672 individus, Scénario PA+ : 40 simulations avec une population d'apprentissage de 1077 individus. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte , EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

La filière trutticole française vendant plus de 700 millions d'œufs embryonnés par an (communication SYSAAF), un gain de 15% du nombre d'œufs par génération représenterait à l'échelle de la filière un gain cumulatif d'environ 38 millions d'œufs embryonnés. En considérant une valeur moyenne de 12€ les 1 000 œufs vendus (SYSAAF, communication personnelle), le gain économique potentiel pour les entreprises de sélection trutticole serait alors de 380 k€ par an.

### 3.5.2. Coût et gain économiques de l'amélioration génétique du nombre d'œufs de la ponte

Si nous avons montré que la sélection génomique était techniquement plus efficace qu'une sélection sur pedigree à taille de population de référence constante, il est important d'évaluer si elle peut aussi être économiquement efficace à relativement court terme, car c'est à cette seule condition qu'elle pourra être réellement mise en œuvre. Il s'agit donc non seulement d'évaluer les recettes espérées de cette sélection, mais également les coûts économiques engendrés par le schéma de sélection.

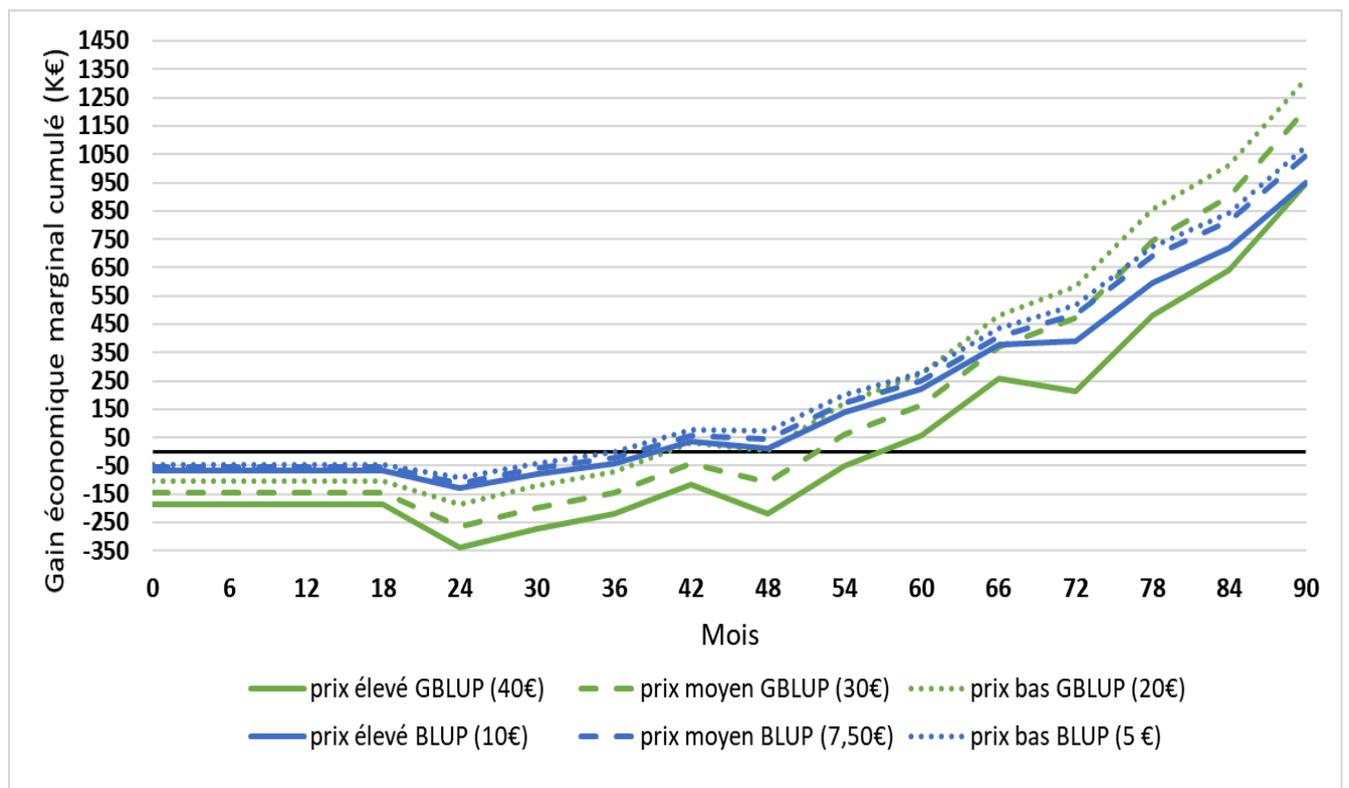
Pour évaluer l'intérêt économique de la mise en œuvre d'une sélection génomique sur les caractères de reproduction femelle, nous calculons le gain économique marginal (ou bénéfice attendu) permis par une sélection sur 3 générations du nombre d'œufs de la ponte (EN) soit par BLUP, soit par GBLUP. A chaque génération, 3 000 candidats à la sélection sont génotypés. A la génération 1, 1 000 collatéraux sont génotypés et phénotypés pour constituer la population de référence initiale (PR1). Puis aux générations suivantes, 500 nouveaux collatéraux sont génotypés et phénotypés pour accroître la population de référence. Le progrès génétique réalisé est considéré constant d'une génération à l'autre. Les différents coûts à considérer sont le prix de l'assignation de parenté en évaluation BLUP versus le prix d'un génotypage MD pour l'évaluation GBLUP et le coût du phénotypage (25€ par individu) de la population de référence. A chaque génération, la recette marginale est due à l'augmentation du nombre d'œufs embryonnés produits par deux cohortes de 10 000 femelles issues de parents améliorateurs (Figure 22). On considère un prix de vente de 12€ pour 1 000 œufs supplémentaires. Le progrès génétique sur le nombre d'œufs est considéré identique pour les 3 pontes d'une même cohorte. Mais seulement 80% des femelles d'une cohorte passe en deuxième ponte et 70% des femelles en deuxième ponte ont une troisième ponte. Le taux de fécondation des œufs est supposé constant pour les 3 pontes et égal à 75%.

Unité de production		Temps (mois)	Noyau de sélection
Cohorte 1	Cohorte 2		
		0	Naissance de la génération 1 issue de parents sélectionnés (PR1=1 000)
		6	
		12	
		18	
1 <sup>er</sup> ponte (g1)		24	Naissance de la génération 2 issue de la génération 1 (PR2=PR1+500)
		30	
2 <sup>ème</sup> ponte (g1)	1 <sup>er</sup> ponte (g1)	36	
3 <sup>ème</sup> ponte (g1)		42	
1 <sup>er</sup> ponte (g2)	2 <sup>ème</sup> ponte (g1)	48	Naissance de la génération 3 issue de la génération 2 (PR3=PR2+500)
	3 <sup>ème</sup> ponte (g1)	54	
2 <sup>ème</sup> ponte (g2)	1 <sup>er</sup> ponte (g2)	60	
3 <sup>ème</sup> ponte (g2)		66	
1 <sup>er</sup> ponte (g3)	2 <sup>ème</sup> ponte (g2)	72	Naissance de la génération 4 issue de la génération 3 (PR4=PR3+500)
	3 <sup>ème</sup> ponte (g2)	78	
2 <sup>ème</sup> ponte (g3)	1 <sup>er</sup> ponte (g3)	84	
3 <sup>ème</sup> ponte (g3)		90	

**Figure 22 :** Schéma sur 3 générations de parents sélectionnés produisant chacune 2 cohortes de 10 000 femelles reproductrices en première ponte.

Trois scénarios de prix sont considérés pour la puce MD : un prix élevé de 40 €, un prix moyen de 30 € et un prix bas de 20 € par individu. Le coût de l'assignation de parenté est comptabilisé à 25% du coût d'un génotypage MD, soit 10 €, 7,50 € ou 5 € selon le scénario.

Pour l'ensemble des scénarios (Figure 23), les coûts de la mise en place d'une sélection BLUP sont rentables dès les premiers 42 mois. Pour le scénario GBLUP à coût de génotypage bas, la rentabilité est atteinte au même stade, alors qu'il faut attendre jusqu'à 60 mois pour le scénario GBLUP à coût de génotypage élevé. Outre un délai au retour sur investissement, la SG demande un investissement initial beaucoup plus important que la sélection sur pedigree. Quand le génotypage est à bas coût, la SG est plus rentable que la sélection sur pedigree au bout de 66 mois ; il faut attendre 24 mois de plus si le génotypage est à coût élevé. Notons que l'évolution des technologies devrait encore permettre une certaine baisse des coûts de génotypage dans les années à venir, mais très vraisemblablement plus importante pour du génotypage MD que pour des puces d'assignation à quelques dizaines de marqueurs. Les évolutions des prix seront plus en faveur de la SG et améliorant encore plus sont intérêt.



**Figure 23 :** Gain économique marginal cumulé sur 90 mois d'une sélection sur pedigree (BLUP : bleu) ou génomique (GBLUP : vert) du nombre d'œufs de la ponte, selon le coût du génotypage.

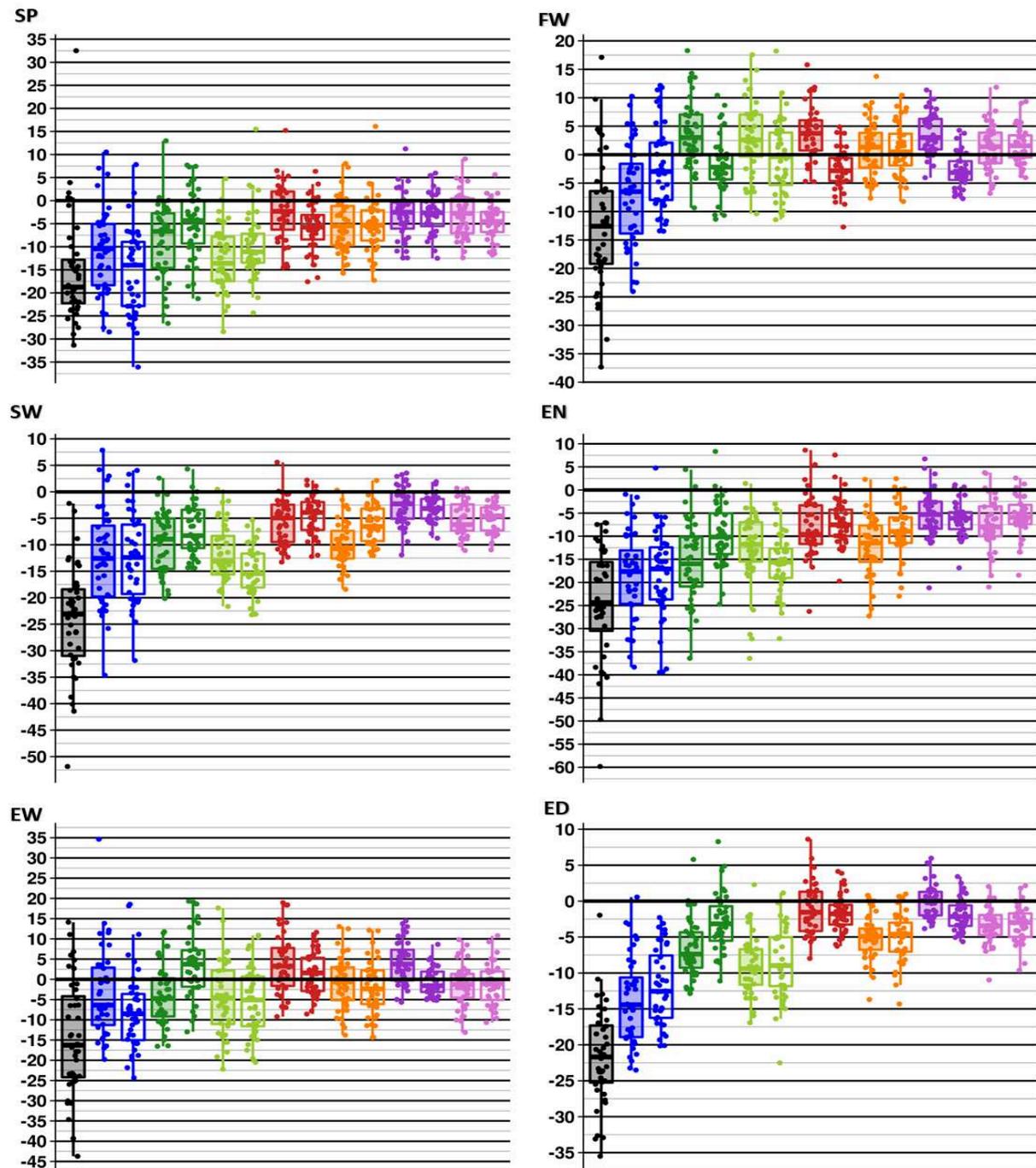
## 4. Résumé

La truite arc-en-ciel est la principale filière piscicole en France, les caractères de reproduction femelle jouent un rôle important dans l'économie des entreprises de sélection avec la vente des œufs fécondés ou de caviar. Les objectifs de cette étude sont triples : connaître les paramètres génétiques des caractères de reproduction femelle, préciser l'architecture génétique de ces caractères par l'identification des QTL et évaluer l'efficacité attendue d'une sélection sur pedigree (BLUP) ou d'une sélection génomique (GBLUP) de ces caractères. Pour cela, 1343 truites ont été génotypées sur 57 000 SNP (Single Nucleotide Polymorphism) avec la puce Axiom™ Trout Genotyping array et phénotypées pour évaluer sept caractères : le poids de la femelle avant la ponte (FW) et après la ponte (PW), le poids de la femelle à deux ans, le poids de la ponte et le nombre d'œufs (SW, EN), le poids et la taille moyenne d'un œuf (EW, ED). L'estimation des paramètres génétiques a été faite par AIREML en modèle animal multi-caractères. Les analyses d'association pan-génomique pour détecter des QTLs ont été conduites par analyse multi-marqueurs GBLUP ou BayesC $\pi$ . Pour différents scénarios de densité de marqueurs, de taille de population d'apprentissage et d'apparentement à la population de candidats à la sélection, l'efficacité de la sélection a été appréciée par la précision de la sélection ( $r$ ) estimée comme la corrélation entre les valeurs génétiques prédites et les phénotypes corrigés ( $y^*$ ) pour les effets fixes, divisée par la racine carrée de l'héritabilité. En outre, le coefficient de régression des  $y^*$  sur les valeurs génétiques prédites a été calculé comme mesure de biais des évaluations génétiques et génomiques. Les sept caractères étudiés ont des héritabilités moyennes variant de 0,27 à 0,44. Les caractères EN et SW sont génétiquement très corrélés (0,86) entre eux, EW et ED le sont encore plus (0,99). Si EN est peu corrélé aux autres caractères, SW est génétiquement bien corrélé à EW et ED (0,49 et 0,59) mais pas significativement aux poids FW et PW des reproductrices. Tous les caractères sont très polygéniques, seulement 2 à 4 QTLs expliquant chacun plus de 1% (mais moins de 3%) de la variance génétique du caractère. Une sélection assistée sur marqueurs ne peut donc être efficace, mais une sélection génomique GBLUP permettrait des gains de précision de 10 à 37% par rapport à la sélection BLUP sur pedigree. La sélection sur pedigree a, quant à elle, des précisions quel que soit le caractère considéré proche de 0,5 pour une population d'apprentissage d'environ 1 100 collatéraux des candidats à la sélection. Des pertes de précision fortes (-71% en BLUP et -43% en GBLUP) sont par ailleurs observées quand populations d'apprentissage et de validation sont moins apparentées. L'utilisation de génotypages à basse

densité avec 6 000 à 10 000 marqueurs pourrait permettre des précisions de la sélection presque aussi élevées (0 à 5% maximum de perte de précision selon les cas testés) qu'une sélection génomique avec la puce commerciale actuelle à 30 000 marqueurs utiles.

Le chapitre suivant nous permettra d'une part de mieux préciser l'intérêt des puces LD et de populations de référence constituées d'individus de cohortes successives, et d'autre part d'estimer les gains de précision attendus pour la sélection génomique des caractères de production.

## 5. Annexes



**Annexe - figure 1:** Boîtes à moustaches de la perte de précision (%) entre le BLUP, les puces LD (GBLUP) versus un GBLUP avec la puce MD pour chaque caractère avec une population d'apprentissage de 1077 individus (PA+).

Chaque point d'une boîte à moustaches représente l'une des 40 simulations du scénario. Les boîtes à moustaches transparentes représentent les puces LD avec un tri sur la  $MAF > 0,25$  et les boîtes à moustaches colorées représente les puces avec un tri sur la  $MAF > 0,05$ . Les couleurs représentent la densité de marqueurs et les génotypes utilisés pour le choix des marqueurs : BLUP : noir, 1K : bleu foncé (VD), 3K : vert foncé (VD) - clair (AQ), 6K : rouge (VD) - orange (AQ), 10K : violet (VD) – rose (AQ). SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte , EW : poids d'un œuf, ED : diamètre d'un œuf.

Puce LD	SD		FW		SW		EN		EW		ED	
	r	b	r	b	r	b	R	b.	r	b	r	b
BLUP	0,48 (0,12)	1,01 (0,25)	0,48 (0,11)	1,05 (0,29)	0,41 (0,10)	0,97 (0,27)	0,43 (0,09)	0,96 (0,21)	0,41 (0,09)	0,93 (0,20)	0,47 (0,08)	0,96 (0,19)
1K-05-VD	0,49 (0,11)	0,85 (0,21)	0,49 (0,11)	0,87 (0,23)	0,43 (0,10)	0,85 (0,20)	0,49 (0,08)	0,88 (0,14)	0,47 (0,09)	0,84 (0,17)	0,51 (0,07)	0,83 (0,14)
1K-25-VD	0,47 (0,12)	0,81 (0,22)	0,50 (0,10)	0,90 (0,22)	0,44 (0,10)	0,86 (0,20)	0,50 (0,08)	0,89 (0,15)	0,46 (0,09)	0,81 (0,17)	0,53 (0,07)	0,84 (0,13)
3K-05-VD	0,50 (0,13)	0,91 (0,24)	0,55 (0,11)	1,01 (0,23)	0,44 (0,10)	0,92 (0,21)	0,51 (0,07)	0,95 (0,15)	0,48 (0,09)	0,88 (0,17)	0,56 (0,07)	0,93 (0,15)
3K-25-VD	0,52 (0,13)	0,93 (0,23)	0,52 (0,10)	0,95 (0,22)	0,47 (0,09)	0,96 (0,20)	0,52 (0,07)	0,95 (0,13)	0,52 (0,09)	0,93 (0,17)	0,58 (0,07)	0,94 (0,14)
6K-05-VD	0,53 (0,12)	0,97 (0,22)	0,55 (0,10)	1,02 (0,22)	0,48 (0,09)	1,00 (0,20)	0,53 (0,07)	0,99 (0,14)	0,52 (0,09)	0,95 (0,17)	0,59 (0,07)	0,98 (0,15)
6K-25-VD	0,52 (0,12)	0,93 (0,23)	0,52 (0,10)	0,95 (0,21)	0,49 (0,09)	0,98 (0,19)	0,54 (0,06)	0,97 (0,13)	0,51 (0,09)	0,91 (0,17)	0,59 (0,07)	0,96 (0,14)
10K-05-VD	0,53 (0,12)	0,96 (0,22)	0,55 (0,10)	1,02 (0,22)	0,50 (0,09)	1,02 (0,19)	0,55 (0,06)	1,01 (0,14)	0,52 (0,09)	0,94 (0,17)	0,59 (0,07)	0,98 (0,14)
10K-25-VD	0,54 (0,12)	0,94 (0,22)	0,52 (0,10)	0,94 (0,21)	0,50 (0,09)	0,97 (0,18)	0,55 (0,06)	0,97 (0,12)	0,50 (0,09)	0,88 (0,18)	0,59 (0,07)	0,95 (0,14)
MD 30K	0,57 (0,12)	0,92 (0,21)	0,53 (0,10)	0,96 (0,21)	0,55 (0,08)	0,96 (0,14)	0,58 (0,06)	0,94 (0,12)	0,50 (0,10)	0,88 (0,17)	0,60 (0,07)	0,95 (0,14)

**Annexe - tableau 3 :** Précision moyenne (r) et coefficient de dispersion moyen (b) d'une évaluation GBLUP pour les caractères de reproduction calculés sur 40 simulations du scénario PA- et différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

Dans le nom des puces LD, K indique le nombre de marqueurs en milliers et 05 et 25 indiquent la MAF considérée pour le choix des SNPs et k. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs dans la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

RESEARCH ARTICLE

Open Access

# Genetic architecture and genomic selection of female reproduction traits in rainbow trout



J. D'Ambrosio<sup>1,2</sup>, R. Morvezen<sup>2</sup>, S. Brard-Fudulea<sup>3</sup>, A. Bestin<sup>2</sup>, A. Acin Perez<sup>4</sup>, D. Guéméné<sup>3</sup>, C. Poncet<sup>5</sup>, P. Haffray<sup>2</sup>, M. Dupont-Nivet<sup>1</sup> and F. Phocas<sup>1\*</sup> 

## Abstract

**Background:** Rainbow trout is a significant fish farming species under temperate climates. Female reproduction traits play an important role in the economy of breeding companies with the sale of fertilized eggs. The objectives of this study are threefold: to estimate the genetic parameters of female reproduction traits, to determine the genetic architecture of these traits by the identification of quantitative trait loci (QTL), and to assess the expected efficiency of a pedigree-based selection (BLUP) or genomic selection for these traits.

**Results:** A pedigreed population of 1343 trout were genotyped for 57,000 SNP markers and phenotyped for seven traits at 2 years of age: spawning date, female body weight before and after spawning, the spawn weight and the egg number of the spawn, the egg average weight and average diameter. Genetic parameters were estimated in multi-trait linear animal models. Heritability estimates were moderate, varying from 0.27 to 0.44. The female body weight was not genetically correlated to any of the reproduction traits. Spawn weight showed strong and favourable genetic correlation with the number of eggs in the spawn and individual egg size traits, but the egg number was uncorrelated to the egg size traits. The genome-wide association studies showed that all traits were very polygenic since less than 10% of the genetic variance was explained by the cumulative effects of the QTLs: for any trait, only 2 to 4 QTLs were detected that explained in-between 1 and 3% of the genetic variance. Genomic selection based on a reference population of only one thousand individuals related to candidates would improve the efficiency of BLUP selection from 16 to 37% depending on traits.

**Conclusions:** Our genetic parameter estimates made unlikely the hypothesis that selection for growth could induce any indirect improvement for female reproduction traits. It is thus important to consider direct selection for spawn weight for improving egg production traits in rainbow trout breeding programs. Due to the low proportion of genetic variance explained by the few QTLs detected for each reproduction traits, marker assisted selection cannot be effective. However genomic selection would allow significant gains of accuracy compared to pedigree-based selection.

**Keywords:** Fish, Heritability, GWAS, QTL, Body weight, Spawning date, Spawn weight, Egg weight, Egg size, Fecundity

\* Correspondence: [florence.phocas@inrae.fr](mailto:florence.phocas@inrae.fr)

<sup>1</sup>Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Rainbow trout (*Oncorhynchus mykiss*) is a worldwide cultured salmonid species with numerous breeding programs implemented over the last 50 years in closed populations. Breeding goals, i.e., the number, the nature and the importance given to the selected traits, are an essential feature of a breeding program because this determines the direction and extent of genetic trends in the population under selective breeding. The traits of interest can be recorded either directly on candidates for selection, as in the case of growth and morphology traits or on their sibs when the trait measurement requires a fish to be killed, such as assessment of disease resistance or processing yield. Until the 2000s, commercial lines have been mainly selected for growth traits by mass selection and external morphology assisted by ultrasound to improve gutted yield [1]. Since then, sib-based selection to improve more efficiently gutted and fillet yields and disease resistance traits have been more and more emphasized [2, 3].

The egg productivity of rainbow trout has so far received limited attention in commercial selective breeding because its high relative fecundity (1500 to 2000 eggs / kg of body weight) was not considered as a limiting factor [4]. However these traits play an important role in the economics of a breeding company because of the sale of eyed eggs. In addition, the production of caviar from trout eggs is now a developing market, so an increase in the egg number per ton of fish produced and the limitation of egg size would allow an increase in this production.

Reproduction is the basis of all animal production systems. Maintaining or improving reproductive efficiency is essential to the productivity of animal farming. A high reproductive capacity is also one of the main levers for the genetic improvement of the species. It is therefore important to estimate the performance and the genetic trends for reproduction traits, especially since genetic antagonisms are feared with production traits in all species [5].

High selection for production traits in closed and small broodstock populations of rainbow trout have also induced significant levels of inbreeding [6]. Previous works showed significant inbreeding depression effects in female body weight and spawn weight [7] as well as in egg number and spawning age of rainbow trout [8]. Therefore, it is important to evaluate the performance for those female reproduction traits and the possibilities of selecting them in current broodstock.

Early literature reported estimates of genetic parameters for egg production trait in rainbow trout [9–12] or Atlantic salmon [12] based on full-sib analysis of variance of a limited number of families and populations. Only three publications (in rainbow trout: Su et al. [13,

14]; in Coho salmon: Gall and Neira [15]) reported estimates based on REML procedures and animal models that prevent the well-known issue of biased estimation under full-sib analysis of variance.

Since early 1990's, several breeding programs were developed in France based on combined mass selection on growth and sib selection to improve processing traits and, more recently, disease resistance [1, 16]. The aims of the study were to estimate genetic parameters, to identify quantitative trait loci (QTL) and to assess the efficiency of genomic selection (GS) compared to pedigree-based BLUP selection for female reproduction traits in one of these French rainbow trout selected line. Concerning this last objective, change in selection efficiency was investigated according to two main factors of variation: the size of the reference population and the degree of kinship between reference and candidate populations for selection. Traits under study concerned female body weight, spawning date, fecundity and egg size. As far as we know, it is the first report on the efficiency of (genomic) selection for fish reproduction traits.

## Results

### Genetic parameters

Estimates of genetic parameters are summarized in Table 1. All heritability estimates were in a medium range of values going from 0.27 to 0.44.

Our estimates of correlations for the two measures of female body weight with (FW) or without (PW) consideration of the spawning and coelomic liquid weights showed that the two measures were essentially describing the female own weight since we estimated a phenotypic correlation between FW and PW of 0.96 with a genetic correlation close to unity. Regarding correlations between reproduction traits, it should be first mentioned that genetically speaking, EN and SW on one hand, and EW and ED on the other hand corresponded to almost the same traits with genetic correlations ( $r_g$ ) estimated to 0.86 ( $\pm 0.05$ ) and 0.99 ( $\pm 0.02$ ), respectively. All those high correlations are partly due to the auto-correlations originating from methods of deriving these traits from common measures of biological components, as shown by the high environmental correlations observed between those pairs of traits (0.82 between EN and SW and 0.55 between EW and ED in our study).

Although EN and SW were highly correlated in our study, those two traits did not behave similarly with ED and EW. On the contrary, ED and EW could be considered as two different measures of a unique biological egg size trait, those two traits being associated in a very similar manner to all the other traits in the analysis. SW was significantly and positively associated with egg size traits ( $r_g$  in the range 0.5–0.6), while EN was genetically uncorrelated to egg size traits ( $r_g$  in the range 0.01–0.10)

**Table 1** Heritability (on the diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) for female reproduction and weight traits

Trait	SD	FW	PW	SW	EN	EW	ED
SD	0.27 (0.05)	0.12 (0.15)	0.03 (0.15)	0.34 (0.15)	0.22 (0.16)	0.46 (0.15)	0.51 (0.12)
FW	0.04 (0.04)	0.30 (0.06)	0.99 (0.01)	0.12 (0.15)	0.13 (0.16)	-0.01 (0.16)	-0.01 (0.15)
PW	0.04 (0.04)	0.96 (0.01)	0.33 (0.06)	-0.06 (0.15)	-0.02 (0.16)	-0.10 (0.16)	-0.12 (0.14)
SW	0.05 (0.04)	0.34 (0.03)	0.17 (0.03)	0.32 (0.06)	0.86 (0.05)	0.49 (0.13)	0.59 (0.10)
EN	0.10 (0.04)	0.31 (0.03)	0.17 (0.03)	0.83 (0.01)	0.24 (0.05)	0.01 (0.17)	0.10 (0.15)
EW	0.06 (0.04)	0.09 (0.03)	0.03 (0.03)	0.33 (0.03)	-0.18 (0.03)	0.27 (0.06)	0.99 (0.02)
ED	0.13 (0.04)	0.12 (0.03)	0.05 (0.03)	0.38 (0.03)	0.01 (0.03)	0.70 (0.01)	0.44 (0.07)

Standard errors are given in brackets. When adjusting SW and EN for a constant FW, their genetic correlations with FW dropped to respectively -0.26 ( $\pm 0.14$ ) and -0.17 ( $\pm 0.16$ ), while the phenotypic correlations were close to zero with respective estimates of -0.05 ( $\pm 0.04$ ) and 0.06 ( $\pm 0.04$ )  
SD Spawning date; FW Female body weight; SW Spawn weight; EN Egg number; EW average egg weight; ED Average egg diameter

in our study. Both EN and SW traits were genetically weakly correlated to SD (0.22 and 0.34 respectively). Intermediate positive genetic correlations were also estimated between egg size traits and SD.

Regarding the correlations between SD and PW (or FW), we did not observed any significant phenotypic or genetic associations in our study. The female body weights were also not genetically correlated to any of the egg size traits (EW and ED) or egg quantity traits (EN and SW). Considering FW as a covariate in the

modelling of EN and SW led to negative genetic correlations between FW and the adjusted EN\* and SW\* while regressing towards 0 the corresponding phenotypic correlations.

#### Genetic architecture

With the only exception of SD, no QTL explained over 3% of the genetic variance (Table 2). Therefore, the genetic architecture of all female reproduction traits appeared to be highly polygenic (Fig. 1 and Supplementary

**Table 2** Summary statistics for GWAS for female reproduction traits based on GBLUP and BayesC $\pi$  methods

Trait	Peak SNP			GBLUP		BayesC $\pi$		
	Omy	Identifier	Position (bp)	$-\log(P\text{-value})$	Confidence interval (Mb)	LogBF	Credibility interval (Mb)	Var <sup>a</sup> (%)
SD	6	Affx-88,920,061	38,611,684	11.5	38.61–38.61	18.4	38.61–38.61	7.2
SD	11	Affx-88,936,987	44,670,771			7.4	43.17–45.32	1.9
SD	15	Affx-88,942,247	5,640,348			6.8	4.86–5.64	1.2
SD	27	Affx-88,907,414	8,854,196			6.1	7.04–8.95	1.7
SW	1	Affx-88,918,810	79,053,674			9.4	77.68–79.05	1.3
SW	2	Affx-88,951,720	5,801,797	6.3	5.80–6.18	9.2	5.80–6.18	2.9
SW	2	Affx-88,927,743	17,660,020	6.1	16.29–17.66	8.4	16.29–17.66	1.9
SW	12	Affx-88,950,456	5,349,425	6.9	4.56–6.22	9.2	4.56–6.80	2.5
EN	2	Affx-88,951,720	5,801,797	6.5	5.80–6.18	9.6	5.80–6.18	3.2
EN	8	Affx-88,953,704	50,018,409	5.9	49.64–50.02	8.6	49.64–50.02	1.4
EN	12	Affx-88,950,456	5,349,425	6.2	4.23–6.80	8.0	4.56–6.80	2.7
EW	1	Affx-88,932,157	48,339,143	5.4	48.34–48.60			
EW	1	Affx-88,950,822	64,632,546	7.0	60.50–64.63	6.3	61.59–64.63	1.2
EW	1	Affx-88,915,911	67,153,692	7.2	66.61–68.88			
EW	1	Affx-88,957,820	68,664,516			6.4	66.54–68.88	2.0
ED	1	Affx-88,941,218	68,345,854	6.0	68.35–68.66			
ED	1	Affx-88,942,198	71,813,379	7.1	70.85–71.81	8.5	70.85–71.81	1.9
ED	2	Affx-88,930,958	17,234,307			6.6	16.14–18.25	1.5
ED	8	Affx-88,908,356	47,852,976			6.4	47.84–49.70	1.6
ED	12	Affx-88,930,328	68,199,037			7.5	67.33–68.79	1.1

<sup>a</sup>Var: % of genetic variance explained by all the SNPs included in the QTL credibility interval

SD Spawning date; SW Spawn weight; EN Egg number; EW Egg average weight; ED Egg average diameter

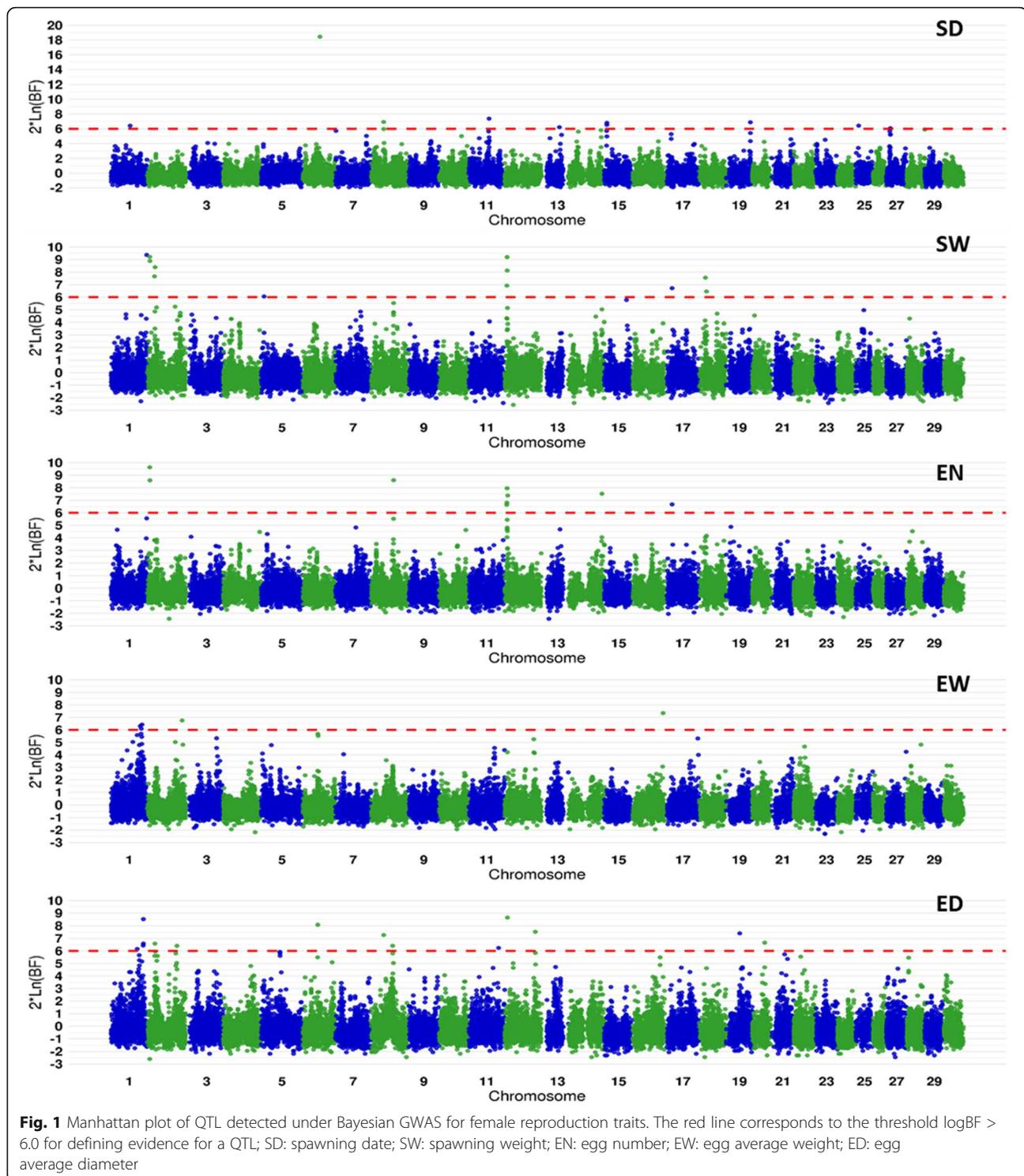


Figure 1, Additional File 1). Only 2 to 4 QTLs were detected that explained at least 1% of the genetic variance for any of the five reproduction traits (Table 2). Ten of the twelve QTLs that were at 1% chromosome-wide significant under GBLUP analysis corresponded to QTLs with strong evidence under the Bayesian approach. In

most cases, the same peak SNPs were given by both studies. The consistency of both GWAS was a bit less clear for EW and ED traits. In total, the Bayesian approach allowed the detection of 17 QTLs that explained at least 1% of the genetic variance, but only nine had a very strong evidence ( $\log_{10}BF > 8$ ).

### Genomic selection

Efficiency of genomic selection in comparison to pedigree BLUP selection is presented in Fig. 2 and Supplementary Table 1, Additional File 2 for both size scenarios (T- and T+) in terms of training population size (Table 3).

Whatever the trait and the scenario, we noticed a slight tendency to overestimate ( $b < 1$ ) the variation of GEBVs compared to the genetic variance although all inflation coefficients were statistically not different from 1 (see Supplementary Table 1, Additional File 2). The GBLUP accuracy was higher than the BLUP accuracy in most of the 40 validation samples (Fig. 2); less than 10% of the samples failed to respect this rule in average over all the traits. In 60% of the 40 samples, GBLUP was more accurate than BLUP for any of the six traits; loss of accuracy of GBLUP was less than 10% in 80% of the remaining samples.

For T- scenario, the mean accuracy of GEBVs varied from 0.50 for EW to 0.60 for ED whereas the corresponding accuracies of EBVs were 0.41 and 0.47 (see Supplementary Table 1, Additional File 2); when comparing GBLUP versus BLUP, the highest increase in accuracy (+ 35%) was obtained for SW\* while the lowest increase (+ 10%) was observed for FW. The same trend was observed for T+ scenario when comparing GBLUP versus BLUP. The mean accuracy of GEBVs varied from 0.55 for EW to 0.66 for ED whereas the corresponding accuracies of EBVs were 0.50 and 0.60 (see Supplementary Table 1, Additional File 2); the highest increase in

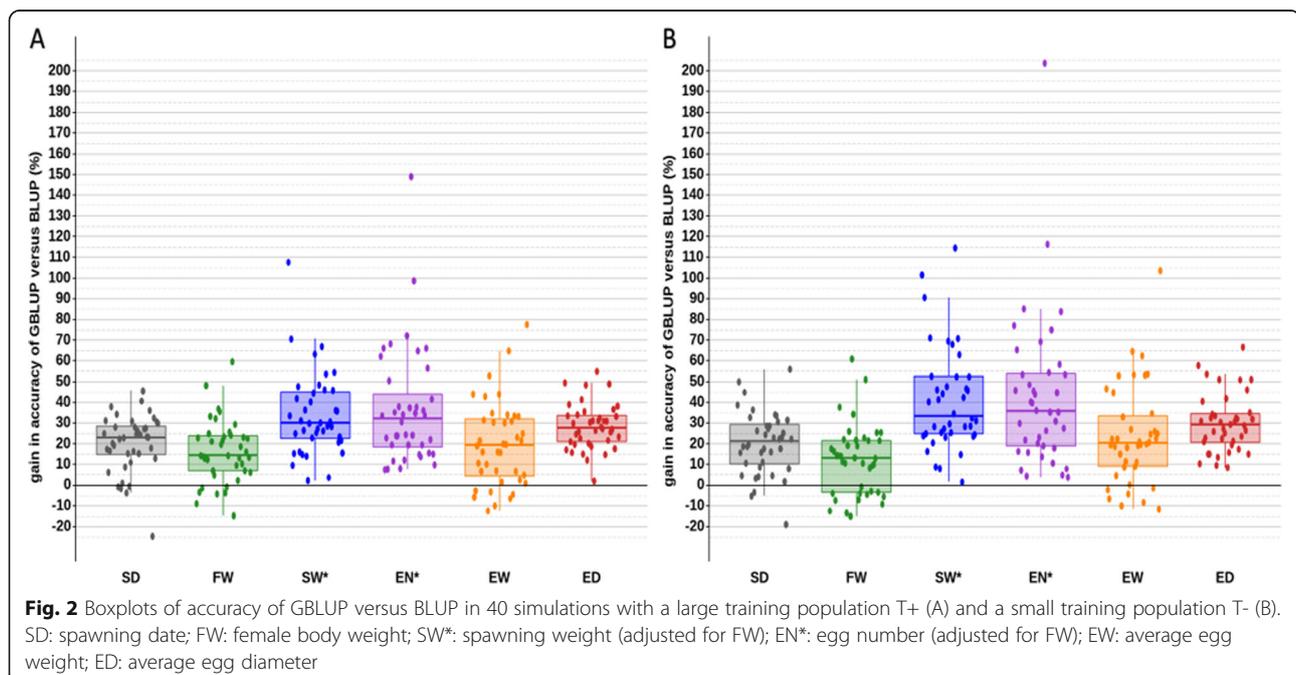
accuracy (+ 32%) was obtained for SW\* while the lowest increase (+ 16%) was observed for FW.

For T1 and T2 scenarios, the mean accuracy of GEBVs varied from 0.24 to 0.39 whereas the corresponding accuracies of EBVs were 0.01 and 0.23 (see Supplementary Table 2, Additional File 2). When comparing selection accuracy (Fig. 3) depending on the degree of kinship between the training and the validation populations (T1 or T2 versus T-), we observed that the accuracy was drastically reduced when the training population was less related to the validation population and that the accuracy loss was more severe for BLUP (- 71% in average over the 6 traits) than for GBLUP (- 43%).

### Discussion

#### Genetic parameters

The estimated heritability values for FW and PW in the present study are within the range (0.27–0.43) of values reported in recent studies in the same species for female body weight between 13 and 25 months [17–19]. In a study that accounted for random full-sib family effects in an animal model, Su et al. [13] estimated lower heritabilities for PW with values ranging from 0.09 for a control line to 0.26 for a line selected for egg size while the estimate was 0.13 for the line selected for yearling body weight; the full-sib family effect was very significant accounting for 22, 13 and 10% of the phenotypic variance for the three lines, respectively. In our study, no significant full-sib family effect was detected whatever the trait under consideration. A maternal random effect was also



**Table 3** Number and cohort origin of animals in training and validation sets for cross-validation tests

Scenario	T+	T-	T1	T2
Training set	1077 fish randomly chosen in C1 + C2	672 fish randomly chosen in C1 + C2	726 fish in C1	620 fish in C2
Validation set	269 remaining fish in C1 + C2	same 269 fish as in scenario T+	620 fish in C2	726 fish in C1

tested for all traits, but were never significant. It is likely due to the mixing of eyed-eggs from all the different families and their rearing in common tanks.

Heritability estimated for SD in our study was moderate in comparison to the large values (0.49 to 0.87 with a pooled heritability of 0.65) estimated by Su et al. [13] for their three rainbow trout lines bred derived from the same experimental base population of University of California. However, Gall and Neira [15] estimated a moderate heritability (0.24) for SD in Coho salmon that may be partially explained by the fact all the genetic variation was not accounted for because the spawned females represented a selected group of earliest spawners. The same explanation may also be considered to understand our own moderate estimate of heritability for SD in rainbow trout.

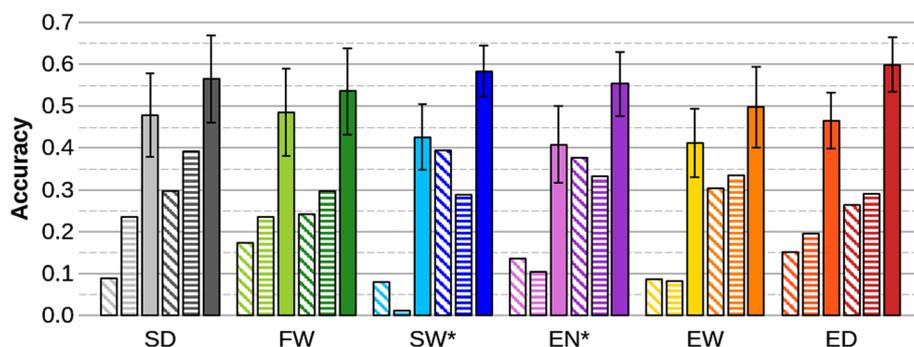
All our heritability estimates for rainbow trout reproductive traits (SD but also PW, SW, EN and EW) were in general close to those of Gall and Neira [15] in Coho salmon, although their heritability for EN was higher (0.42 vs 0.24 in our study) and more consistent with their heritability of SW than in our study. Considering EW as a measure of egg size, Gall and Neira [15] found a similar estimate of heritability as ours. In Su et al. [13], egg size was measured as a volume (in pl), neither corresponding to EW or ED. They found high estimates of heritability for their egg size trait with a pooled heritability of 0.60 across their three rainbow trout lines. Rather than measuring SW, Su et al. [13] measured the egg volume (ml) by allowing the eggs to settle in a volumetric cylinder after water hardening; egg size was determined

by dividing 30 by the count of eggs in a 30 ml sample. Egg number was then obtained by dividing egg volume by egg size and its pooled heritability was very high (0.55) compared to ours and in a lesser extent to Gall and Neira [15]'s one.

Regarding correlations between reproduction traits, the high correlation we estimated between EN and SW was also observed in Coho salmon [15]. However these authors estimated a large negative genetic correlation (-0.63) between EN and EW for Coho salmon while we did not observed any significant correlations between EN and egg size traits in our study. Our result was only consistent with Su et al. [13]' one for this particular point. Therefore the question is still pending whether or not selection for egg size may cause or not a decrease in egg number in salmonids or if the negative correlation that we observed is only population specific.

Considering genetic correlations between egg size traits and SD, our estimates were consistent with the estimate (0.51) in Su et al. [13]' study while Gall and Neira [15] reported a weaker (0.16) genetic correlation in Coho salmon. The positive, but weak, genetic correlations we estimated between egg quantity traits (EN and SW) and SD were also reported by Su et al. [13] and Gall and Neira [15] between EN and SD. However in Coho salmon the genetic correlation estimated between SW and SD was null [15].

Regarding the correlations between SD and female body weights, our results were in contradiction with the significant positive genetic correlation (0.51) reported by Su et al. [13] but very consistent with those for Coho



**Fig. 3** Selection accuracy for BLUP (light colors) or GBLUP (dark colors) considering the training scenarios T1 (diagonal hatching bars), T2 (horizontal hatching bars) and T- (full bars). For scenario T-, the sampling standard deviation over 40 replicates is indicated by a black segment above and below the mean. SD: spawning date; FW: female body weight; SW\*: spawning weight (adjusted for FW); EN\*: egg number (adjusted for FW); EW: average egg weight; ED: average egg diameter

salmon [15]. This may be due to the inclusion of only early spawning females in our study as in Gall and Neira' one.

Although female body weights did not significantly correlated neither to egg quantity traits (EN, SW) or egg size traits (EW, ED) in our study, significant positive genetic correlations were found for PW with EN (0.47), egg size (0.51) and egg volume (0.67) in US lines of rainbow trout [13]. Gall and Neira [15] reported also in Coho salmon that PW correlated moderately with EN (0.32) or EW (0.37), and more strongly with SW (0.56). Therefore our results are not consistent with the two previous studies in salmonid species regarding genetic associations between female body weight and egg production traits.

Earlier studies indicated that selection for body weight at any age would result in a correlated response for body weight at other ages including age at slaughter and age at spawning for females [20] and that genetic correlations between yearling weight, post-spawning weight of females and egg size, egg number and egg volume were moderate but positive in rainbow trout [14]. These findings led to the assumption that breeding programs designed to increase body size would result in improving egg production traits [14]. However our results do not really validate this assumption since FW and PW do not appear to be genetically correlated - neither positively nor negatively - to any of the reproduction traits (SD, SW, EN, EW or ED). It is therefore important to consider direct selection for improving egg production traits in rainbow trout breeding programs.

#### Genetic architecture

The credibility/confidence intervals associated with the QTLs were large in most cases, which precluded the meaningful identification of potential underlying candidate genes that may explain the phenotypic variations observed for female reproduction traits. Nevertheless, based on the functional information given in the human gene database GeneCards® [21, 22] and described phenotypes in mutant mice or worms, we were able to propose five candidate genes for female reproduction traits (3 for SD, 1 for ED and 1 shared by EN and SW). No phenotypes were described for these candidate genes in the zebrafish database ZFIN [23]. As far as we know, our study is the first report for QTLs and candidate genes playing for fecundity and egg size traits in rainbow trout.

Fitness traits such as spawning date and body weight are major factors in the life history of salmonid fishes. Despite the fact that some recent QTLs studies have focused on growth traits in rainbow trout [17, 19] the only significant SNP we detected for female body weight was not in the vicinity of any QTL regions reported for trout body weight. This SNP was not presented in Table 2

because the detected association was considered as a spurious one. There was indeed no evidence for a QTL under GBLUP analysis and only a single SNP was detected on Omy1 with a  $\log_{10}BF > 6$  in an intergenic region between *crocc2* and *slco1f1* genes that are not annotated as playing a role in growth function.

In our study, the highest significant ( $p$ -value  $< 10^{-6}$  at the genome level) SNP was found for SD on Omy6. It corresponds to the only significant QTL for SD under GBLUP analysis and this single SNP explained over 7% of the genetic variance under BayesC approach. The SNP Affx-88,920,061 is located within the *importin-11* gene (alias Ran-Binding Protein 11) that plays a receptor role in nuclear protein import. The phenotypes observed in a MGI mouse strain (ID MGI:5617259) with a mutation in this gene may help to understand the effects of a variant allele may have on spawning date in rainbow trout. The double mutant homozygote is a pre-weaning lethal phenotype in mouse and the heterozygote exhibits decreasing levels of iron and glucose levels in the blood. In addition, the heterozygote male had an abnormal eye (lens) morphology. However we may wonder whether the association is a spurious one due to a very high rate of mendelian errors at this SNP position when comparing genotypes from progeny and their parents. Indeed, 104 mendelian errors were observed at this SNP while the medium number of mendelian errors was only 3 for the 4067 SNPs (out of 27,799 SNPs) with at least one mendelian error detected.

While the genetic architecture underlying SD is still largely unknown, a lot of QTLs affecting SD or age at sexual maturation have been detected in salmonid species during the last 20 years [24–30]. Due to the low density of markers and the non standardisation of linkage group names in early studies, it is difficult to report whether our QTLs may have been detected in other studies. Nevertheless, as far as we know, no QTL for SD in salmonid species has been reported in the neighborhood of the highly significant SNP we observed on Omy6. Under Bayesian GWAS, three others QTLs for SD were detected on Omy11, Omy15 and Omy27. Each of them explained between 1.2 and 1.9% of the genetic variance of SD. There is no obvious candidate gene for the large QTL region on Omy11. On the contrary, it is worth mentioning that the peak SNP on Omy15 is positioned within the *ARHGEF4* (Rho Guanine Nucleotide Exchange Factor 4) gene that acts as guanine nucleotide exchange factor (GEF) for RHOA, RAC1 and CDC42 GTPases. MGI mutant phenotypes for *ARHGEF4* concern in particular immune system and metabolism with decreased hemoglobin content, decreased IgE level, decreased IgG1 level, decreased T cell number, increased mature B cell number, increased circulating alkaline phosphatase level, increased circulating total protein

level and decreased circulating triglyceride level. Despite the large credibility interval for the last QTL on *Omy27*, we may also propose *NR2E1* as a convincing candidate gene because the peak SNP is located just behind this estrogen-related receptor gamma-like gene. This gene is an orphan receptor that binds DNA as a monomer to hormone response elements and is in particular involved in the regulation of retinal development and essential for vision. It may be involved in retinoic acid receptor regulation in retinal cells. MGI mutant phenotypes for *NR2E1* have abnormal optic nerve and retina morphology, abnormal brain morphology, decreased body size and total body fat amount and decreased female fertility. Interestingly, we can hypothesize that two of the four QTLs detected for SD may be associated to abnormal eye morphology and defects in vision that may render trout less sensitive to the photoperiod stimuli.

Concerning the fecundity traits EN and SW, two QTLs sharing the same SNP peaks on *Omy2* (Affx-88,951,720) and *Omy12* (Affx-88,950,456) were detected, confirming that the two traits are biologically very close. The QTL on *Omy2* explained 3.2 and 3.0% of the genetic variance for EN and SW, respectively. The QTL on *Omy12* explained about 2.5–2.7% of the genetic variance for each trait. For EN, a third QTL was detected on *Omy8* that explained 1.4% of the genetic variance. No obvious gene candidate could be proposed for this QTL. For SW, a third QTL was detected on *Omy2* that explained 1.9% of the genetic variance and a last QTL was detected on *Omy1* that explained 1.3% of the genetic variance. These last two QTLs for SW had large credibility intervals (> 1 Mb) and no obvious candidate gene could be proposed. Concerning the common peak SNP for the QTL on *Omy2* shared by EN and SW, its location is in-between the *MPRD* gene (cation-dependent mannose-6-phosphate receptor-like) and the *PHC1* gene (polyhomeotic-like protein 1). This *PHC1* gene is a homolog of the *Drosophila* polyhomeotic gene, which is a member of the Polycomb group of genes. It is a component of a Polycomb group (PcG) multiprotein PRC1-like complex, a complex class required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development [31]. MGI homozygous mutant phenotypes for the *PHC1* gene exhibit perinatal lethality, posterior skeletal transformations and defects in neural crest derived tissues, including ocular abnormalities, cleft palate, parathyroid and thymic hypoplasia and cardiac anomalies. We hypothesize that this gene may have a role in rainbow trout fecundity.

For egg size traits measured by EW and ED respectively, we will focus on the three QTLs on *Omy1* showing the highest consistency of results across GWAS (Table 2). Among those QTLs, the same QTL region on *Omy1*, spanning between 66.539 Mb and 68.881 Mb was

detected for the two traits, although the peak SNPs were different across GWAS (67.15 Mb for GBLUP and 68.66 Mb for BayesC $\pi$ ). No obvious candidate gene could be proposed in this large QTL region that explained 2% of the genetic variance for EW.

Regarding EW, there was a strong evidence for a distinct QTL explaining 1.2% of the genetic variance, in the region spanning from 60.498 Mb and 64.633 Mb on *Omy1* with the peak SNP very close to 64.633 Mb in an uncharacterized protein (LOC110527930). No candidate gene could be proposed within this QTL region, but in the close vicinity of this QTL region, let us mention the presence of the *prkg2* gene (located between 64.660 and 64.676 Mb on *Omy1*) whose role is important in oocyte maturation in mammals and zebrafish [32]. This *prkg2* gene may be therefore suggested as a functional candidate gene.

Regarding ED, there was evidence for another QTL on *Omy1* that explained nearly 2% of the genetic variance in the region spanning between 70.848 Mb and 71.813 Mb. The peak SNP (at 71.813 Mb) is close to the position (71.675–71.699 Mb) of the *WAPLA* gene (wings apart-like protein homolog) which is a straightforward gene candidate for explaining this last QTL on *Omy1*. Indeed this gene is a regulator of meiotic chromosome structure and function, playing a role in sister chromatid cohesion, cohesin association with chromatin, DNA double strand break repair and polar body positioning following meiotic divisions during oogenesis [33, 34]. Worm *C. elegans* mutants have an egg-laying defect and reduced brood size with 21% displaying embryonic lethality whilst 28% arrest at the larval stage [34].

### Genomic selection

The implementation of genomic selection in a breeding program for female reproduction traits would allow a gain in accuracy of 16 to 32% (depending on the trait studied) compared to BLUP selection when considering a reference population of about 1100 individuals. For a smaller reference population of 670 individuals, genomic selection will also be more efficient than pedigree-based selection. In the literature, several studies in salmonids [17, 35–38] reported moderate to strong gains in accuracy (+ 11% to 110%) for genomic selection compared to BLUP selection of depending on the genetic architecture of the traits and the size of the reference populations. Increasing by 60% the size of the training population (starting from 670 individuals) in a fish line whose effective population size is estimated around 50 [6], increased the accuracy of GEBVs by 6% to 11% depending on the trait considered. Considering Goddard [39] prediction equation of GS accuracy, we could have expect an increase of accuracy of about 20% in our rainbow trout line when increase our training population from

670 to 1070 animals. Significantly higher gain in accuracy (about +80%) was described for gestation length ( $h^2 = 0.4$ ) in a pig line when increasing by 130% the training population size (starting from 550 sows) in a Large White line with effective population size estimated to be about 100 [40].

We showed a severe loss of accuracy when the degree of kinship between the training and the validation populations was less related and this was more important for BLUP than GBLUP. It is well known that close relationships between animals in the training and validation sets increase the accuracy of genomic predictions compared to the ones derived for an independent validation population [41]. Under scenarios T1 and T2, our estimates of GS accuracy are close to the theoretical estimates derived from Goddard [39]' formula, which makes sense since one of the basic assumption beyond his formula is that GS accuracy comes only from linkage disequilibrium across the whole population and not from any linkage association and family structure. The fact that the loss of accuracy was less severe for GBLUP than for BLUP had also been shown on poultry [42]. This phenomenon has also been quantified in a salmon population dedicated to genomic selection for sea lice resistance [36]. What is less known is that these strong decreases in accuracy are also associated with strong bias in the predictions (see Supplementary Table 2, Additional File 2). While the variance of GEBVs was not significantly overestimated when the training and validation sets were closely related (see Supplementary Table 1, Additional File 2), it became either strongly inflated or deflated when relationships became more distant. The same observation held for BLUP predictions, even in a larger extent. This may be a strong issue to correctly predict genetic trends or performing optimal multitrait index selection since the magnitude of the inflation varied a lot across traits. Nevertheless, when performing selection within cohort based on a training population including sibs of the candidates, it does not appear to be a problem.

## Conclusions

In our study, all female reproduction and weight traits were moderately heritable with spawn weight showing strong and favorable genetic correlations with number of eggs in the spawn and individual egg size traits (egg average diameter or weight). On the contrary, number of eggs in the spawn was uncorrelated to egg size traits and female body weight (just before or after spawning) was not genetically correlated to any of the reproduction traits. Therefore it is unlikely that selection for growth will induce any indirect improvement for female reproduction traits. It is thus important to consider

direct selection for spawn weight for improving egg production traits in rainbow trout breeding programs.

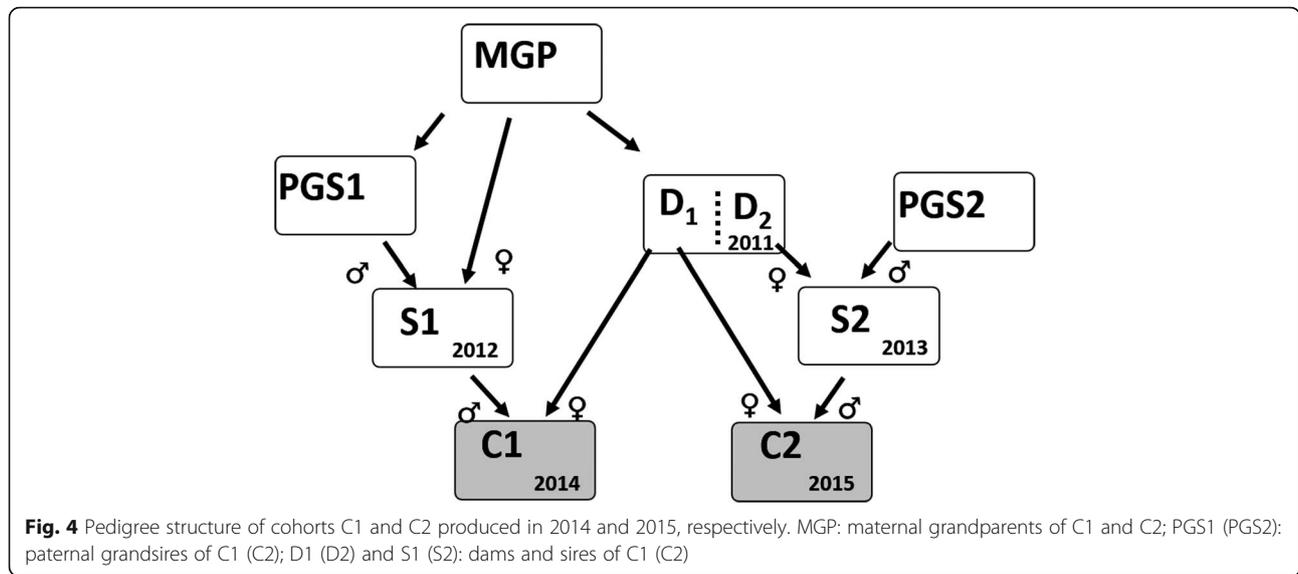
The GWAS results suggested that female reproduction traits are highly polygenic. Only six QTLs over the 19 identified across the five traits studied explained at least 2% of the genetic variance. These results suggest that gene-assisted selection will be useless for improving reproduction traits. However, genomic selection based on a reference population of only one thousand individuals related to candidates may improve the efficiency of BLUP selection from 16 to 37% depending on female reproduction traits. Cross-validation test of genomic prediction highlighted the clear increase in prediction accuracy compared to that of pedigree-based prediction in almost all population samples. The accuracy of GBLUP was the highest when training and validation sets were closely related but the relative advantage over pedigree-based prediction within a population was the largest when relationships were more distant.

## Methods

### Phenotypes

Phenotypes were collected at 2 years of age in females from two successive cohorts produced in 2014 and 2015, hereafter named C1 and C2, and composing the 9th generation of selection of the trout breeding company "Viviers de Sarrance". Those cohorts constituted the broodstock of the company and came from two related paternal cohorts S1 and S2 and from different groups of dams D1 and D2 of the same maternal cohort produced in 2011 (Fig. 4). The animals used in the study were reared at the French farm "Viviers de Sarrance" (Pisciculture Labedan, 64,490 Sarrance, France). They were released after the data and egg collection and they reproduced a second time the year after the study.

Raw phenotypes collected were the weight of the ready-to-spawn female (FW), its post-spawning weight (PW), the weight of the total egg mass hereafter called the spawn weight (SW), the length of 50 eggs aligned along a graduated rule, the weight and the number of eggs in a sampling spoon of 2.5 ml, the spawning week number in the calendar year and the presence of over-mature eggs in the spawn. All performance records which were more than 4 standard deviations from the mean in absolute value were considered as outliers and discarded from the study. Analyzed traits were the raw phenotypes FW, PW and SW and four derived phenotypes: egg average diameter (ED), egg average weight (EW), egg number of the spawn (EN) and spawning date (SD). ED was calculated as the length of 50 aligned eggs divided by 50. EW was derived as the ratio of the weight of eggs to the number of eggs in the sample of 2.5 ml. EN was calculated as the ratio of SW to EW. SD corresponded 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 after weeks) within cohort. In total, the phenotypes of 1517 fish were considered in the study (Table 4).

**Genotypes**

Among the 1517 phenotyped fish, 1346 fish (726 and 620 individuals from C1 and C2 cohorts, respectively) were genotyped for 57,501 SNPs (Single Nucleotide Polymorphism markers) with the Axiom™ Trout Genotyping array [43] at the INRAE genotyping Platform Gentyane. Most of the fish have their parents also genotyped for the 57,501 SNPs. Only 45 phenotyped fish had ungenotyped dams and 2 other phenotyped fish had ungenotyped sires. Those fish were retained for the genetic analysis considering either unknown dam or unknown sire in their pedigree. The remaining 1301 fish were produced by 83 dams and 71 sires within 7 factorial plans of 12 dams mated to 10 or 11 sires. The genomic relationships among fish were in average 0.05 between cohorts C1 and C2, 0.08 within cohort C1 and 0.07 within cohort C2. The highest genomic relationship

values were 0.30 across cohorts, 0.69 and 0.66 within cohort C1 and C2, respectively.

Quality controls of genotyped SNPs were performed as described in D'Ambrosio et al. [6] in particular to remove SNPs with probe polymorphism and multiple locations on the genome assembly (accession number: GCF\_002163495.1). Only the 29,799 SNPs with a call rate higher than 0.97, a test of deviation from Hardy-Weinberg equilibrium with a *p*-value > 0.0001 and a minor allele frequency higher than 0.05 were retained for the analysis. All missing genotypes for the 29,799 SNPs were imputed using the FImpute software [44].

**Pedigree-based BLUP-animal model and estimation of genetic parameters**

Mixed linear BLUP - animal models were used to get the estimated breeding values (EBV) and to estimate genetic parameters based on pedigree information.

For any trait *i* among the six reproduction traits considered in the study, the following statistical model was considered to describe the vector of performance  $y_i$  of the 1346 fish:

**Table 4** Summary statistics of 2-year female reproduction and weight traits in the rainbow trout broodstock

Trait	Number	Mean	SD	Min	Max
Spawning date (SD, week rank)	1517	2.52	1.54	1	5
Ready-to-spawn female body weight (FW, g)	1516	2067	448	1010	3580
Post-spawning body weight (PW, g)	1515	1845	408	840	3116
Spawn weight (SW, g)	1517	187	71	12	398
Egg number (EN, #)	1507	4710	1743	577	10,434
Average egg weight (EW, mg)	1501	39.9	6.5	20.0	68.1
Average egg diameter (ED, mm)	1504	4.03	0.19	3.32	4.72

SD Standard deviation

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \quad (1)$$

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

For all the traits, the cohort group (C1 or C2) was considered as the main fixed effect. For the traits FW, SW, EN and ED, an additional fixed effect due to the presence of overmature eggs was significant and therefore considered in the models. The week of spawning had also a significant effect on the traits FW, PW, SW, EN, EW and ED and was considered as a covariate factor nested within cohort.

Because the targeted breeding goal was to increase egg production at a constant weight of female, additional genetic and genomic analyses of SW and EN were performed considering those two traits adjusted (SW\* and EN\*) for female weight (considered as a covariate factor within cohort).

Tracing back over 8 generations the pedigree of the 1346 phenotyped fish, the vector  $\mathbf{u}_i$  corresponded to the breeding values of 15,265 individuals related through the pedigree relationship matrix  $\mathbf{A}$ .

To estimate all variance components by Average Information Restricted Maximum Likelihood algorithm [45], eight multi-trait models were run considering four traits in joint analyses. With the female body weight (BW) representing FW in a first set of 4 multi-trait analyses and PW in a second set of 4 multi-trait analyses, the following combinations of traits were considered in the joint analyses: [BW, SD, EW, SW], [BW, SD, EW, EN], [BW, SD, ED, EN] and [BW, SD, ED, SW]. For a given BW, results for any genetic or phenotypic parameter were the averages of 4 estimates, except for the correlations between FW and PW, SW and EN, EW and ED that were estimated under unique bivariate models due to convergence issues related to the very high correlations between those pairs of traits.

The EBV were estimated using BLUPf90 package and the variance components using AIREMLf90 program [46].

#### Genomic BLUP model and QTL detection

Genomic BLUP (GBLUP) is an analogous approach to BLUP considering a genomic related matrix  $\mathbf{G}$  [47] in place of a pedigree matrix  $\mathbf{A}$ . The performance modeling is the same as for the BLUP model described in eq. (1), but only the performance of the genotyped animals can be integrated into a conventional GBLUP analysis:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{Z}_i\mathbf{g}_i + \mathbf{e}_i \quad (2)$$

with the vector  $\mathbf{g}_i$  corresponding to the breeding values of 1346 phenotyped and genotyped individuals related through the genomic relationship matrix  $\mathbf{G}$ .

The  $\mathbf{G}$  matrix and the genomic estimated breeding values (GEBV) were estimated using BLUPf90 package [48].

Genome-wide association studies (GWAS) were performed considering the GBLUP models to identify QTL. The postGSF90 module [49] of the BLUPf90 package makes it possible to obtain the estimated effects of the SNPs ( $\hat{\mathbf{a}}_i$ ) from the genomic breeding values  $\hat{\mathbf{g}}_i$  predicted for the genotyped animals according to the equation:

$$\hat{\mathbf{a}}_i = \mathbf{d} \mathbf{Z}' [\mathbf{Z} \mathbf{d} \mathbf{Z}']^{-1} \hat{\mathbf{g}}_i \quad (3)$$

where  $\mathbf{d}$  is the vector of weights associated with the SNP effects.

A region of the genome was considered to be a QTL when the  $-\log_{10}(p\text{-value})$  for a SNP of this region was equal or greater than 5.0 (which corresponds to a chromosome-wide significance threshold of 1% derived as  $-\log_{10}(0.01/(n/30))$  after Bonferroni correction with  $n = 29,799$  the total number of SNPs included in the analysis). The peak SNP was considered as very significant when its  $-\log_{10}(p\text{-value})$  was at least equal to  $-\log_{10}(0.01/n) = 6.5$ , corresponding to a genome-wide significance threshold of 1% after Bonferroni correction. Considering a drop-off value of 1.5 log unit of p-value [50], the QTL confidence interval was delimited by integrating all the SNPs into a 1 Mb sliding window around the peak SNP as long as the new SNPs exhibited a  $-\log_{10}(p\text{-value})$  greater than the peak value  $-1.5$ .

A Bayes C $\pi$  strategy [51] was used to confirm those QTLs and to better estimate the variance explained by those QTLs. Therefore, a general linear mixture model was defined in which a fraction  $\pi$  of the 30 K SNPs was assumed to have a non-zero effect at each cycle of the MCMC algorithm. A total of 100,000 cycles were performed, with a burn-in period of 5000 cycles. Results were saved every 20 cycles. Convergence was assessed by visual inspection of plots of the posterior density of genetic and residual variances and by deriving high correlations ( $r > 0.99$ ) between GEBVs estimated from different chains of the MCMC algorithm. Assuming for  $\pi$  a beta distribution  $B(\alpha, \beta)$  with  $\alpha = 300$  and  $\beta = 29,800$ , the  $\pi$  value was kept almost constant at 1%, corresponding to approximately 300 SNPs selected at each iteration among the 29,800 markers. By trial-and-error, this  $\pi$  value was considered as a good compromise in our variable selection algorithm between the high degree of polygeny of the quantitative traits under study and the limited number of individuals ( $n \sim 1300$ ) in our dataset that led to consider  $p = 300 < n$  to correctly estimate p SNP effects simultaneously. We used the BESSiE software [52] to perform the Bayesian analyses.

The degree of association between each SNP and phenotypes was assessed with the Bayes Factor (BF) that

involves  $\pi$  and  $P_i$ , the probability of the  $i$ th SNP to have a non-zero effect:  $BF = \frac{P_i/(1 - P_i)}{\pi/(1 - \pi)}$

As proposed by Kass and Raftery [53], the logBF was computed as twice the natural logarithm of the BF and the threshold  $\log BF \geq 6$  was used for defining evidence for a QTL. As proposed by Michenet et al. [54], a credibility interval was built around the peak SNP integrating to the QTL region the SNPs with  $\log BF \geq 3$  that were located close to the peak SNP using a sliding window of 1 Mb on both sides of the peak SNP. Following Michenet et al. [54], we considered that a peak SNP with  $6 \leq \log BF < 8$  corresponded only to a putative QTL unless the QTL region explained at least 1% of the genetic variance; in that case, the QTL was considered as having a strong effect on the trait of interest.

Candidate genes located within the confidence or credibility intervals estimated using either GBLUP or BC $\pi$  analysis were listed from the NCBI *Oncorhynchus mykiss* Annotation Release 100 (GCF\_002163495.1).

#### Criteria of validation of selection efficiency

To assess the BLUP or GBLUP selection efficiency, cross-validation tests were performed for different scenarios varying the training population size and its relatedness to the validation population. Efficiency was assessed by the accuracy of selection and the inflation coefficient of EBV as a measure of selection bias. The accuracy of selection was derived as the correlation between (G) EBVs of individuals in the validation set and their corrected phenotypes (adjusted for the covariate and fixed effects) divided by the square root of the trait heritability [55]. The inflation coefficient was derived as the regression coefficient of the corrected phenotypes on the (G)EBVs. In the absence of selection bias, this coefficient is expected to be equal to 1; the coefficient value is below 1 in case of EBV over-dispersion (inflation) and the value is above 1 in case of EBV under-dispersion.

To test the training size effect, 40 replicates of Monte-Carlo 'leave-one-group-out' cross-validation tests [56] were run. For a given replicate, 269 fish for the validation set, 672 fish for the small training set (T-) and 1077 fish for the large training set (T+) were randomly chosen across the 1346 individuals phenotyped and genotyped. For the T+ and T- scenarios, accuracy of selection and inflation coefficient were derived as the mean over the 40 replicates of the correlation and regression coefficients previously described between (G) EBV and corrected phenotypes for the validation population.

To test the relationship effect between the training and validation populations, the C1 and C2 fish were alternatively used as the training and validation sets, the scenarios T1 and T2 corresponding to C1 and C2 fish in the training sets, respectively (Table 3).

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-06955-7>.

**Additional file 1 Supplementary Figure 1.** Manhattan plot of QTL detected under GBLUP-based GWAS for female reproduction traits. The red line corresponds to the genome-wide significance threshold at 1% after Bonferroni correction and the blue line corresponds to the chromosome-wide significance threshold at 1% after Bonferroni correction. SD: spawning date; SW: spawn weight; EN: egg number; EW: average egg weight; ED: average egg diameter

**Additional file 2 Supplementary Table 1.** Mean and standard deviation over 40 replicates (in brackets) of selection accuracy ( $r$ ) and inflation coefficient ( $b$ ) of EBVs and GEBVs for training scenarios T+ and T-. SD: spawning date; FW: female body weight; SW\*: spawn weight (adjusted for FW); EN\*: egg number (adjusted for FW); EW: average egg weight; ED: average egg diameter. **Supplementary Table 2.** Selection accuracy ( $r$ ) and inflation coefficient ( $b$ ) of EBVs and GEBVs for training scenarios T1 and T2. SD: spawning date; FW: female body weight; SW\*: spawn weight (adjusted for FW); EN\*: egg number (adjusted for FW); EW: average egg weight; ED: average egg diameter.

#### Abbreviations

BLUP: Best linear unbiased predictor; EBV: Estimated breeding value; GEBV: Genomic estimated breeding value; GBLUP: Genomic best linear unbiased predictor; GS: Genomic selection; GWAS: Genome-wide association study; MAF: Minor allele frequency; Omy: Chromosome abbreviation for the species *Oncorhynchus mykiss*; QTL: Quantitative trait loci; REML: Restricted maximum likelihood; SNP: Single nucleotide polymorphism

#### Acknowledgements

We thank the staff of the French breeding company Viviers de Sarrance for their participation in the fish rearing and phenotyping and also for providing biological samples.

#### Authors' contributions

JDA performed the curation, the analysis and interpretation of the data and was involved in the script writing, the interpretation of the results and the writing of the original draft. RM and SBF were involved in the sampling of the animals, the data investigation and the supervision of the work. AB was involved in the curation and the investigation of the data. AAP performed the sampling and phenotyping of animals and was involved in the data investigation. CP managed the animal genotyping and was involved in the data investigation. DG was involved in the funding acquisition and conception of the work. PH managed the funding acquisition and the administration of the project and was involved in the conception and supervision of the work. MD was involved in the conception and supervision of the work. FP was involved in the conception, choice of design and statistical methodology, results interpretation and supervision of the work, and wrote the original draft. All authors revised the draft and approved the final manuscript.

#### Funding

This work was supported by the ANRT (N°2017/0239) the European Maritime and Fisheries Fund (grant: RFEA47 0016 FA 1000016) and FranceAgrimer (grant: n°2017-0239) for funding J. D'Ambrosio's PhD thesis, the animal phenotyping and genotyping, the data analysis and the communication of the results. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

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](mailto:florence.phocas@inrae.fr)) and Ana Acin-Perez ([ana@sarrance.com](mailto:ana@sarrance.com)) on request via a material transfer agreement and with permission of the breeding company « Viviers de Sarrance ».

The accession numbers for the rainbow trout reference sequence assembly and the corresponding gene bank assembly were GCF\_002163495.1 and GCA\_002163495.1 in the National Center for Biotechnology Information database that can be accessed at the weblink: [https://www.ncbi.nlm.nih.gov/genome/196?genome\\_assembly\\_id=319782](https://www.ncbi.nlm.nih.gov/genome/196?genome_assembly_id=319782)

Functional information for genes was extracted from the human gene database GeneCards® at the weblink: <https://www.genecards.org/>

#### Ethics approval and consent to participate

This study used on-field data and fin clips collected by the breeding company « Viviers de Sarrance » as part of their commercial breeding program. Fish were reared at the French farm « Les Viviers de Sarrance » (Pisciculture Labedan, 64490 Sarrance, France) under conditions complying with the Directive 98/58/CE on the protection of animals kept for farming purposes. Article 1.5 in the EU Directive 2010/63/ EU on the protection of animals used for scientific purposes excludes practices undertaken for the purposes of recognised animal husbandry and identification from the scope covered by the Directive 2010/63/ EU. As such, the project was not subjected to oversight by an institutional ethic committee.

#### Consent for publication

Not applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France. <sup>2</sup>SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, 35042 Rennes cedex, France. <sup>3</sup>SYSAAF, Section Avicole, Centre INRAE Val de Loire, 37380 Nouzilly, France. <sup>4</sup>Viviers de Sarrance, Pisciculture Labedan, 64490 Sarrance, France. <sup>5</sup>Université Clermont-Auvergne, INRAE, GDEC, 63039 Clermont-Ferrand, France.

Received: 4 April 2020 Accepted: 27 July 2020

Published online: 14 August 2020

#### References

- Haffray P, Bugeon J, Rivard Q, Quittet B, Puyo S, Allamelou JM, et al. Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 2013;410–411:236–44. <https://doi.org/10.1016/j.aquaculture.2013.06.016>.
- Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Consortium A, et al. A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquac Int*. 2016;24:1287–307.
- Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, et al. Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18:1–23.
- Jalabert B, Fostier A. La truite arc-en-ciel: De la biologie à l'élevage. 2010. <http://sbiproxy.uqac.ca/login?url=http://international.scholarvox.com/book/45008232>.
- Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ. Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livest Prod Sci*. 1998;56:15–33.
- D'Ambrosio J, Phocas F, Haffray P, Bestin A, Brard-Fudulea S, Poncet C, et al. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genet Sel Evol*. 2019;51:26.
- Kincaid HL. Inbreeding in fish populations used for aquaculture. *Aquaculture*. 1983;33:215–27.
- Su GS, Liljedahl LE, Gall GAE. Effects of inbreeding on growth and reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1996;142:139–48.
- Gall GA. Genetics of reproduction in domesticated rainbow trout. *J Anim Sci*. 1975;40:19–28.
- Gall GAE, Gross SJ. A genetics analysis of the performance of three rainbow trout broodstocks. *Aquaculture*. 1978;15:113–27.
- Gall GAE, Huang N. Heritability and selection schemes for rainbow trout: female reproductive performance. *Aquaculture*. 1988;73:57–66.
- Gjedrem T, Haus E, Halseth V. Genetic variation in reproductive traits in Atlantic salmon and rainbow trout. *Aquaculture*. 1986;57:369.
- Su GS, Liljedahl LE, Gall GAE. Genetic and environmental variation of female reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1997;154:115–24.
- Su GS, Liljedahl LE, Gall GAE. Genetic correlations between body weight at different ages and with reproductive traits in rainbow trout. *Aquaculture*. 2002;213:85–94.
- Gall GAE, Neira R. Genetic analysis of female reproduction traits of farmed coho salmon (*Oncorhynchus kisutch*). *Aquaculture*. 2004;234:143–54.
- Vandeputte M, Haffray P. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front Genet*. 2014;5(DEC):1–8.
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front Genet*. 2016;7(NOV). <https://doi.org/10.3389/fgene.2016.00203>.
- Leeds TD, Vallejo RL, Weber GM, Gonzalez-Pena D, Silverstein JT. Response to five generations of selection for growth performance traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 2016;465:341–51. <https://doi.org/10.1016/j.aquaculture.2016.08.036>.
- Neto RVR, Yoshida GM, Lhorente JP, Yáñez JM. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Mol Gen Genomics*. 2019;294:563–71. <https://doi.org/10.1007/s00438-018-1518-2>.
- Crandell PA, Gall GAE. The genetics of body weight and its effect on early maturity based on individually tagged rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1993;117:77–93.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards : kmegrat g information about genes , prote and diseases. *Trends Genet*. 1997;13:4.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics*. 1998;14:656–64.
- Sprague J, Bayraktaroglu L, Bradford Y, Conlin T, Dunn N, Fashena D, et al. The Zebrafish Information network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res*. 2008;36(SUPPL. 1):768–72.
- Sakamoto T, Danzmann RG, Okamoto N, Ferguson MM, Ihssen PE. Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1999;173:33–43.
- O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM. Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *J Hered*. 2003;94:273–84.
- Leder EH, Danzmann RG, Ferguson MM. The candidate gene, clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *J Hered*. 2006;97:74–80.
- Moghadam HK, Poissant J, Fotherby H, Haidle L, Ferguson MM, Danzmann RG. Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Mol Gen Genomics*. 2007;277:647–61.
- Haidle L, Janssen JE, Gharbi K, Moghadam HK, Ferguson MM, Danzmann RG. Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*). *Mar Biotechnol*. 2008;10:579–92.
- Küttner E, Moghadam HK, Skúlason S, Danzmann RG, Ferguson MM. Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus*). *Mol Gen Genomics*. 2011;286:67–79.
- Gutierrez AP, Lubieniecki KP, Fukui S, Wither RE, Swift B, Davidson WS. Detection of quantitative trait loci (QTL) related to Grilising and late sexual maturation in Atlantic Salmon (*Salmo salar*). *Mar Biotechnol*. 2014;16:103–10.
- Grimaud C, Bantignies F, Pal-Bhadra M, Ghana P, Bhadra U, Cavalli G. RNAi components are required for nuclear clustering of polycomb group response elements. *Cell*. 2006;124:957–71.
- Li J, Wang Y, Zhou W, Li X, Chen H. The role of PKG in oocyte maturation of zebrafish. *Biochem Biophys Res Commun*. 2018;505:530–5.
- Stanvitch G, Moore LL. Cin-4, a gene with homology to topoisomerase II, is required for centromere resolution by cohesin removal from sister kinetochores during mitosis. *Genetics*. 2008;178:83–97.
- Crawley O, Barroso C, Testori S, Ferrandiz N, Silva N, Castellano-Pozo M, et al. Cohesin-interacting protein WAPL-1 regulates meiotic chromosome

- structure and cohesion by antagonizing specific cohesin complexes. *Elife*. 2016;5:1–26.
35. Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, et al. 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*. 2017; 49:1–13.
  36. Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, et al. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol*. 2016;48:1–11.
  37. Ødegård J, Moen T, Santi N, Korsvoll SA, Kjølglum S, Meuwisse THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front Genet*. 2014;5(NOV):1–8.
  38. Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015;16:1–9. <https://doi.org/10.1186/s12864-015-2117-9>.
  39. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136:245–57.
  40. Hidalgo AM, Bastiaansen JWM, Lopes MS, Calus MPL, de Koning DJ. Accuracy of genomic prediction of purebreds for cross bred performance in pigs. *J Anim Breed Genet*. 2016;133:443–51.
  41. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol*. 2010;42:1–12.
  42. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, et al. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet Sel Evol*. 2011;43:1–8.
  43. Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, et al. The development and characterization of a 57K SNP array for rainbow trout. *Mol Ecol Resour*. 2015; 15:662–72. <https://doi.org/10.1111/1755-0998.12337>.
  44. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
  45. Gilmour AR, Thompson R, Cullis BR. Average Information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 1995;51:1440–1450.
  46. Misztal I, Tsuruta S, Strabel T, Druet T, Lee D. BLUPF90 and related programs (BGF90). *Proc 7th World Congr Genet Appl Livest Prod*. 2002;28:743.
  47. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23. <https://doi.org/10.3168/jds.2007-0980>.
  48. Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Athens Univ Georg. 2014;142p.
  49. Aguilar I, Tsuruta S, Legarra A. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. 10th WCGALP. 2014;Vancouver Canada.
  50. Li H. A quick method to calculate QTL confidence interval. *J Genet*. 2011;90. <https://doi.org/10.1007/s12041-011-0077-7>.
  51. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12:186.
  52. Boerner V, Tier B. BESSIE: a software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. *Genet Sel Evol*. 2016;48:1–5.
  53. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90:773–95.
  54. Michenet A, Barbat M, Saintilan R, Venot E, Phocas F. Detection of quantitative trait loci for maternal traits using high-density genotypes of blonde d'Aquitaine beef cattle. *BMC Genet*. 2016;17:1–13. <https://doi.org/10.1186/s12863-016-0397-y>.
  55. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. *Genetics*. 2008;180:611–8.
  56. Kuhn M, Johnson K. Chapter 4: over-fitting and model tuning. In: *Applied Predictive Modeling*. New York: Springer; 2013. p. 61–89.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



## 6. Références

- Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, et al. Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18(1):1–23. <https://doi.org/10.1186/s12864-017-3557-1>
- AGRESTE. Enquête Aquaculture 2018 - Recensement Pisciculture, culture d'algues et de cyanobactéries 2018 [internet : <https://agreste.agriculture.gouv.fr/agreste-web/disaron/ChdAgri2003/detail/>]. Vol. 3, Chiffres et données. 2020 May.
- Aguilar I, Tsuruta S, Legarra A. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. 10th WCGALP. 2014;Vancouver(Canada). <https://doi.org/10.13140/2.1.4801.5045>
- Boerner V, Tier B. BESSiE: A software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. *Genet Sel Evol*. 2016;48(1):1–5. <https://doi.org/10.1186/s12711-016-0241-x>
- Callet T, Médale F, Larroquet L, Surget A, Aguirre P, Kerneis T, et al. Successful selection of rainbow trout (*Oncorhynchus mykiss*) on their ability to grow with a diet completely devoid of fishmeal and fish oil, and correlated changes in nutritional traits. *PLoS One*. 2017;12(10):1–21. <https://doi.org/10.1371/journal.pone.0186705>
- Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Aquatrace Consortium, et al. A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquac Int*. 2016;24(5):1287–307. <https://doi.org/10.1007/s10499-016-9985-0>
- Chevassus B, Quillet E, Krieg F, Hollebecq M-G, Mambrini M, Fauré A, et al. Enhanced individual selection for selecting fast growing fish: the “PROSPER” method, with application on brown trout (*Salmo trutta fario*). *Genet Sel Evol*. 2004;36(6):643–61. <https://doi.org/10.1051/gse:2004022>
- Crandell PA, Gall GAE. The genetics of body weight and its effect on early maturity based on individually tagged rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1993;117(1–2):77–93. [https://doi.org/10.1016/0044-8486\(93\)90125-I](https://doi.org/10.1016/0044-8486(93)90125-I)
- Crawley O, Barroso C, Testori S, Ferrandiz N, Silva N, Castellano-Pozo M, et al. Cohesin-interacting protein WAPL-1 regulates meiotic chromosome structure and cohesion by antagonizing specific cohesin complexes. *Elife*. 2016;5(FEBRUARY2016):1–26. <https://doi.org/10.7554/eLife.10851>
- D'Ambrosio J, Morvezen R, Brard-Fudulea S, Bestin A, Acin Perez A, Guéméné D, et al. Genetic architecture and genomic selection of female reproduction traits in rainbow trout. *BMC Genomics*. 2020;21(1):558. <https://doi.org/10.1186/s12864-020-06955-7>
- Gall GA. Genetics of reproduction in domesticated rainbow trout. *J Anim Sci*. 1975;40(1):19–28. <https://doi.org/10.2527/jas1975.40119x>
- Gall GAE, Gross SJ. A genetics analysis of the performance of three rainbow trout broodstocks. *Aquaculture*. 1978;15(2):113–27. [https://doi.org/10.1016/0044-8486\(78\)90057-1](https://doi.org/10.1016/0044-8486(78)90057-1)
- Gall GAE, Huang N. Heritability and selection schemes for rainbow trout: Female reproductive performance. *Aquaculture*. 1988;73(1–4):57–66. [https://doi.org/10.1016/0044-8486\(88\)90041-5](https://doi.org/10.1016/0044-8486(88)90041-5)
- Gall GAE, Neira R. Genetic analysis of female reproduction traits of farmed coho salmon

- (*Oncorhynchus kisutch*). *Aquaculture*. 2004;234(1–4):143–54. <https://doi.org/10.1016/j.aquaculture.2004.01.029>
- Gjedrem T, Haus E, Halseth V. Genetic variation in reproductive traits in Atlantic salmon and rainbow trout. *Aquaculture*. 1986;57(1–4):369. [https://doi.org/10.1016/0044-8486\(86\)90229-2](https://doi.org/10.1016/0044-8486(86)90229-2)
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). *Front Genet*. 2016;7(NOV). <https://doi.org/10.3389/fgene.2016.00203>
- Gutierrez AP, Lubieniecki KP, Fukui S, Withler RE, Swift B, Davidson WS. Detection of Quantitative Trait Loci (QTL) Related to Grilising and Late Sexual Maturation in Atlantic Salmon (*Salmo salar*). *Mar Biotechnol*. 2014;16(1):103–10. <https://doi.org/10.1007/s10126-013-9530-3>
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12. <https://doi.org/10.1186/1471-2105-12-186>
- Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol*. 2010;42(1):1–12. <https://doi.org/10.1186/1297-9686-42-5>
- Haffray P, Bugeon J, Rivard Q, Quittet B, Puyo S, Allamelou JM, et al. Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 2013;410–411:236–44. <https://doi.org/10.1016/j.aquaculture.2013.06.016>
- Haffray P, Vandeputte M, Petit V, Pincet C, Chatain B, Chapuis H, et al. Minimizing maternal effect in salmonid families mixed since eyed stages and a posteriori DNA-pedigreed. *Livest Sci*. 2012;150(1–3):170–8. <https://doi.org/10.1016/j.livsci.2012.08.017>
- Haidle L, Janssen JE, Gharbi K, Moghadam HK, Ferguson MM, Danzmann RG. Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*). *Mar Biotechnol*. 2008;10(5):579–92. <https://doi.org/10.1007/s10126-008-9098-5>
- Huang Y, Hickey JM, Cleveland MA, Maltecca C. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet Sel Evol*. 2012;44(1):1. <https://doi.org/10.1186/1297-9686-44-25>
- Jalabert B, Fostier A. *La truite arc-en-ciel De la biologie à l'élevage*. Editions Quae, editor. *La truite arc-en-ciel : de la biologie à l'élevage*. 2010.
- Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90(430):773–95. <https://doi.org/10.1080/01621459.1995.10476572>
- Küttner E, Moghadam HK, Skúlason S, Danzmann RG, Ferguson MM. Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus*). *Mol Genet Genomics*. 2011;286(1):67–79. <https://doi.org/10.1007/s00438-011-0628-x>
- Leder EH, Danzmann RG, Ferguson MM. The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *J Hered*. 2006;97(1):74–80. <https://doi.org/10.1093/jhered/esj004>
- Leeds TD, Vallejo RL, Weber GM, Gonzalez-Pena D, Silverstein JT. Response to five generations of selection for growth performance traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 2016;465:341–51. <https://doi.org/10.1016/j.aquaculture.2016.08.036>
- Li H. A quick method to calculate QTL confidence interval. *J Genet*. 2011;90. <https://doi.org/10.1007/s12041-011-0077-7>

- Li J, Wang Y, Zhou W, Li X, Chen H. The role of PKG in oocyte maturation of zebrafish. *Biochem Biophys Res Commun.* 2018;505(2):530–5. <https://doi.org/10.1016/j.bbrc.2018.09.124>
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Univ Georg Athens, USA. 2018;125.
- Misztal I, Tsuruta S, Strabel T, Druet T, Lee D. BLUPF90 and related programs (BGF90). *Proc 7th World Congr Genet Appl to Livest Prod.* 2002;2.
- Moghadam HK, Poissant J, Fotherby H, Haidle L, Ferguson MM, Danzmann RG. Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): Comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Mol Genet Genomics.* 2007;277(6):647–61. <https://doi.org/10.1007/s00438-007-0215-3>
- Moser G, Khatkar MS, Hayes BJ, Raadsma HW. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet Sel Evol.* 2010;42(1):1–15. <https://doi.org/10.1186/1297-9686-42-37>
- Neto RVR, Yoshida GM, Lhorente JP, Yáñez JM. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Mol Genet Genomics.* 2019;294(3):563–71. <https://doi.org/10.1007/s00438-018-1518-2>
- O'Malley KG, Sakamoto T, Danzmann RG, Ferguson MM. Quantitative trait loci for spawning date and body weight in rainbow trout: Testing for conserved effects across ancestrally duplicated chromosomes. *J Hered.* 2003;94(4):273–84. <https://doi.org/10.1093/jhered/esg067>
- Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwisse THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front Genet.* 2014;5(NOV):1–8. <https://doi.org/10.3389/fgene.2014.00402>
- Palaiokostas C, Vesely T, Kocour M, Prchal M, Pokorova D, Piackova V, et al. Optimizing genomic prediction of host resistance to Koi herpesvirus disease in carp. *Front Genet.* 2019;10(JUN):1–9. <https://doi.org/10.3389/fgene.2019.00543>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007;81(3):559–75. <https://doi.org/10.1086/519795>
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards : kmegrat g information about genes , prote and diseases. *Trends Genet.* 1997;13(163):4.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics.* 1998;14(8):656–64. <https://doi.org/10.1093/bioinformatics/14.8.656>
- Sakamoto T, Danzmann RG, Okamoto N, Ferguson MM, Ihssen PE. Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture.* 1999;173(1–4):33–43. [https://doi.org/10.1016/S0044-8486\(98\)00463-3](https://doi.org/10.1016/S0044-8486(98)00463-3)
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 2014;15(1). <https://doi.org/10.1186/1471-2164-15-478>
- Schurink A, Janss LL, Heuven HC. Bayesian Variable Selection to identify QTL affecting a simulated quantitative trait. *BMC Proc.* 2012;6(SUPPL. 2):S8. <https://doi.org/10.1186/1753-6561-6-S2-S8>
- Sprague J, Bayraktaroglu L, Bradford Y, Conlin T, Dunn N, Fashena D, et al. The Zebrafish Information Network: The zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.* 2008;36(SUPPL. 1):768–72.

- <https://doi.org/10.1093/nar/gkm956>
- Stanvitch G, Moore LL. cin-4, a gene with homology to topoisomerase II, is required for centromere resolution by cohesin removal from sister kinetochores during mitosis. *Genetics*. 2008;178(1):83–97. <https://doi.org/10.1534/genetics.107.075275>
- Su GS, Liljedahl LE, Gall GAE. Genetic and environmental variation of female reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 1997;154(2):115–24. [https://doi.org/10.1016/S0044-8486\(97\)00050-1](https://doi.org/10.1016/S0044-8486(97)00050-1)
- Su GS, Liljedahl LE, Gall GAE. Genetic correlations between body weight at different ages and with reproductive traits in rainbow trout. *Aquaculture*. 2002;213(1–4):85–94. [https://doi.org/10.1016/S0044-8486\(01\)00809-2](https://doi.org/10.1016/S0044-8486(01)00809-2)
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, Taggart JB, et al. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol*. 2016;48(1):1–11. <https://doi.org/10.1186/s12711-016-0226-9>
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015;16(1):1–9. <https://doi.org/10.1186/s12864-015-2117-9>
- Tsairidou S, Hamilton A, Robledo D, Bron JE, Houston RD. Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. *G3 Genes, Genomes, Genet*. 2020;10(2):581–90. <https://doi.org/10.1534/g3.119.400800>
- Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Sixin L, Parsons JE, et al. 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*. 2018;(February):1–12. <https://doi.org/10.1111/jbg.12335>
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)*. 2012;94(2):73–83. <https://doi.org/10.1017/S001667231200027>
- Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J. Genomic selection using low density marker panels with application to a sire line in pigs. *Genet Sel Evol*. 2013;45(1):1–11. <https://doi.org/10.1186/1297-9686-45-28>



## Chapitre 4: Architecture génétique et sélection des caractères de production

### 1. Introduction

Le principal caractère sélectionné est la croissance à différents âges, et notamment aux grandes tailles, car la production de truite a évolué rapidement de la truite portion vers de la grande truite pour produire des filets et vers de la très grande truite pour le fumage. Les rendements (rendement carcasse, rendement filet) et la teneur en lipides du muscle sont aussi des caractères très importants à sélectionner pour améliorer la rentabilité économique de la filière et la qualité des produits. La sélection généalogique a permis à la fois d'améliorer la précision de la sélection sur la croissance en intégrant données individuelles et données familiales, mais aussi permis d'inclure les caractères mesurés sur des collatéraux comme le rendement de découpe.

Dans ce contexte, le bénéfice principal escompté de la sélection génomique est d'améliorer l'efficacité de cette sélection sur apparentés en permettant de classer non plus seulement les familles des candidats à la sélection, mais aussi les individus à l'intérieur des familles.

De nombreuses études ont montré des estimations d'héritabilités limitées à fortes (entre 0,20 et 0,60) pour les caractères de production chez les salmonidés, tels que le rendement du filet, le poids du filet, le poids à la découpe ou le poids éviscéré et étêté (Kause et al. 2002, 2007; Neira et al. 2004; Powell et al. 2008; Haffray et al. 2012a, 2018; Tsai et al. 2015a; Leeds et al. 2016; Neto et al. 2019). Des études d'association pan-génomiques (GWAS) ont aussi montré que ces caractères de production sont très polygéniques (Gonzalez-Pena et al. 2016; Neto et al. 2019; Ali et al. 2020).

Le rendement en filet est un caractère majeur à considérer dans les programmes de sélection pour augmenter la rentabilité de la production, même s'il a une variabilité faible, que son phénotypage précis est complexe et sa sélection difficile car il ne peut être mesuré directement sur les candidats à la sélection (Vandeputte et al. 2019). Ces auteurs montrent qu'une sélection pour le rendement en filet est toutefois possible de façon optimale en combinant la résiduelle du poids éviscéré et étêté corrigé par le poids à la découpe (mesurée sur les collatéraux) et le rapport entre l'épaisseur de la paroi abdominale et la profondeur de la cavité péritonéale (mesurée sur les candidats à la sélection par ultrasons).

Le premier objectif de l'étude présentée dans ce chapitre est d'estimer dans une lignée du sélectionneur Aqualande les paramètres génétiques pour les caractères de découpe (poids total à la découpe, poids éviscéré et le poids éviscéré et étêté) et les caractères de qualité de la chair (couleur et surface du tissu adipeux de la darne), ces derniers caractères ayant encore peu été étudiés. Le second objectif de notre étude est de comparer l'efficacité de la sélection génomique par rapport à la sélection sur pedigree pour ces caractères et d'étudier les leviers d'action pour mettre en œuvre une SG efficace non seulement techniquement mais économiquement au travers du choix des outils de génotypage (densité de marqueurs) et des populations de référence (taille et structure).

## 2. Matériels et méthodes

### 2.1. Caractéristiques de la population étudiée

Les cohortes étudiées sont la 9<sup>ème</sup> génération (génération 9) et la 10<sup>ème</sup> génération (génération 10) de sélection de la lignée A du groupe Aqualande. Dans les deux cas, les cohortes sont monosexes femelles. Cette lignée est principalement sélectionnée sur la croissance et le poids à 18 mois par la méthode PROSPER.

Pour la génération 9, la fécondation a eu lieu le 2 décembre 2014 et l'abattage et le chantier de découpe du 24 au 26 mai 2016 à l'âge de 18 mois. Les animaux de cette génération ont été abattus par un bain avec adjonction de gaz carbonique, puis saignés. Ils ont été mis en caisse et ont eu 24h de maturation à 4°C avant découpe.

La génération 10 a été produite le 5 décembre 2016 et l'abattage et le chantier de découpe ont eu lieu du 23 au 26 avril 2018 à l'âge de 17 mois. Les animaux de la génération 10 ont été euthanasiés par commotion cérébrale puis saignés d'après le guide des bonnes pratiques animales défini par l' Union Européenne concernant les animaux de laboratoire (2010/63/UE). Ils ont été découpés rapidement après.

### 2.1.1. Phénotypes bruts collectés

La collecte des données brutes a concerné 3463 femelles : 1954 pour la génération 9 et 1509 pour la génération 10.

Les phénotypes recueillis sur chaque animal ont été le poids total à l'abattage, le poids des viscères et le poids de la tête.

Une darne a été découpée en avant de la nageoire dorsale selon le protocole défini dans Marty-Mahé et al. (2004) en position A5. Une image numérique de cette darne a été effectuée pour chaque animal. Les mesures prises ultérieurement sur l'image de la darne ont été faites de manières différentes entre les deux générations.

Pour la génération 9, les darnes ont été congelées pour être analysées *a posteriori* : 238 darnes ont été congelées en tunnel de surgélation et les 1709 restantes en chambre froide, ce qui est plus rapide à faire. Les darnes ont été décongelées soit entre le 30<sup>ème</sup> et 31<sup>ème</sup> jour, soit entre le 57<sup>ème</sup> et 66<sup>ème</sup> jour indépendamment de la méthode de congélation, et nettoyées avant la capture d'une image à l'aide d'un scanner A3 (Epson 10000 XL avec une résolution de 300 dpi) décrit plus précisément dans l'étude de Collewet et al. (2013).

Pour la génération 10, les darnes fraîches ont été photographiées durant le chantier à l'aide d'un appareil photo numérique sur pied (Canon EOS 1000 D 10M Pixels avec un dispositif spécial comprenant une tente de prise de vue (Literoom Photoflex©) pour éviter les reflets et avoir un environnement lumineux stable.

### 2.1.2. Phénotypes élaborés

Les statistiques de base des phénotypes des animaux génotypés étudiés sont résumées dans le [Tableau 16](#).

Le poids éviscéré (GCW) est calculé comme le poids de la femelle à la découpe moins le poids des viscères. Le poids éviscéré et étêté (HGCW) est calculé comme le poids de la femelle à la découpe moins le poids des viscères et de la tête.

Le rapport entre HGCW et le poids à l'abattage (BW) a été calculé pour contrôler les ratios incohérents liés à des erreurs de pesée lors du chantier de phénotypage. Ce contrôle était en particulier nécessaire pour les données enregistrées en génération 10 où une des balances utilisées lors du chantier de pesées était défectueuse. Les poids ou les ratios dont les valeurs

étaient supérieures en valeur absolue à la moyenne de la génération plus 4 écarts-types ont été éliminés du jeu de données analysé (1 donnée éliminée sur GCW et 40 sur le ratio HGCW/BW pour les deux générations). Pour la génération 9 où les pesées avaient été faites avec des balances toutes opérationnelles, le rapport HGCW/BW était compris entre 0,72 et 0,84 au maximum. Pour la génération 10, il a en conséquence été décidé d'éliminer les données correspondant à un rapport HGCW/BW inférieur à 0,70 ou supérieur à 0,90. Les 40 individus ne passant pas le filtre du rapport HGCW/BW ont été supprimés de l'analyse.

Les images numériques de darne ont été analysées en utilisant la méthode décrite par Marty-Mahé et al. (2004). Les images ont été traitées numériquement pour les transformer dans le système chromatique  $L^*a^*b^*$  (CIE 1976) (avec  $L^*$  : la clarté entre 0 et 100,  $a^*$  : valeur sur un axe vert-rouge et  $b^*$  : valeur sur un axe bleu-jaune) et une segmentation des images a été réalisée avec Visilog 7.3 for Windows © (Figure 24). La surface de darne et la surface adipeuse de la darne (ATA) comprenant les tissus adipeux, les muscles rouges et les myoseptes (voir graphique 1) ont été calculées à partir du traitement numérique réalisé sur l'image de la darne. Seulement deux phénotypes colorimétriques ont été retenus pour l'analyse génétique d'après la caractérisation des couleurs en système  $L^*a^*b^*$  : le degré de rouge ( $a^*$ ) de la chair (FCA) et la clarté ( $L^*$ ) de la chair (FCL).

#### Génération 9

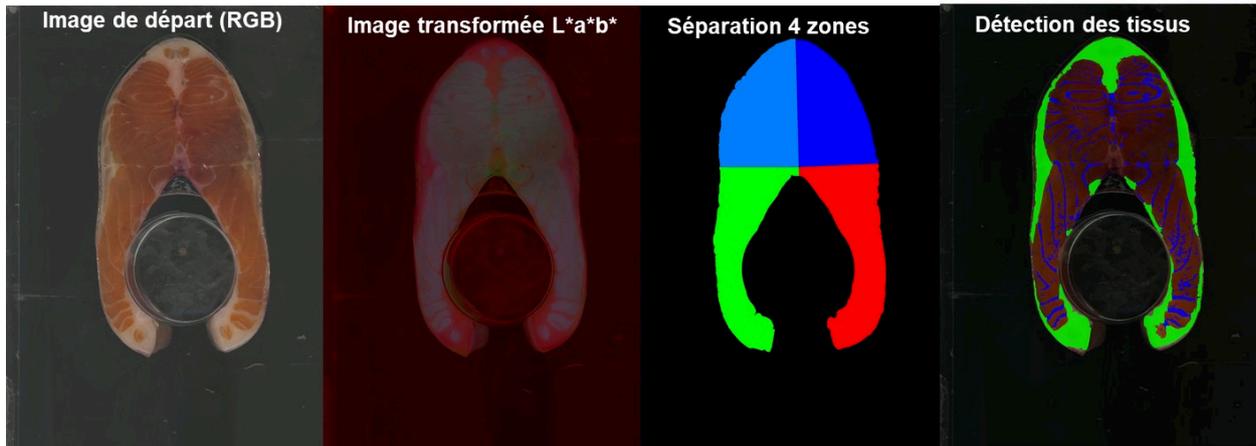
Caractères	N	Moyenne	Ecart-type	Min	Max	CV (%)
Poids à l'abattage (BW, g)	1 315	2 722	604	1 004	4 389	22,2
Poids à l'abattage éviscéré (GCW, g)	1 315	2 495	554	917	4 002	22,2
Poids à l'abattage éviscéré et étêté (HGCW, g)	1 315	2 105	480	753	3 418	22,8
Surface de tissus adipeux sur la darne (ATA, mm <sup>2</sup> )	1 315	1 414	399	175	2 832	28,2
Couleur de la chair A* (FCA)	1 315	27,05	2,18	18,50	32,60	8,1
Couleur de la chair L* (FCL)	1 315	44,59	2,32	35,10	52,80	5,2

#### Génération 10

Caractères	N	Moyenne	Ecart-type	Min	Max	CV (%)
Poids à l'abattage (BW, g)	1 424	2 143	469	658	3 363	21,9
Poids à l'abattage éviscéré (GCW, g)	1 424	1 959	425	621	3 073	21,7
Poids à l'abattage éviscéré et étêté (HGCW, g)	1 424	1 729	383	518	2 741	22,2
Surface de tissus adipeux sur la darne (ATA, mm <sup>2</sup> )	1 424	1 504	340	426	2 887	22,7
Couleur de la chair a* (FCA)	1 424	35,65	1,85	26,14	40,38	5,2
Couleur de la chair L* (FCL)	1 424	49,62	1,52	44,90	55,99	3,0

**Tableau 16 :** Statistiques descriptives des phénotypes élaborés mesurés dans les deux générations des animaux passant les contrôles de qualité (génotypage et phénotypage).

CV : coefficient de variation



**Figure 24 :** Exemple de traitement numérique à partir d'une image de la darne pour définir les caractères étudiés. © Jerome Bugeon.

Image 1 : Photo d'une darne. Image 2 : traitement de l'image pour convertir en couleur L\*a\*b\*. Image 3 : traitement numérique pour identifier 4 zones de la darne (bleu clair : dos gauche, bleu foncé : dos droit, vert : ventre gauche, rouge : ventre droit). Image 4 : identification des différents tissus de la darne (vert : tissus adipeux et muscle rouge (parties qui forment un "triangle" sur chaque côté de la darne), bleu : myosepte, marron : chair).

## 2.2. Génotypages

3168 individus phénotypés ont eu une quantité et qualité suffisantes d'ADN pour être génotypés avec la puce Axiom® Trout Genotyping array de 57 501 SNPs, ci-après nommée puce 57K (D'Ambrosio et al. 2019). Les génotypages de 3 individus ont été éliminés de l'analyse, car moins de 90% des SNPs avaient pu être génotypés. Après un contrôle de qualité, sont conservés les génotypages de :

- 1 316 femelles de la génération 9 avec leurs 177 parents également génotypés (79 mères et 98 pères) + 1 mère sans descendants génotypés.
- 1 492 femelles de la génération 10 avec leurs 182 parents également génotypés (99 pères et 83 mères).

Le nombre final d'individus génotypés et phénotypés par génération et la structuration familiale sont résumés dans le [Tableau 17](#).

Dans la population des 3168 individus phénotypés et génotypés, des contrôles qualité sont ensuite appliqués aux 38 350 SNPs exploitables de la puce 57K avec Plink 1.9 (Purcell et al. 2007) : call-rate > 97%, MAF > 5% et écart non extrême aux fréquences de l'équilibre d'Hardy-

Weinberg avec un seuil de p-value  $> 0,0001$ . Cela conduit à conserver seulement les génotypes de 29 652 SNPs pour l'analyse.

	Génération 9	Génération 10
<b>Pères</b>	100 (dont 2 pères non génotypés)	99
<b>Mères</b>	81 (dont 2 mères non génotypées)	84 (dont 1 mère non génotypée)
<b>Couples Père x Mère</b>	643	645
<b>Descendants par couple</b>	moyenne : 2±1 min : 1 max : 9	moyenne: 2±1 min : 1 max : 9
<b>Demi-sœurs</b>	moyenne : 27±5 min : 7 max : 38	moyenne: 29±7 min : 10 max : 54
<b>Descendants par père</b>	moyenne : 13±4 min : 3 max : 19	moyenne: 14±6 min : 3 max : 31
<b>Descendants par mère</b>	moyenne : 16±5 min : 1 max : 23	moyenne: 17±6 min : 4 max : 30

**Tableau 17 :** Structurations familiale des 2 739 individus (1 315 génération 9 et 1424 génération 10) passant les contrôles de qualité (génotypage et phénotypage).

### 2.2.1. Puce moyenne densité (MD)

La puce MD (ou puce 30K) avec 29 652 SNPs utiles est notre puce de référence pour évaluer l'efficacité de la sélection génomique. Pour tous les individus retenus pour l'analyse, l'information complète de leur génotype aux 29 652 SNPs est obtenue par imputation des génotypes manquants. L'imputation est réalisée à l'aide du logiciel FImpute (Sargolzaei et al. 2014) version 2.2. FImpute utilise une approche de fenêtre coulissante de SNPs le long du génome et l'information du pedigree et des haplotypes entre les individus aux génotypes à imputer et les individus aux génotypes connus pour imputer les informations manquantes (Sargolzaei et al. 2014).

Le test des incompatibilités mendéliennes sur les 29 652 marqueurs a permis de contrôler la qualité du pedigree fournie selon la méthodologie décrite dans le chapitre 3 (paragraphe 2.2.1). Pour la génération 9, 3 individus avaient l'un des parents mal assigné et 3 individus avaient leurs deux parents mal assignés. Les bons parents ont pu être retrouvés sauf pour 2 individus où l'un des parents a été déclaré inconnu dans le pedigree. Pour la génération 10, 20 individus

avaient l'un des parents mal assigné et 3 individus les deux parents mal assignés. Les bons parents n'ont pu être retrouvés que pour 5 individus. Les 18 individus restants ont eu le père déclaré inconnu car ce dernier ne semble pas avoir été génotypé. Le nombre d'individus avec des parents inconnus est mentionné dans le [Tableau 17](#).

### 2.2.2. Puce basse densité (LD)

Des génotypes LD ont été simulés à partir des génotypes de la puce MD. Différentes méthodes existent pour choisir les marqueurs: soit aléatoirement, soit à partir de critères de MAF, déséquilibre de liaison (DL), répartition équi-distante le long du génome ou concentrée au voisinage de QTLs (Moser et al. 2010; Ødegård et al. 2014; Tsai et al. 2015b; Vallejo et al. 2018).

Dans notre étude, on a choisi deux approches, toutes deux considérant un tri plus ou moins fort localement sur le DL car, dans le chapitre 2, nous avons montré que le DL se maintient à longue distance dans la population d'Aqualande ( $r^2 > 0,173$  à 1Mb).

La première approche a donc été un simple tri sur le DL pour éliminer les marqueurs très redondants localement le long du génome en considérant une fenêtre coulissante de longueur fixée (en kb). La seconde approche a été d'effectuer un choix préalable des SNPs basé sur une MAF minimum de 25% (au lieu du minimum initial de 5%) avant le tri sur le DL local, afin de ne retenir que les marqueurs les plus informatifs. Les génotypes LD ont été obtenus avec la commande (-- inder-parwise) de Plink 1.9 (Purcell et al. 2007) qui dans une fenêtre coulissante, d'une longueur  $f$  en kb, retire les SNPs avec un DL (mesure  $r^2$ ) dépassant le seuil donné dans la fenêtre coulissante.

Plusieurs types de puce LD ont été générés et/ou utilisés :

- Le choix des tailles  $f$  des fenêtres et des seuils de  $r^2$  est effectué de manière à générer des puces LD d'environ 1 000 (1K), 3 000 (3K), 6 000 (6K) ou 10 000 (10K) SNPs. Le choix des paramètres  $f$  et  $r^2$  et les statistiques de distance entre marqueurs successifs des puces considérées sont résumés dans le [Tableau 18](#) ci-dessous.

Puce	N. SNP	distance entre SNPs successifs (en kb)					para. Plink	
		1 <sup>er</sup> quartile	médiane	moyenne	3 <sup>ème</sup> quartile	max	f	r <sup>2</sup>
LD 1k-05-AQ	996	340	987	1 800	2 334	21 041	1 800	0,128
LD 1k-25-AQ	996	368	1 003	1 786	2 149	25 297	1 800	0,185
LD 3k-05-AQ	3 011	121	323	611	712	17 185	600	0,261
LD 3k-25-AQ	3 005	126	311	602	668	17 982	600	0,393
LD 6k-05-AQ	5 986	62	166	309	364	11 184	300	0,395
LD 6k-05-co	6 167	59	157	297	338	13 725	-	-
LD 6k-25-AQ	5 993	63	167	304	357	15 525	300	0,587
LD 10K-05-AQ	10 017	39	101	185	222	10 817	180	0,54
LD 10k-25-AQ	10 002	39	102	182	220	13 294	180	0,816
MD 30K	29 799	10	34	62	79	11 176	-	-

**Tableau 18 :** Nombre de SNPs (N. SNP), statistiques de distances (en kb) entre SNPs successifs et paramètres utilisés pour le choix des marqueurs (para. Plink) de chaque puce construite à partir de la population AQ.

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison ; f et r<sup>2</sup> sont les paramètres utilisés pour retenir le nombre souhaité de marqueurs à l'aide de Plink. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD mis en commun.

- Par ailleurs, des génotypes LD de la population Aqualande (AQ) ont aussi été simulés à partir de choix de marqueurs effectués antérieurement sur la population de Viviers de Sarrance (VD) pour constituer des puces LD d'environ 3 000 (3K), 6 000 (6K) et 10 000 (10K) SNPs. Les caractéristiques de ces puces sont décrites dans le chapitre 3 (Tableau 5). Le nombre de marqueurs polymorphes avec une MAF > 5% dans la population AQ pour ces puces établies à partir de la population VD est indiqué dans le Tableau 19. Les statistiques de distance moyenne entre SNPs successifs sont similaires entre les puces conçues à partir des génotypes AQ et VD. En revanche, la distance maximale entre deux SNPs successifs peut être notablement différente (plusieurs Mb de différence) entre ces puces.
- Les nombres de marqueurs des puces établis à partir de la population VD qui ont une MAF supérieure 5% dans la population AQ sont de :
  - 2 637 et 2 738 respectivement pour les puces 3K-05-VD et 3K-25-VD,
  - 5 104 et 5 396 respectivement pour les puces 6K-05-VD et 6K-25-VD,
  - 8 544 et 8 900 respectivement pour les puces 10K-05-VD et 10K-25-VD.
- Enfin, une puce 6K commune (6k-05-co) a été simulée à partir de puces 13K établies sur chacune des populations AQ et VD avec le paramètre f=135 kb et un DL maximum de 0,70 et 0,63, respectivement. Les 6167 marqueurs en commun entre ces deux puces 13K ont été choisis pour constituer la puce 6K-05-co.

Les génotypes sur puces LD ont été sélectionnés à partir des génotypages MD avant imputation des génotypes manquants. Les génotypes des SNPs manquants sur puces LD ont alors été imputés sans considérer une population de référence génotypée en MD (ceci afin de simuler une situation où tous les génotypages disponibles ne seraient que des génotypes LD).

### 2.2.3. Répartition des marqueurs le long du génome

La répartition des marqueurs le long du génome est différente selon le type de puce considéré. Pour faciliter la lecture de ce paragraphe, le terme « couverture du génome » va faire référence à l'ensemble des zones de 1 Mb ayant au moins 1 marqueur considéré. A même nombre total de marqueurs, les zones couvertes par les marqueurs peuvent être différentes d'une puce à une autre selon la méthode définie pour choisir les SNPs. Si on découpe le génome assemblé en zones successives de 1 Mb, nous avons 1 854 zones avec au moins 1 SNP utile, avec en moyenne  $21 \pm 10$  SNPs par zone en considérant les 38K SNPs utiles de la puce 57K. La puce 30K couvre 1 836 de ces zones, soit 99% de couverture par rapport aux SNPs disponibles, avec en moyenne  $16 \pm 8$  SNPs par zone. Parmi ces 1836 zones couvertes par la puce MD, environ 94, 86, 69 et 39% sont couvertes par les puces LD, respectivement de densité 10K, 6K, 3K et 1K (Tableau 19).

Lorsqu'une pré-sélection sur la MAF est réalisée, cela limite le nombre de marqueurs disponibles et augmente le  $r^2$  entre les marqueurs. En effet, pour les puces LD AQ avec une pré-sélection sur la MAF, le  $r^2$  moyen est deux fois supérieur par rapport aux puces LD AQ sans pré-sélection sur la MAF (à marqueurs constants). Cette augmentation du  $r^2$  est moins forte pour les puces LD VD.

Pour la puce 10K-25-AQ, 0,63% et 12% des marqueurs successifs ont un  $r^2$  supérieur à 0,80 et 0,60, respectivement, alors que le  $r^2$  max est de 0,54 pour 10K-05-AQ. Les marqueurs successifs avec un très fort  $r^2$  sont plus redondants ce qui limite leur intérêt, de même que les marqueurs avec une MAF trop faible sont peu discriminants.

Puce LD	N SNP	Quartile 0,25	Médiane	Moyenne	Quartile 0,75	% couverture du génome par rapport à la puce MD	N SNP par zone
1K-05-AQ	996	0,00	0,02	0,03	0,05	39,22	1,4±0,7
1K-25-AQ	996	0,01	0,03	0,05	0,08	39,65	1,4±0,7
3K-05-AQ	3 011	0,01	0,05	0,07	0,11	70,75	2,3±1,5
3K-25-AQ	3 005	0,02	0,08	0,12	0,20	68,68	2,4±1,5
3K-05-VD	2 637	0,02	0,08	0,15	0,20	68,63	2,1±1,3
3K-25-VD	2 738	0,02	0,08	0,16	0,23	69,88	2,1±1,3
6K-05-AQ	5 986	0,02	0,07	0,10	0,16	86,82	3,8±2,5
6K-25-AQ	5 993	0,03	0,14	0,19	0,32	86,82	3,8±2,4
6K-05-co	6 167	0,02	0,08	0,14	0,20	85,19	3,9±2,6
6K-05-VD	5 104	0,03	0,09	0,17	0,25	86,71	3,2±2,1
6K-25-VD	5 396	0,03	0,11	0,19	0,28	86,55	3,4±2,2
10K-05-AQ	10 017	0,02	0,08	0,13	0,20	94,77	5,8±3,7
10K-25-AQ	10 002	0,07	0,23	0,28	0,44	93,08	5,9±3,5
10K-05-VD	8 544	0,03	0,10	0,20	0,29	93,90	5,0±3,1
10K-25-VD	8 900	0,04	0,14	0,23	0,36	93,95	5,2±3,1

**Tableau 19 :** Déséquilibre de liaison ( $r^2$ ) entre les marqueurs successifs des différentes puces LD à partir des génotypes de la population AQ et pourcentage de couverture des puces par rapport à la puce MD et nombre moyen de SNP par zone.

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison. AQ et VD indiquent à partir de quelle population ont été effectués les choix des marqueurs. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD.

Dans la suite de cette partie, nous allons focaliser notre étude sur les différences de répartition des SNPs le long du génome entre les diverses puces 6K construites.

- **6K-05-AQ vs 6K-25-AQ :**

Pour la comparaison des deux puces 6K-AQ avec un tri minimal de 5 ou 25% , il est à noter que 7% des zones de 1 Mb couvertes par la puce MD ne sont couvertes par aucune de ces deux puces (Figure 25-A). Les deux puces couvrent à 80% les mêmes zones, mais seulement 60% des marqueurs sont communs aux deux puces. On note également une hétérogénéité des couvertures selon les chromosomes. Par exemple, la couverture est très similaire entre les deux puces sur le chromosome 20, alors que cette couverture est relativement différente sur le chromosome 2. Les zones les moins couvertes par les deux puces par rapport à la puce MD se situent généralement dans des régions au DL élevé comme au centre des chromosomes métacentriques (ex : chromosomes 4, 5, 6) ou sur des chromosomes avec un DL globalement élevé (ex : chromosomes 5 et 20).

- **6K-05-AQ vs 6K-05-VD :**

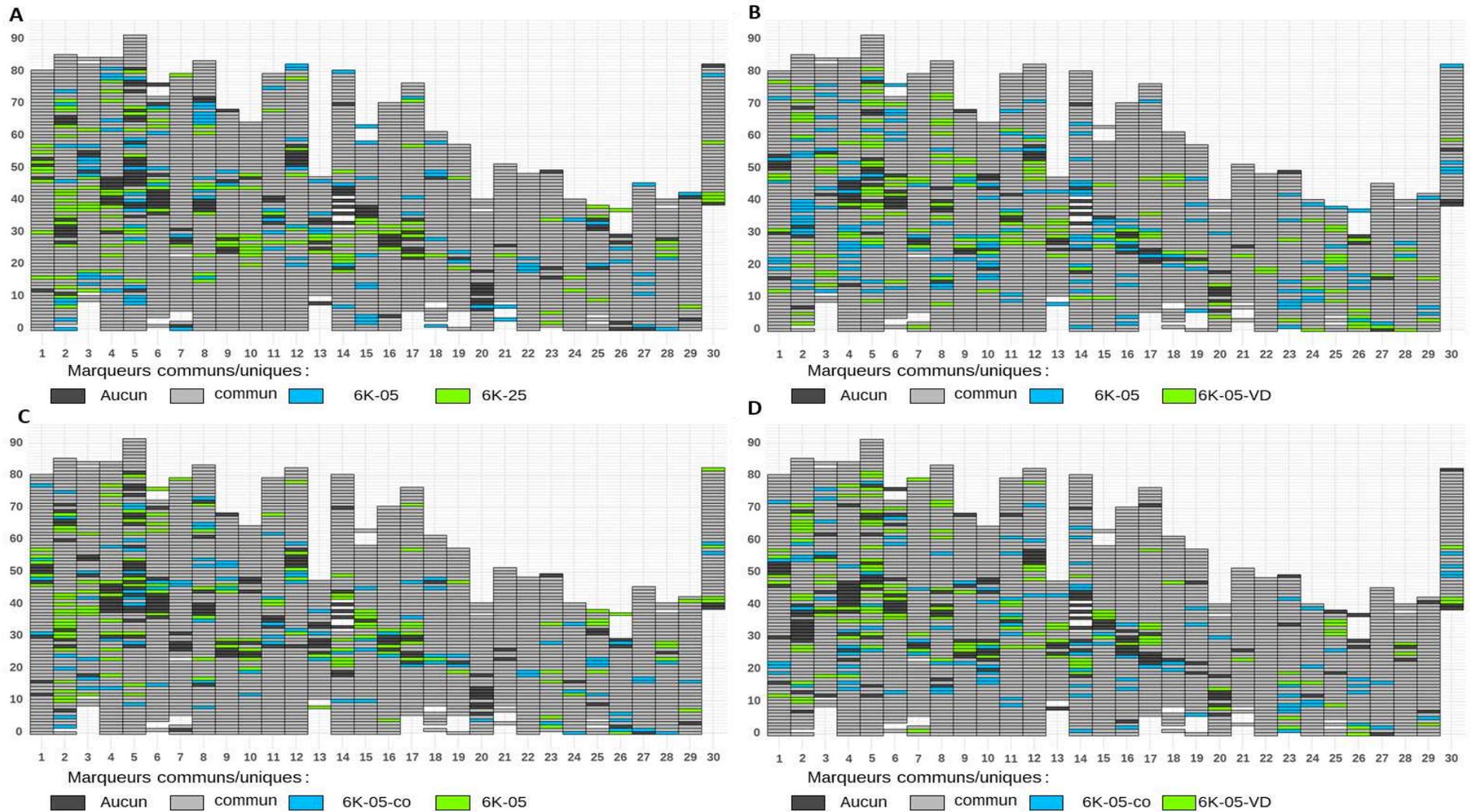
Dans le cas d'un choix sur une MAF supérieure à 5%, les deux puces 6K ne couvrent pas 5% des zones de 1Mb couvertes par la puce MD ([Figure 25-B](#)). 78% des zones couvertes sont identiques entre les deux puces, mais seulement 35% des marqueurs sont communs aux deux puces. On note des régions de plusieurs Mb couvertes très différemment par les deux puces. Par exemple entre 30 et 42 Mb sur le chromosome 2, 9 zones de 1 Mb ne sont pas couvertes par la puce 6K-05-VD alors que seulement 2 zones ne sont pas couvertes par la puce 6K-05-AQ. Sur les 92 segments de 1Mb constitutifs du chromosome 5, seulement 55 zones sont communément couvertes par les deux puces, 20 zones uniquement par la puce 6K-05-VD et 7 zones uniquement par la puce 6K-05-AQ et 9 zones ne sont couvertes par aucune des deux puces.

- **6K-05-co vs 6K-05-AQ et 6K-05-VD :**

La couverture commune entre la puce 6K-05-co et les puces 6K-05 est de 80% et 79% pour les puces établies sur les populations AQ et VD, respectivement. La proportion de marqueurs communs entre la puce 6K-05-co et ces deux puces est de 51% et 60% respectivement.

On observe dans la [Figure 25-C](#) qui compare la puce 6K-05-co et la puce 6K-05-AQ que les zones non couvertes par la puce 6k-05-co ont une taille de plusieurs Mb (couleur verte) alors que les zones non couvertes par la puce 6k-05-AQ (couleur bleue) ont en général une taille de seulement de 1 Mb. Nous retrouvons ce phénomène dans la comparaison entre la puce 6K-05-co et la puce 6K-05-VD ([Figure 25-D](#)). Il semblerait que la construction de la puce 6K-05-co à partir des marqueurs commun entre AQ et VD ait amplifié une couverture non-homogène et pourrait accumuler les défauts des puces 6k-05-AQ et 6k-05-VD.

A nombre de marqueurs quasi-constants, nous voyons donc que les diverses puces construites couvrent différemment le génome par rapport à la répartition le long du génome des 30K SNPs de la puce MD. Nous verrons dans la suite de ce chapitre les impacts potentiels sur la précision de l'évaluation génomique.



**Figure 25 :** Couverture du génome par les puces LD de densité 6K (absence ou présence d'au moins 1 marqueur par zone de 1Mb couverte par la puce MD).

A : Comparaison puce 6K-05-AQ et 6K-25-AQ, B : Comparaison puce 6K-05-AQ et 6K-05-VD, C : Comparaison puce 6K-05-co et 6K-05-AQ, D : Comparaison puce 6K-05-co et 6K-05-VD. Chaque barre verticale représente un chromosome (en abscisse) et chaque rectangle une zone de 1Mb (en ordonnée). La couleur représente l'absence ou présence de marqueurs dans chaque zone pour les deux puces comparées (gris foncé : absence dans les deux, gris clair : présence dans les deux puces, bleu et vert : présence pour une seule des deux puces (voir légende)).

## 2.3. Modèles et méthodes d'analyse génétique

### 2.3.1. Modèle animal et évaluation BLUP sur pedigree

Pour prédire les valeurs génétiques (EBV) et estimer les paramètres génétiques, un modèle animal, basé sur l'information du pedigree et la méthodologie BLUP, est utilisé :

$$y_i = X_i\beta_i + Z_iu_i + e_i \quad (1)$$

où  $y_i$  est le vecteur des performances pour le caractère  $i$ ,  $X_i$  la matrice d'incidence pour le vecteur des effets fixes  $\beta_i$ ,  $Z$  la matrice d'incidence pour le vecteur  $u_i$  des valeurs génétiques des animaux et  $e_i$  le vecteur des résidus du modèle pour le caractère  $i$ .

Le pedigree des 2 739 animaux avec performances est constitué de 17 235 individus enregistrés dans les généalogies des 10 générations précédentes établies par empreintes génétiques.

Pour tous les caractères (BW, GCW, HGCW, ATA, FCA et FCL), l'effet fixe de la génération a été inclus dans le modèle. Pour les caractères GCW et HGCW, le poids à la découpe BW a été considéré en covariable car l'objectif de sélection est d'améliorer le rendement en filet en augmentant le poids de carcasse à poids de découpe constant. Pour ATA, la surface totale de la darne a été considérée en covariable, car l'objectif de sélection est d'améliorer la teneur en gras à surface constante. Pour FCA et FCL, l'effet du type de congélation (tunnel de surgénélation, chambre froide, pas de congélation) et la durée de congélation (5 niveaux : non congelé, 29-35 J, 36-45 J, 46-50 J et 51-66 J de congélation) sont considérés comme effets fixes.

Les EBV et les composantes de variance sont estimées à l'aide du package BLUPf90 et AIREMLf90 (Misztal et al. 2002, 2018). Des analyses bi-caractères ont permis d'estimer intra-génération les corrélations génétiques entre les caractères considérés deux à deux. De plus, les corrélations génétiques entre les caractères de production (BW, GCW et HGCW) ont été estimées en cumulant l'information des deux générations car seuls ces trois caractères ont été phénotypés de la même manière sur les deux générations.

### 2.3.2. Modèle animal et évaluation GBLUP basée sur la parenté génomique

Le GBLUP ou BLUP génomique est très similaire au BLUP, la matrice de parenté calculée à partir du pedigree ( $A$ ) étant simplement remplacée par une matrice de parenté génomique ( $G$ ). Les valeurs génomiques prédites (GEBV) sont obtenues à l'aide du logiciel BLUPf90 (Misztal et al. 2018). Les mêmes effets fixes que pour les modèles BLUP décrits au paragraphe précédent ont été utilisés.

Les GEBV permettent de recalculer les effets estimés des SNPs grâce au package postGSF10 (Aguilar et al. 2014) à l'aide de la formule suivante (Wang et al. 2012) :

$$\hat{\mathbf{a}} = \mathbf{DZ}'[\mathbf{ZDZ}']^{-1}\hat{\mathbf{u}}_g \quad (2)$$

avec  $\mathbf{D}$  la matrice diagonale des poids associés aux effets de SNPs ( $D=I$  en GBLUP),  $\mathbf{Z}$  la matrice d'incidence qui relie les génotypes aux SNPs des individus génotypés au vecteur  $\hat{\mathbf{a}}$  des effets estimés des SNPs et  $\hat{\mathbf{u}}_g$  le vecteur des GEBV des individus génotypés ( $\hat{\mathbf{u}}_g = \mathbf{Z}\hat{\mathbf{a}}$ ).

Ensuite, on peut estimer les GEBV d'individus génotypés non inclus dans l'évaluation à partir des effets estimés des SNPs comme la somme des effets de leurs génotypes aux SNPs.

## 2.4. Comparaison de l'efficacité de la sélection selon différents scénarios

Pour estimer l'efficacité de la sélection sur pedigree (BLUP) ou génomique (GBLUP), nous avons effectué une étude de validation croisée. Pour cela, nous avons divisé notre jeu de données en deux populations : une population d'apprentissage (PA) et une population de validation (PV).

### 2.4.1. Les scénarios

Plusieurs scénarios ont été testés pour évaluer l'effet de la taille et la structuration de la population d'apprentissage sur l'efficacité de la sélection. Le récapitulatif des scénarios est présenté dans le [Tableau 20](#).

Deux scénarios intra- générations ont été considérés avec 40 échantillonnages de Monte-Carlo « leave-one-group-out » (tirage aléatoire avec remises) pour chaque générations :

- Le scénario **SW1** a consisté à tirer aléatoirement 960 individus dans la génération 9 pour constituer la PA, les 355 individus restants de la génération 9 constituant alors la PV.
- Le scénario **SW2** a consisté à tirer aléatoirement 960 individus dans la génération 10 pour constituer PA, les 464 individus restants de la génération constituant alors PV.

Cinq scénarios mélangeant en population d'apprentissage des individus des deux générations ont également été considérés. Dans ces cas-là, la génération 10 a été subdivisée en 4 populations de 356 individus pouvant tour à tour servir de PV : pv1, pv2, pv3 et pv4 (en respectant un équilibre dans les contributions de toutes les familles dans chacune des populations de validation). Les cinq scénarios définis sont les suivants :

- **SB1** : 960 individus choisis aléatoirement parmi les 1 315 de la génération 9 pour constituer PA. Pour une population de validation donnée, 10 échantillonnages de population d'apprentissage ont été réalisés. Au total, il y a donc eu 40 simulations ( $4 \times 10$ ) de ce scénario qui ont été effectuées.
- **SB2** : les 1 315 individus de la génération 9 ont constitué PA. Il n'y a donc eu que 4 simulations sous ce scénario, correspondant aux 4 populations de validation.
- **SB3** : les 1 315 individus de la génération 9 et une des sous-populations de 356 individus de la génération 10 ont constitué une PA de 1671 individus pour prédire à tour de rôle les 3 sous-populations de validation restantes. Au total, il y a donc 12 ( $=4 \times 3$ ) simulations réalisées sous ce scénario.
- **SB4** : les 1 315 individus de la génération 9 et 712 individus de la génération 10 (soit deux des 4 sous-populations considérées tour à tour :  $pv1+pv2$ ,  $pv1+pv3$ ,  $pv1+pv4$ ,  $pv2+pv3$ ,  $pv2+pv4$  ou  $pv3+pv4$ ) constituent une PA de 2 027 individus au total. Sous ce scénario, 12 simulations ( $6 \times 2$ ) sont également réalisées.
- **SB5** : les 1 315 individus de la génération 9 et 1 068 individus de la génération 10 (soit 3 des 4 sous-populations définies) constituent une PA de 2 383 individus au total. Seulement 4 simulations peuvent être effectuées sous ce scénario.

Nom	Pop. d'apprentissage			Pop. de validation	Nombre de simulations
	Génération 9	Génération 10	Total		
SB1	960	0	960	356	40
SB2	1 315	0	1 315	356	4
SB3	1 315	356	1 671	356	12
SB4	1 315	712	2 027	356	12
SB5	1 315	1 068	2 383	356	4
SW1	960	0	960	355	40
SW2	0	960	960	464	40

**Tableau 20 :** Récapitulatif du nombre d'individus considérés dans les populations d'apprentissage (issus des générations 9 et 10) et de validation (issus de la génération 10 sauf pour le scénario SW1) et du nombre de simulations réalisées pour chaque scénario  
Pop. : population

#### 2.4.2. Efficacité de la sélection

On utilise les génotypes et les phénotypes (ou seulement les phénotypes en BLUP) de PA pour établir les équations de prédiction. Ensuite, on estime les valeurs génétiques (EBV ou GEBV) des animaux de PV. La précision de la sélection est estimée à partir de la corrélation entre les (G)EBV estimées pour les individus de la population de validation PV et leurs phénotypes corrigés ( $y^*$ ) pour les effets fixes d'environnement, qui sont alors considérés comme un « proxy » des valeurs génétiques vraies en standardisant par la racine carrée de l'héritabilité  $h^2$  du caractère. La précision de la sélection ( $r$ ) est ainsi approchée par la formule :

$$r = \frac{\text{cor}(y^*, (G)EBV)}{\sqrt{h^2}} \quad (6)$$

La pente ( $b$ ) de la droite de régression des phénotypes  $y^*$  sur les (G)EBV permet de savoir si les (G)EBV surestiment ou sous-estiment la variabilité des valeurs génétiques vraies :

$$b = \text{cor}(y^*, (G)EBV) * \left( \frac{\sigma_{y^*}}{\sigma_{(G)EBV}} \right) \quad (7)$$

Ce coefficient  $b$  est aussi appelé coefficient de dispersion ou inflation. Pour qu'une évaluation soit sans biais,  $b$  doit être égal à 1. Si  $b < 1$ , la variabilité des (G)EBV est surestimée (surdispersion ou inflation) et si  $b > 1$ , la variabilité est sous-estimée (sous-dispersion ou déflation).

Pour chaque scénario étudié, la moyenne et l'écart-type d'échantillonnage des coefficients  $r$  et  $b$  ont été calculés sur l'ensemble des simulations réalisées (Tableau 20).

### 3. Résultats & discussion

#### 3.1. Paramètres génétiques

##### 3.1.1. Héritabilité des caractères

Pour les caractères de découpe, les héritabilités sont moyennes à fortes, variant de 0,25 à 0,51 (Tableau 21).

Génération 9 et 10						
Trait	BW	GCW*	HGCW*	ATA*	FCA	FCL
<b>BW</b>	<b>0,25 (0,04)</b>	0,01 (0,11)	-0,03 (0,11)			
<b>GCW*</b>	0,00 (0,03)	<b>0,54 (0,05)</b>	0,82 (0,03)			
<b>HGCW*</b>	0,00 (0,03)	0,74 (0,01)	<b>0,51 (0,05)</b>			

Génération 9						
Trait	BW	GCW*	HGCW*	ATA*	FCA	FCL
<b>BW</b>	<b>0,23 (0,05)</b>	0,02 (0,16)	0,00 (0,16)	-0,32 (0,14)	0,27 (0,17)	-0,25 (0,22)
<b>GCW*</b>	0,00 (0,04)	<b>0,52 (0,07)</b>	0,77 (0,06)	0,23 (0,12)	0,27 (0,14)	0,05 (0,19)
<b>HGCW*</b>	0,00 (0,03)	0,67 (0,02)	<b>0,48 (0,07)</b>	0,10 (0,13)	0,16 (0,15)	0,14 (0,20)
<b>ATA*</b>	-0,30 (0,04)	0,07 (0,03)	-0,01 (0,03)	<b>0,44 (0,07)</b>	0,62 (0,10)	-0,09 (0,19)
<b>FCA</b>	0,53 (0,02)	0,06 (0,03)	0,07 (0,03)	0,51 (0,02)	<b>0,26 (0,06)</b>	-0,14 (0,22)
<b>FCL</b>	-0,04 (0,03)	-0,01 (0,03)	0,03 (0,03)	0,10 (0,03)	-0,14 (0,03)	<b>0,15 (0,05)</b>

Génération 10						
Trait	BW	GCW*	HGCW*	ATA*	FCA	FCL
<b>BW</b>	<b>0,30 (0,06)</b>	-0,05 (0,14)	-0,10 (0,15)	-0,22 (0,17)	0,49 (0,12)	0,12 (0,14)
<b>GCW*</b>	-0,01 (0,04)	<b>0,51 (0,07)</b>	0,88 (0,03)	0,06 (0,15)	-0,16 (0,13)	0,00 (0,13)
<b>HGCW*</b>	-0,01 (0,04)	0,84 (0,01)	<b>0,47 (0,07)</b>	0,06 (0,15)	-0,15 (0,14)	-0,04 (0,13)
<b>ATA*</b>	0,02 (0,04)	0,02 (0,03)	0,03 (0,03)	<b>0,27 (0,06)</b>	0,11 (0,17)	-0,02 (0,15)
<b>FCA</b>	0,64 (0,02)	-0,04 (0,04)	-0,01 (0,04)	0,30 (0,03)	<b>0,35 (0,06)</b>	0,01 (0,14)
<b>FCL</b>	0,10 (0,03)	-0,01 (0,03)	0,00 (0,03)	-0,09 (0,03)	-0,08 (0,03)	<b>0,47 (0,07)</b>

**Tableau 21** : héritabilités des caractères sur la diagonale (en gras), corrélations génétiques au-dessus de la diagonale, corrélations phénotypiques en dessous de la diagonale.

Les erreurs standard sont indiquées entre parenthèses. BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids à la découpe éviscéré et étêté, ATA : surface adipeuse de la darne, FLA : couleur rouge A de la chair, FCL : clarté L de la chair.

\*Si GCW et HGCW ne sont pas corrigés pour BW, leurs corrélations génétiques et phénotypiques avec BW sont de 0,99 ( $\pm 0,01$ ) dans chacune des générations. \*Si ATA n'est pas corrigé pour la surface de la darne, ses corrélations génétique et phénotypique avec BW sont respectivement de 0,63 ( $\pm 0,10$ ) et 0,79 ( $\pm 0,012$ ) pour la génération 9 et de 0,87 ( $\pm 0,04$ ) et 0,86 ( $\pm 0,01$ ) pour la génération 10.

L'héritabilité de 0,25 obtenue dans notre étude pour le poids à 17-18 mois est dans l'intervalle de 0,20 à 0,52 des estimations fournies dans la littérature pour des poids entre 12 et 18 mois chez la truite arc-en-ciel et autres espèces de salmonidés toutes résumées dans le [Tableau 22](#).

Les héritabilités fortes estimées proches de 0,50 pour GCW et HGCW sont également cohérentes avec les estimations antérieures en truite arc-en-ciel : Haffray et al. (2012) estiment des héritabilités de 0,49 et 0,55 en pBLUP pour GCW et HGCW respectivement, dans la même population que dans notre étude à 17 mois tandis que Gonzales-Pena et al. (2016) estiment une héritabilité de 0,62 pour HGCW dans le cas d'une découpe à 13 mois.

Etudes	Espèce	h <sup>2</sup>	Mois / poids lors de l'étude
Gonzalez-Pena et al. 2016	Truite arc-en-ciel	0,32	13 mois
Kause et al. 2002	Truite arc-en-ciel	0,20	30 mois
Kause et al. 2007	Truite arc-en-ciel	0,29	1 à 2 kg (6 générations étudiées)
Haffray et al. 2012	Truite arc-en-ciel	0,37	17 mois
Leeds et al. 2016	Truite arc-en-ciel	0,27	13 mois
Neto et al. 2019	Truite arc-en-ciel	0,33	18 mois
Neira et al. 2004	Saumon coho	0,13 à 0,24	12 mois (selon la lignée)
Powell et al. 2008	Saumon Atlantique	0,50	23 - 24 mois

**Tableau 22** : héritabilités (h<sup>2</sup>) estimées pour le poids chez les espèces de salmonidés. Haffray et al. 2012 : générations antérieures à notre étude.

Pour les caractères mesurés sur la darne, les héritabilités varient considérablement selon la génération considérée, en particulier pour FCL avec des valeurs de 0,15 et 0,47 respectivement pour les générations 9 et 2. Pour FCA et ATA, les héritabilités restent dans des valeurs moyennes autour de 0,30, variant entre 0,26 et 0,35. On note des estimations d'héritabilité plus faibles dans la génération 9 pour FCA et FCL pouvant être expliquées par la congélation de la darne. En effet, après un cycle de congélation et décongélation, la couleur est dégradée car les poissons sont souvent dépigmentés (Jérôme Bugeon, communication personnelle). Par contre pour ATA, l'estimation de l'héritabilité est plus forte dans la génération 9 que dans la génération 10. Il est donc important d'uniformiser le phénotypage des caractères mesurés sur la darne.

Dans une étude précédente concernant une génération antérieure de la population d'Aqualande, l'héritabilité du taux de lipides de la chair mesuré indirectement par Fatmeter a été estimée à 0,55 et 0,30 pour FCA (Haffray et al. 2018). Une étude en cours sur la génération 10 montre une corrélation entre le Fatmeter de 0,55 et 0,24 avec le gras sous-cutané et le gras du myosepte obtenue par vision numérique (Blay et al, 2020).

Kause et al. (2008) chez la truite arc-en-ciel estiment une héritabilité de 0,58 pour FCL qui est plus élevée que notre étude. Des études estiment une héritabilité chez le saumon Atlantique de 0,20 et 0,05 pour FCA et FCL chez le saumon Atlantique (Norris et Cunningham 2004), 0,28 pour le pourcentage de gras mesuré par Fatmeter et 0,16 pour FCA à 23 mois (Powell et al. 2008).

### 3.1.2. Corrélations génétiques et phénotypiques entre les caractères

On observe des corrélations génétiques moyennes, positives ou négatives, entre BW et les caractères mesurés sur la darne. Celles-ci varient légèrement entre les deux générations (Tableau 21).

Les corrélations génétiques entre GCW et HGCW ajustées pour BW avec BW ne sont pas significativement différentes de zéro comme montré précédemment sur la même population (Haffray et al. 2012). Si GCW et HGCW ne sont pas ajustés pour BW, leurs corrélations génétiques avec BW sont de 0,99, ainsi que cela a été relevé dans différentes espèces de salmonidés (Neira et al. 2004; Kause et al. 2007; Powell et al. 2008; Haffray et al. 2012a; Tsai et al. 2015a).

Les corrélations phénotypique et génétique estimées entre GCW et HGCW sont de 0,74 et 0,82 respectivement. Dans une étude précédente sur la même population, Haffray et al. (2012) estiment les corrélations génétiques entre le poids du filet et les caractères GCW et HGCW à des valeurs encore plus fortes (0,79 et 0,98, respectivement), soulignant l'intérêt d'une sélection sur HGCW pour améliorer le poids du filet. Ces résultats sont difficilement comparables à ceux trouvés dans la littérature pour d'autres populations car ils correspondent à un rendement de découpe calculé par le ratio du poids eviscéré et étêté sur le poids à la découpe alors que nous avons travaillé sur la résiduelle du poids éviscéré et étêté ajusté par le poids à la découpe.

GCW et HGCW ne sont que peu corrélés aux caractères de qualité de la chair qui sont mesurés sur la darne (Tableau 21).

Ces caractères de qualité de chair sont également peu corrélés entre eux tant phénotypiquement que génétiquement, à l'exception de ATA et FCA qui s'avèrent modérément corrélés (Tableau

21). Ces résultats ne vont pas dans le sens d'études antérieures montrant un lien entre la clarté de la darne (FCL) et sa teneur en gras (Bugeon et al, 2010; Marty-Mahé et al, 2004).

Dans toutes les espèces animales (Homme : Wood et al. 2014; Bovin : Bouwman et al. 2018 ; Porc : Qiao et al. 2015), les caractères de croissance sont des caractères très polygéniques dont l'expression est expliquée par des centaines, voire des milliers de gènes. Gonzales-Pena et al. (2016), Neto et al. (2019) et Ali et al. (2020) ont montré que les caractères de croissance étaient également très polygéniques chez la truite arc-en-ciel. Ainsi, Gonzales-Pena et al. (2016) décrivent les poids à 10 et 13 mois et le HGCW comme des caractères polygéniques avec seulement quelques QTLs expliquant au maximum 1 à 2% de la variance génétique. Neto et al. (2019) montrent que le poids à 18 mois est un caractère polygénique avec des QTLs expliquant 1 à 3% de la variance génétique. Aussi, nous n'avons pas conduit de GWAS sur les caractères de poids et rendement de découpe, les évaluations GBLUP réalisées nous confirmant la forte polygénéicité de ces caractères dans la population d'Aqualande. De plus, des analyses GWAS sur les caractères de qualité de chair font l'objet d'un travail mené en 2020 sur les données de la génération 10 par Carole Blay dans le cadre du projet Omega-Truite (Blay et al, 2020). Nous avons donc focalisé notre étude sur l'intérêt de la sélection génomique par rapport à une sélection sur pedigree pour les caractères de production.

## 3.2. Efficacités comparées de la sélection sur pedigree ou génomique intra-génération

### 3.2.1. Evaluation avec la puce MD

Les précisions moyennes et les coefficients de dispersion des évaluations génétique et génomique pour chaque caractère et chacun des deux scénarios intra-génération SW1 et SW2 sont présentés dans le [Tableau 23](#).

Pour SW1, en moyenne, la précision de l'évaluation BLUP varie de 0,28 (FCL) à 0,54 (ATA) et la précision du GBLUP varie de 0,32 (FCL) à 0,69 (ATA), soit un gain de précision de 11% (HGCW) à 29% (ATA) attendu entre les évaluations GBLUP et BLUP. Pour les coefficients de dispersion  $b$ , les valeurs moyennes sont relativement proches de 1 sauf pour FCL ([Annexe -](#)

tableau 4 et Annexe - tableau 5) où une surdispersion des (G)EBV ( $b < 1$ ) est observée. Aucune différence notable n'est à mentionner entre les  $b$  estimés en BLUP et GBLUP.

Pour SW2, en moyenne, la précision de l'évaluation BLUP varie de 0,42 (FCA) à 0,55 (FCL) et la précision du GBLUP varie de 0,46 (FCA) à 0,68 (GCW), soit un gain de précision de 11% (FCA) à 32% (ATA) entre GBLUP et BLUP. Pour les coefficients de dispersion  $b$ , les valeurs moyennes sont toutes relativement proches de 1 en BLUP (Annexe - tableau 4 et Annexe - tableau 5).

Les précisions obtenues pour les évaluations intra-génération BLUP et GBLUP des caractères de poids et rendement de découpe sont plus élevées dans la génération 10 que dans la génération 9 : de 6 à 8% pour BW et GCW, de 1% (BLUP) et 17% (en GBLUP) pour HGCW entre la génération 10 par rapport à la génération 9. Toutefois, les gains de précision obtenus en GBLUP versus BLUP pour les caractères de découpe sont très similaires pour les deux scénarios intra-génération (1-2% de différence) sauf pour HCGW où le gain de précision est de 11% pour SW1 et 29% pour SW2.

Trait	BLUP				GBLUP			
	SW1		SW2		SW1		SW2	
	r	b	r	b	r	b	r	b
BW	0,43 (0,07)	0,99 (0,19)	0,46 (0,07)	1,11 (0,22)	0,49 (0,08)	0,87 (0,17)	0,52 (0,05)	0,98 (0,14)
GCW	0,51 (0,04)	1,01 (0,13)	0,54 (0,04)	1,00 (0,09)	0,63 (0,04)	0,95 (0,10)	0,68 (0,05)	0,96 (0,07)
HGCW	0,50 (0,06)	0,98 (0,14)	0,51 (0,05)	0,98 (0,11)	0,56 (0,06)	0,86 (0,10)	0,66 (0,05)	0,96 (0,09)
ATA	0,52 (0,05)	1,04 (0,13)	0,48 (0,04)	0,82 (0,12)	0,67 (0,06)	1,04 (0,12)	0,63 (0,05)	0,82 (0,11)
FCA	0,41 (0,07)	0,88 (0,17)	0,42 (0,06)	1,02 (0,18)	0,47 (0,06)	0,77 (0,11)	0,46 (0,06)	0,87 (0,15)
FCL	0,42 (0,08)	0,70 (0,24)	0,55 (0,06)	1,20 (0,17)	0,48 (0,08)	0,61 (0,17)	0,62 (0,05)	1,11 (0,11)

**Tableau 23 :** Précision moyenne (r) et coefficient moyen de dispersion (b) des 40 simulations d'évaluations BLUP et GBLUP pour les scénarios SW1 et SW2.

*Les écarts-types d'échantillonnage sont entre parenthèses.* BW : Poids à la découpe, GCW : poids à la découpe éviscère, HGCW : poids à la découpe éviscère et étêté, ATA : surface adipeuse de la darne, FCA : couleur de la chair A, FCL : couleur de la chair L.

Pour les caractères de qualité de chair, en raison des différences de méthode de phénotypage entre les deux générations, les précisions ont été calculées en standardisant les corrélations par la racine carrée de l'héritabilité estimée intra-génération car les estimations de paramètres génétiques étaient très différentes entre les deux générations (cf. [paragraphe 3.1.1.](#)). Si les deux types de mesure pour FCA donnent des précisions similaires entre BLUP et GBLUP, les mesures sur darne fraîche (génération 10) permettent d'obtenir des évaluations (G)BLUP clairement plus précises pour FCL mais un peu moins précises pour ATA que les mesures sur darne congelée (génération 9). En ce qui concerne les coefficients de dispersion des (G)EBV, on note que les mesures sur darne congelée (SW1) engendrent plus de biais (écart de b à 1 en valeur absolue) dans les évaluations (G)BLUP des paramètres de couleur que les mesures sur darne fraîche (SW2), alors que le phénomène inverse est observé pour le caractère ATA. Ces résultats nous amènent à suggérer que si l'objectif de sélection principal pour la qualité de la chair est ATA, une mesure sur darne congelée paraît plus efficace (en termes de biais et précision des évaluations) qu'une mesure sur darne fraîche. Si en revanche c'est la clarté de la darne qui est avant tout recherchée, la mesure sur darne fraîche est à promouvoir. Pour l'évaluation du paramètre de couleur  $a^*$ , les deux types de mesure permettent des évaluations d'efficacité relativement similaire.

À notre connaissance, une seule autre étude en validation croisée (Gonzalez-Pena et al. 2016) chez la truite arc-en-ciel s'intéresse aux estimations de précision de sélection pour des caractères de production. Les auteurs ont estimé un gain de précision de 100% et plus pour la SG ( $r$  de 0,50 à 0,55) versus l'évaluation BLUP ( $r$  de 0,13 à 0,28) pour le poids de filet, le poids éviscéré et étêté ou le rendement en filet (rapport du poids de filet sur le poids à la découpe) avec environ 1 000 individus en population de référence. Cette étude montre des précisions de la SG relativement proches de celles obtenues sur la population d'Aqualande. Le gain de la SG est plus faible dans notre étude liée à une précision beaucoup plus élevée des évaluations sur pedigree par rapport à l'étude de Gonzalez-Pena et al. (2016). Ceci peut être lié au plan factoriel permettant une connexion génétique de tous les individus d'une génération permettant des précisions élevées en BLUP (Haffray et al. 2018).

D'autres études présentent les gains de précision associés à la SG chez le saumon Atlantique : 11% pour le poids à 25 mois ( $r \approx 0,73$  en GBLUP avec environ 1 000 individus en PA ; 50K SNPs, Yoshida et al. 2018), 20% de gain pour le poids ( $r \approx 0,70$  en GBLUP) et la longueur ( $r \approx 0,66$  en GBLUP) à 12 mois (environ 400 individus en PA, avec 112K SNPs, Tsai et al. 2015b).

Bangera et al. (2017b) montrent un gain de précision entre GBLUP et BLUP chez le saumon Atlantique de 6% pour la teneur en gras du filet ( $r \approx 0,22$  en GBLUP), 8% pour le rendement en filet ( $r \approx 0,35$  en GBLUP) et 19% pour le rendement du poids eviscéré et étêté sur le poids ( $r \approx 0,39$  en GBLUP) et 43% pour la pigmentation du filet mesuré par spectométrie de réflectance dans le proche infrarouge ( $r \approx 0,68$  en GBLUP).

### 3.2.2. Evaluation avec les puces LD

Pour faciliter la lecture, nous appelons « puce LD AQ » une puce dont les SNPs ont été choisis à partir du LD et des MAF observés dans la population d'Aqualande et « puce LD VD » une puce dont les SNPs ont été choisis à partir de la population de Viviers de Sarrance (cf [Chapitre 4](#)). Les [Figure 26](#) et [Figure 27](#) montrent le nombre nécessaire de marqueurs à considérer dans les puces LD pour limiter la perte de précision des évaluations GBLUP par rapport à une évaluation réalisée avec la puce MD dans la population d'Aqualande pour les caractères de poids et de rendement de découpe et pour les caractères mesurés sur la darne.

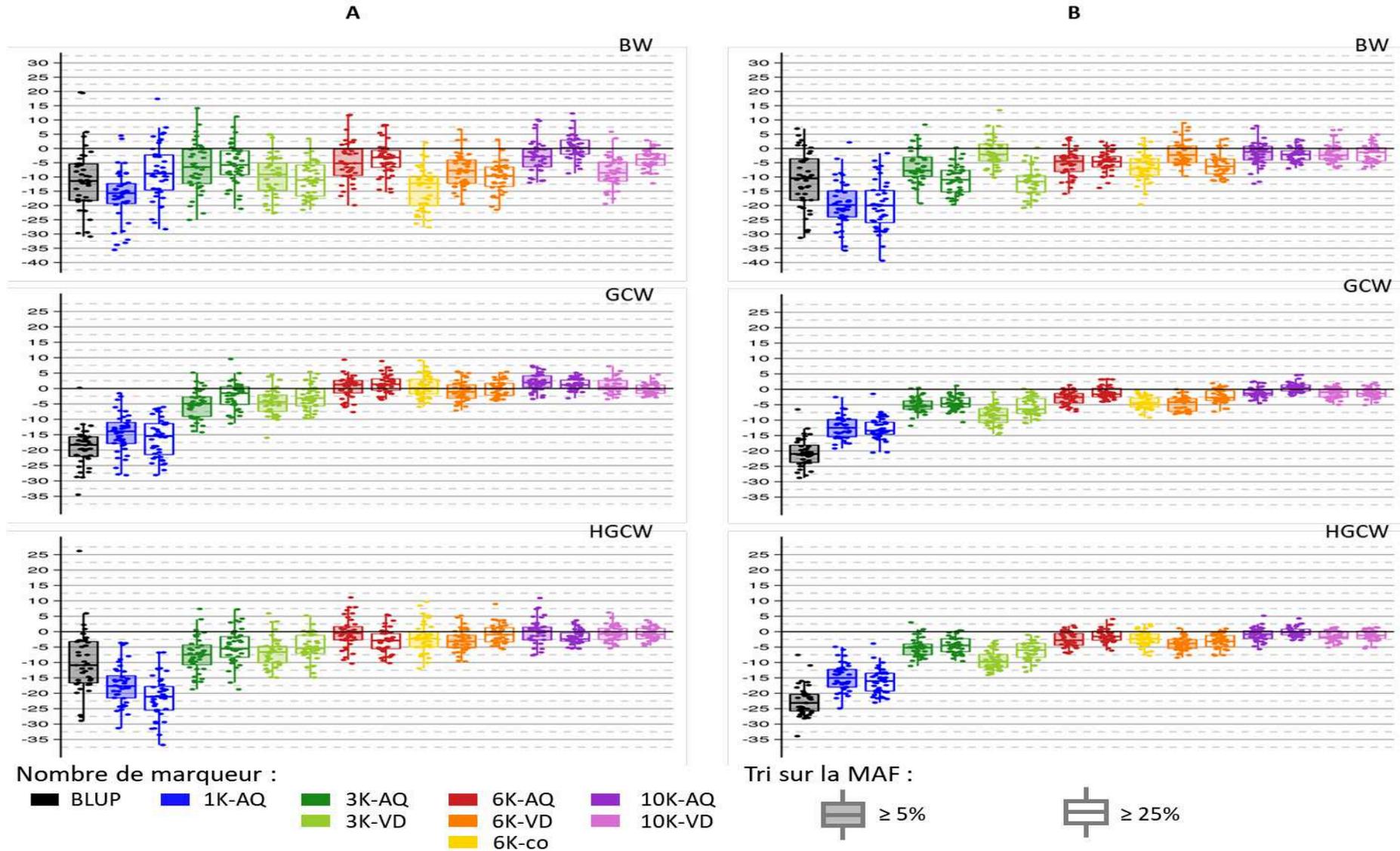
Les résultats sont similaires à ceux décrits dans le chapitre 3 pour les caractères de reproduction évalués dans la population de Viviers de Sarrance. Les précisions moyennes en fonction du type de puce LD considéré sont résumés dans [Annexe - tableau 4](#) pour les caractères de production et dans [Annexe - tableau 5](#) pour les caractères de qualité de chair.

Avec des génotypages LD 1K, on observe une précision plus faible ou au mieux équivalente à celle d'une évaluation BLUP sur pedigree. Avec des génotypages LD 3K, la précision de la SG devient souvent supérieure à celle d'une évaluation BLUP, en particulier quand les puces sont construites sur la population AQ. Globalement de meilleurs résultats sont obtenus quand il n'y a pas une présélection forte des marqueurs sur leur MAF, à l'exception du scénario SW1 avec des puces LD AQ où les précisions obtenues sont légèrement supérieures si les SNPs ont tous des  $MAF > 0,25$ .

En moyenne sur l'ensemble des caractères étudiés, les pertes de précision des évaluations LD par rapport aux évaluations MD sont inférieures à 5% pour tous les génotypages 6K et 10K considérés. Pour les puces 6K-AQ, la précision moyenne par rapport à la puce MD varie entre -5,7 à +1,6% selon le caractère et la génération alors qu'elles varient entre -10 à +1,8% pour

les puces 6K-VD. Pour les puces 10K-AQ, les précisions moyennes par rapport à la puce MD varient entre -3,4 à +2,4% selon le caractère et la génération alors qu'elles varient entre -8,3 à +1,7% pour les puces 10K-VD. Les puces 10K permettent une meilleure précision en moyenne que les puces 6K et une variabilité moins importante des résultats entre les différents caractères et génération.

Il n'y a pas d'argument clair pour privilégier le choix d'un outil intégrant une présélection forte sur la MAF dès lors que l'on construit des puces avec au moins 3 000 marqueurs. Une présélection d'intensité moindre, basée sur un seuil de MAF à 10 ou 15% par exemple, serait toutefois intéressante à considérer dans des études ultérieures.



**Figure 26 :** Boîtes à moustache des pertes de précision (en %) des évaluations BLUP (boîtes à moustaches noires) et GBLUP en LD (boîtes à moustaches colorées) par rapport aux évaluations GBLUP en MD des caractères BW, GCW et HGCW pour chaque scénario intra-génération : A - SW1 et B - SW2.

Les points représentent les résultats des 40 simulations de chaque scénario. BW : poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids à la découpe éviscéré et étêté.

Il est à noter que les génotypages basés sur une puce LD 6K (6K-05-co) construite à partir des SNPs communs aux puces 12K des populations AQ et VD donnent des résultats moins bons (au mieux équivalents) à ceux obtenus à partir des puces construites uniquement avec les informations de la population VD. La méthode choisie pour construire cette puce commune n'a donc pas permis d'améliorer la précision comparée à l'utilisation d'une puce basée uniquement sur le LD et les MAF d'une population autre que la population d'étude. Dans la partie 2.2.2, nous avons pu voir que la répartition des SNPs le long du génome n'est pas similaire entre les puces. Pour la puce 6K-05-co, des zones de 2-4 Mb du génome ne sont pas couvertes. Il semblerait que la puce 6K-05-co cumule les défauts de couverture des puces 6K-05-AQ et 6K-05-VD, la rendant moins performante qu'une puce choisie sur une seule des deux populations.

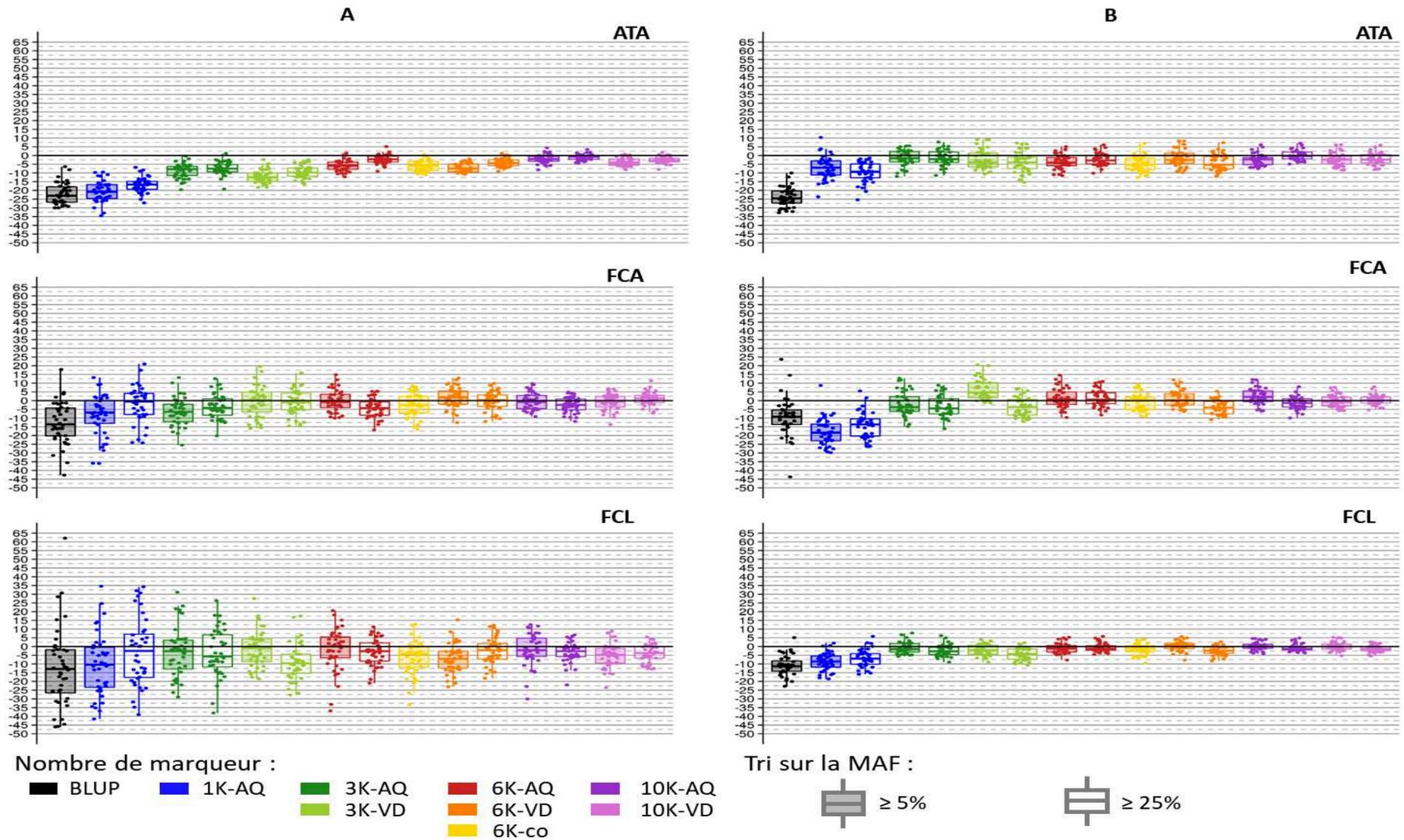
Nos résultats sont cohérents avec les observations déjà faites dans la littérature aquacole. Ainsi, chez la truite arc-en-ciel, Vallejo et al. (2018) montrent une augmentation négligeable de la précision au-delà de 10K (par rapport à la puce MD) pour la résistance à la flavobactériose. Chez le saumon Atlantique, Ødegård et al. (2014) décrivent une augmentation de la précision d'environ 7% entre 4K et 220K alors qu'elle est seulement d'environ 2% entre 20K et 220K pour la couleur du filet et une augmentation de la précision d'environ 3% entre 4K et 220K alors qu'elle est seulement d'environ 1% entre 20K et 220K pour la résistance au pou de mer. Tsai et al. (2015b) montrent un gain de précision de 20% pour la longueur et le poids du saumon à 1 an quand la densité en marqueurs passe de 0,5K à 5K, puis une très faible augmentation en allant jusqu'à 112K. Bangera et al. (2017) montrent une augmentation importante de la précision jusqu'à 3K puis une faible augmentation jusqu'à 10K pour la résistance à la *Piscorickettsia salmonis* par rapport à un génotypage 50K chez le saumon Atlantique.

D'autres études en aquaculture (Tsai et al. 2017; Tsairidou et al. 2019; Yoshida et al. 2018) montrent de très bonnes précisions de la SG avec beaucoup moins de marqueurs (0,25 à 1K SNPs) mais en imputant les génotypes LD en génotypes MD. Les candidats à la sélection et les collatéraux phénotypés sont génotypés avec une puce LD et ensuite imputés en MD grâce aux génotypes MD des parents. L'évaluation génomique est effectuée sur les données imputées en MD. Ensuite, les candidats retenus sont régénotypés avec la puce MD pour assurer la qualité de l'imputation de la génération suivante de candidats et collatéraux.

En résumé, nous concluons que si une puce LD est conçue uniquement pour une population donnée, 6 000 marqueurs semblent suffisants pour être presque aussi efficace qu'avec la puce MD, et, ce, quel que soit le caractère considéré. En revanche, si l'outil de génotypage doit être

multi-populationnel, il est souhaitable de privilégier un outil à 10 000 marqueurs pour avoir des pertes de précision négligeables (c'est-à-dire inférieures à 5%) quelle que soit la situation considérée (caractère, population). Toutefois, un outil à 6 000 marqueurs devrait toujours permettre un gain d'efficacité d'au moins 7% à 27% selon le caractère par rapport à une sélection sur pedigree et un gain de 10% à 28% selon le caractère pour une puce avec 10 000 marqueurs par rapport à une sélection sur pedigree.

Il serait également envisageable d'optimiser le choix des marqueurs pour leurs effets sur les caractères importants à améliorer (Moser et al. 2010; Vallejo et al. 2017), mais ceci limite l'intérêt de l'outil de génotypage ainsi construit à une seule population et un nombre restreint de caractères. En effet, Moser et al. (2010) montrent qu'un nombre limité de SNPs sont intéressants pour construire une puce visant à mieux caractériser deux caractères différents. En moyenne seront partagés seulement 25% des SNPs de deux puces 5K conçues spécifiquement pour améliorer chaque caractère.



**Figure 27 :** Boîtes à moustaches des pertes de précision (en %) des évaluations BLUP (boîtes à moustaches noires) et GBLUP en LD (boîtes à moustaches colorées) par rapport aux évaluations GBLUP en MD des caractères ATA, FCL et FCA pour chaque scénario intra-génération : A - SW1 et B - SW2.

Les points représentent les résultats des 40 simulations de chaque scénario. ATA : surface adipeuse de la darne, FCA : couleur de la chair A, FCL : couleur de la chair L.

### 3.3. Efficacité de la sélection entre générations

Dans cette partie, nous allons étudier l'impact sur la prédiction des valeurs génétiques des candidats à la sélection (appartenant la génération 10) de l'utilisation d'une population d'apprentissage construite :

- soit uniquement avec des individus de la génération 9 ;
- soit en combinant des individus de la génération 9 en complément d'individus de la génération 10.

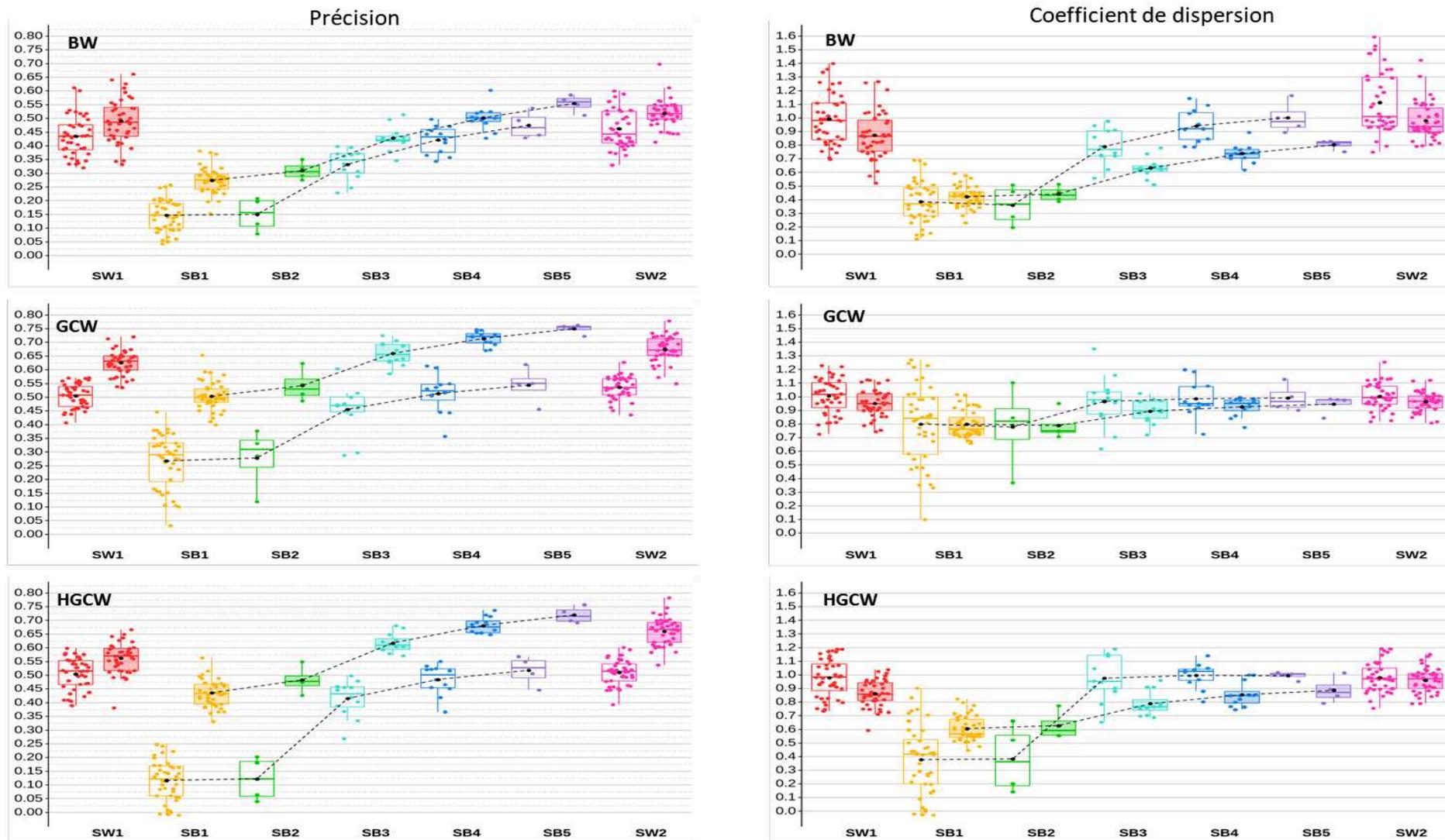
Les individus phénotypés et génotypés de la génération 9 sont uniquement des soeurs des parents de la génération 10, donc seules des tantes et non des parents de la génération 10 sont disponibles dans la génération 9 pour constituer la population d'apprentissage (12,5% de gènes en commun entre une tante et sa nièce (sans consanguinité)).

Pour évaluer l'efficacité de la sélection entre générations, nous envisageons 5 scénarios (SB1 à SB5, cf. partie 2.4). Pour SB1 et SB2, nous considérons une PA constituée respectivement de 960 et 1 315 individus de la génération 9 pour prédire 356 individus de la génération 10. Pour les scénarios SB3, SB4 et SB5, nous avons les 1 315 individus de la génération 9 entièrement en PA et nous complétons la PA avec respectivement 356, 712 et 1 068 individus de la génération 10. Ces 5 scénarios sont comparés au scénario intra-génération SW2.

#### 3.3.1. Evaluation avec la puce MD

##### **Précision et biais des valeurs génétiques prédites à partir des seules informations de la génération précédente**

Dans cette partie, nous comparons les scénarios SB1 et SB2 au scénario intra-génération SW2. Pour les 6 caractères étudiés, les pertes de précision associées aux scénarios SB1 et SB2 par rapport au scénario SW2 ([Tableau 24](#)) sont très conséquentes variant de -48 à -114% pour les évaluations BLUP et de -13 à -68% pour les évaluations GBLUP. Les pertes de précision sont plus conséquentes pour les paramètres de couleur que pour les autres caractères.



**Figure 28 :** Boîtes à moustache de la précision et du coefficient de dispersion d'une évaluation GBLUP avec 30K marqueurs (boîtes à moustaches pleines) et une évaluation BLUP (boîtes à moustaches transparentes) des différents échantillonnages des 7 scénarios testés pour les caractères de poids et découpe. La ligne pointillée représente la moyenne par scénario. BW : poids à la découpe, GCW : poids à la découpe éviscère, HGCW : poids à la découpe éviscère et étêté. g. : génération, scénario SW1 : 960 individus de la g. 9 en PA et individus de la g. 9 en PV, scénario SW2 : 960 individus de la g. 10 en PA et individus de la g. 10 en PV, Scénario SB1 : 960 individus de la g. 9 en PA, SB2 : 1 315 individus de la g. 9 en PA, SB3 : 1 315 individus de la g. 9 et 356 individus de la g. 10 en PA, SB4 : 1 315 individus de la g. 9 et 712 individus de la g. 10 en PA, SB5 : 1 315 individus de la g. 9 et 1 068 individus de la g. 10 en PA

Comme nous l'avons déjà relevé dans le chapitre 3, lorsque les populations d'apprentissage et de validation sont moins apparentées, l'efficacité de la sélection est réduite et les pertes de précision sont nettement plus importantes en BLUP qu'en GBLUP. Les marqueurs génomiques permettent alors de mieux capturer les ressemblances entre les individus des générations 9 et 10 que le pedigree. Les individus phénotypés et génotypés de la génération 9 inclus dans la PA sont uniquement les tantes (et non les parents) de la génération 10. Il ne serait donc pas souhaitable dans un schéma de sélection de phénotyper et génotyper une génération sur deux, la perte de précision étant trop importante quand la PA est seulement constituée de collatéraux des parents de la génération précédente.

En ce qui concerne la dispersion des (G)EBV, les estimations des coefficients  $b$  sont significativement inférieures à 1, variant entre 0,13 et 0,50 en BLUP et 0,41 et 0,63 en GBLUP (Tableau 24). Cela correspond à une forte inflation des (G)EBV qui indique qu'une évaluation des candidats à la sélection à partir des seules performances de la génération antérieure serait très fortement biaisée.

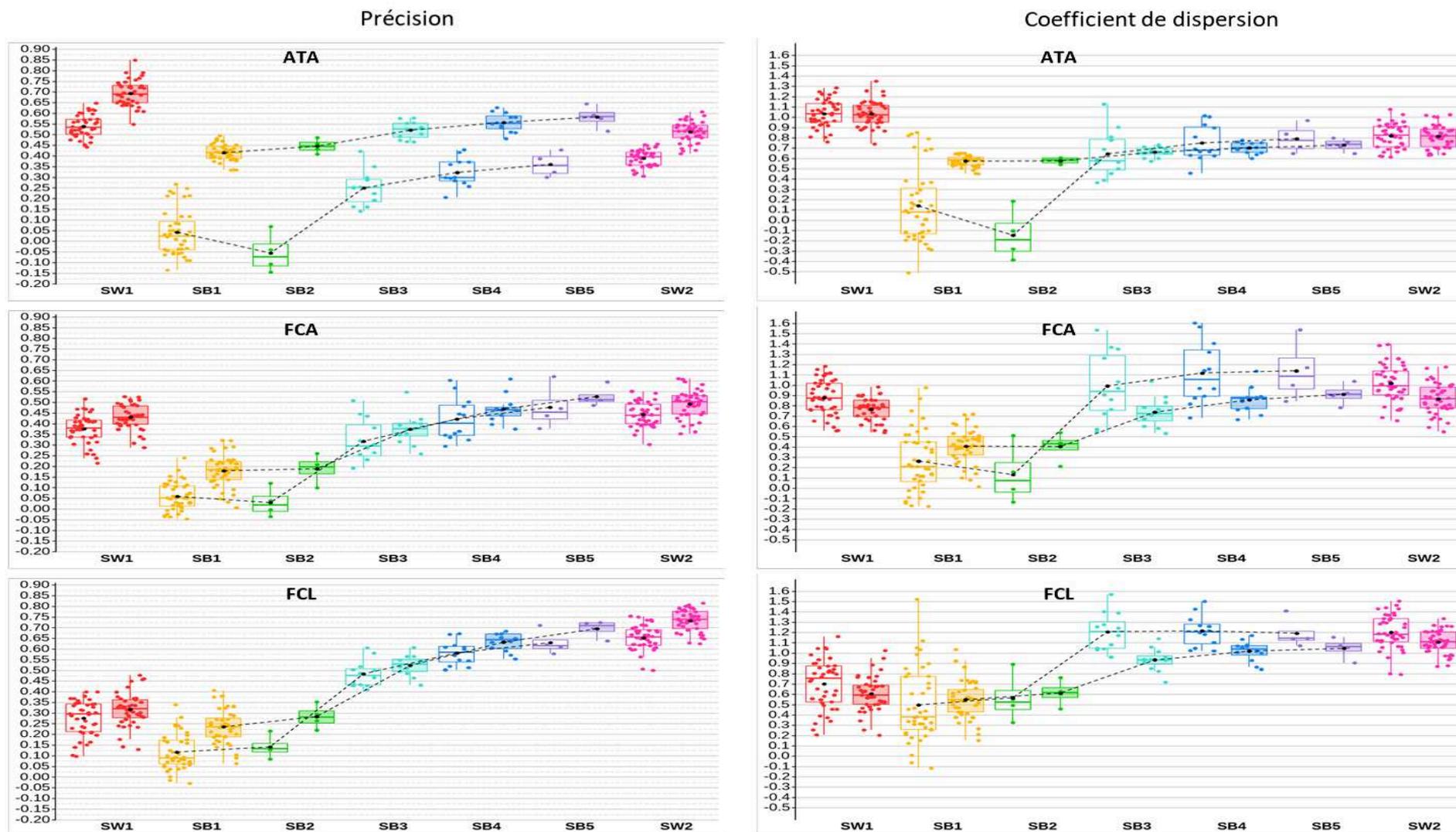
Wolc et al. (2011) montrent également des pertes de précision en sélection aviaire, mais plus faibles que dans notre étude, quand seules les données de générations antérieures sont utilisées pour l'évaluation génomique des candidats à la sélection. Les plus faibles pertes observées dans leurs travaux sont vraisemblablement expliquées par le fait que les parents des candidats à la sélection sont inclus dans leur population d'apprentissage contrairement à notre étude où seules des tantes sont considérées en PA.

Trait	SB1		SB2		SB3		SB4		SB5	
	BLUP	GBLUP	BLUP	GBLUP	BLUP	GBLUP	BLUP	GBLUP	BLUP	GBLUP
BW	-68,1	-47,0	-67,4	-40,3	-28,1	-17,5	-8,6	-3,2	<b>2,8</b>	<b>6,9</b>
GCW	-50,0	-25,3	-48,0	-19,6	-14,9	-2,2	-4,2	<b>5,8</b>	<b>1,5</b>	<b>11,1</b>
HGCW	-77,2	-33,9	-76,1	-26,8	-18,8	-6,4	-5,1	<b>3,3</b>	<b>1,4</b>	<b>9,1</b>
ATA	-89,0	-18,9	-114,0	-13,0	-35,8	1,5	-17,3	<b>8,6</b>	-7,6	<b>13,4</b>
FCA	-86,3	-63,4	-92,9	-61,5	-28,2	-23,9	-4,7	-4,9	<b>7,6</b>	<b>7,0</b>
FCL	-82,1	-67,9	-78,3	-61,3	-25,9	-28,6	-11,0	-13,7	-3,5	-5,1

**Tableau 24 :** Perte ou gain de précision (%) des scénarios SB1 à SB5 par rapport au scénario intra-génération SW2.

BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids éviscéré et étêté.

Scénario SB1 : 960 individus de la génération 9 en PA, SB2 : 1 315 individus de la génération 9 en PA, SB3 : 1 315 individus de la génération 9 et 356 individus de la génération 10 en PA, SB4 : 1 315 individus de la génération 9 et 712 individus de la génération 10 en PA, SB5 : 1 315 individus de la génération 9 et 1 068 individus de la génération 10 en PA.



**Figure 29 :** Boîtes à moustache de la précision et du coefficient de dispersion d'une évaluation GBLUP avec 30K marqueurs (boîtes à moustaches pleines) et une évaluation BLUP (boîtes à moustaches transparentes) des différents échantillonnages des 7 scénarios testés pour les caractères sur la darne.

La ligne pointillée représente la moyenne par scénario. ATA : surface adipeuse de la darne, FCA : couleur de la chair A, FCL : couleur de la chair L. g : génération, scénario SW1 : 960 individus de la g. 9 en PA et individus de la g. 9 en PV, scénario SW2 : 960 individus de la g. 10 en PA et individus de la g. 10 en PV, Scénario SB1 : 960 individus de la g. 9 en PA, SB2 : 1 315 individus de la g. 9 en PA, SB3 : 1 315 individus de la g. 9 et 356 individus de la g. 10 en PA, SB4 : 1 315 individus de la g. 9 et 712 individus de la g. 10 en PA, SB5 : 1 315 individus de la g. 9 et 1 068 individus de la g. 10 en PA

### **Intérêt du cumul d'information intra et entre générations en population d'apprentissage**

Dans cette partie, nous étudions grâce aux scénarios SB3, SB4 et SB5 l'intérêt pour l'évaluation de candidats (génération 10) d'un apport croissant en PA de données de la génération 10 en complément des données de la génération 9. Les précisions et les coefficients de dispersion des (G)EBV attendus pour tous les scénarios sont présentés dans les (Annexe - tableau 6).

Sur les Figure 28 et Figure 29, on peut voir que l'accumulation des données (scénarios SB3 à SB5) permet un gain de précision par rapport à une sélection intra-génération (SW2). Les gains ou pertes de précision des scénarios SB1 à SB5 par rapport au scénario SW2 sont synthétisés dans le Tableau 24. Nous observons un gain de précision de 1% à 3% en BLUP et de 7% à 11% en GBLUP pour les caractères de découpe entre les scénarios SB5 et SW2. L'accumulation des données d'une génération à une autre permet donc un gain de précision pour ces caractères. Le gain est plus important en GBLUP qu'en BLUP, mais, en parallèle, une légère augmentation de l'inflation des GEBV est observée. En revanche, on remarque une perte de précision de certaines évaluations (BLUP ou GBLUP) des caractères mesurés sur la darne entre les scénarios SB5 et SW2, pouvant s'expliquer par le fait que les phénotypes ont été mesurés différemment entre les deux générations. Le même phénomène avait été observé dans le chapitre 3 pour le caractère taille d'œuf en évaluation ssGBLUP. Il est donc important d'avoir une méthode de phénotypage constante et fiable pour pouvoir augmenter la précision par le cumul des données d'une génération à une autre. En cas de modification dans la méthode de phénotypage d'un caractère, il est donc conseillé d'étudier l'impact sur la précision de l'évaluation car il est préférable d'exclure les données anciennes de l'évaluation si des pertes de précision sont observées en cumulant les informations d'une génération à une autre.

Sur les Figure 28 et Figure 29, on voit une augmentation de la précision très nette pour les évaluations BLUP et GBLUP avec l'ajout à PA de collatéraux des candidats à la sélection (SB3, SB4 et SB5) par rapport aux scénarios SB1 et SB2. De même, les estimations des coefficients de dispersion sont beaucoup plus proches de 1 pour les scénarios SB3 à SB5 que pour les scénarios SB1 et SB2. L'évaluation GBLUP est toujours plus performante que l'évaluation BLUP. Le GBLUP avec une PA constituée des données de la seule génération des candidats à la sélection (SW2) est même plus performant qu'un BLUP cumulant l'information de 2 générations (SB5), les gains de précision sont de 9%, 24% et 27% respectivement pour BW, GCW et HGCW.

L'augmentation la plus importante de précision concerne les scénarios SB2 et SB3 : l'ajout de seulement 356 collatéraux de la génération 10 (SB3) à une PA uniquement constituée d'individus de la génération 9 (SB2) permet un gain de précision de 64% à 240% en BLUP et de 22% à 38% en GBLUP pour les caractères de découpe. Pour les caractères mesurés sur la darne, ce gain est beaucoup plus important notamment en BLUP où les précisions sont quasi-nulles pour le scénario SB2 (-0,07 à 0,14) alors qu'elles sont de 0,25 à 0,48 pour le scénario SB3. Ces gains de précision entre SB2 et SB3 supérieurs pour les caractères mesurés sur la darne s'expliquent par le fait que les données de la génération antérieure ne sont pas réellement informatives pour la génération suivante en raison des méthodes de phénotypage trop différentes entre les deux générations.

L'ajout en nombre croissant de collatéraux des candidats à la PA permet systématiquement de gagner en précision et de diminuer l'inflation des (G)EBV (Figure 28 et Figure 29).

Les gains diminuent cependant avec la taille initiale de la population d'apprentissage. En effet, les gains de précision entre les scénarios SB3 et SB4 varient selon les caractères de 13% à 33% en BLUP et de 7 à 25% en GBLUP alors que les gains les plus faibles observés entre les scénarios SB2 et SB3 sont de 63% et 17%, respectivement en BLUP et GBLUP. Les gains de précision entre les scénarios SB4 et SB5 varient, quant à eux, entre 6 à 13% en BLUP et 4 à 12% en GBLUP (Tableau 24).

Weng et al. (2016) montrent qu'un plateau de la précision associé à un accroissement du nombre de générations ancestrales considérées est plus rapidement atteint en évaluation sur pedigree qu'en sélection génomique. Pour les caractères à faible héritabilité la précision maximale est obtenue quand seulement les générations récentes sont ajoutées dans la population d'entraînement en générale la plus récente, alors que pour les caractères à forte héritabilité, la précision maximale est plutôt obtenue entre 5 et 8 générations ancestrales incluses dans la population d'apprentissage.

En conclusion, la sélection génomique est plus précise qu'une sélection sur pedigree à taille de population de référence constante. Le cumul des informations de la génération précédente et de la génération en cours d'évaluation rend encore plus précise une évaluation génomique. Pour réduire le coût de la mise en œuvre de la sélection génomique, il est alors envisageable de limiter le nombre d'individus génotypés et phénotypés à chaque génération. En effet, en cumulant l'information de générations successives, il est possible de réduire le nombre d'individus phénotypés de la génération en cours d'évaluation à précision au moins égale à la précision d'une évaluation sur la génération complète. Ainsi le scénario SB4 (PA avec 1 315

individus de la génération précédente et 756 individus de la génération en cours) permet une évaluation génomique de précision équivalente ou supérieure à celle d'une évaluation avec une PA constituée de 960 individus de la seule génération en cours. Sous l'hypothèse (très vraisemblable au vu de l'ensemble de nos résultats) que le perte de précision associée à la prise en compte de seulement 756 des 1 315 individus de la génération 9 soit négligeable, il est raisonnable de penser que phénotyper et génotyper environ 800 individus par génération permettrait une sélection plus efficace qu'une sélection sur pedigree avec 1 315 collatéraux. Si le coût du génotypage est inférieur au coût du phénotypage, la mise en œuvre de la SG serait alors économiquement rentable.

### 3.3.2. Evaluation avec la puce LD

Pour réduire les coûts de la mise en place de la sélection génomique, on peut envisager de réduire le nombre d'individus génotypés et phénotypés à chaque génération comme précédemment évoqué, mais aussi de réduire le nombre de marqueurs en utilisant pour le génotypage une puce LD moins chère que la puce MD.

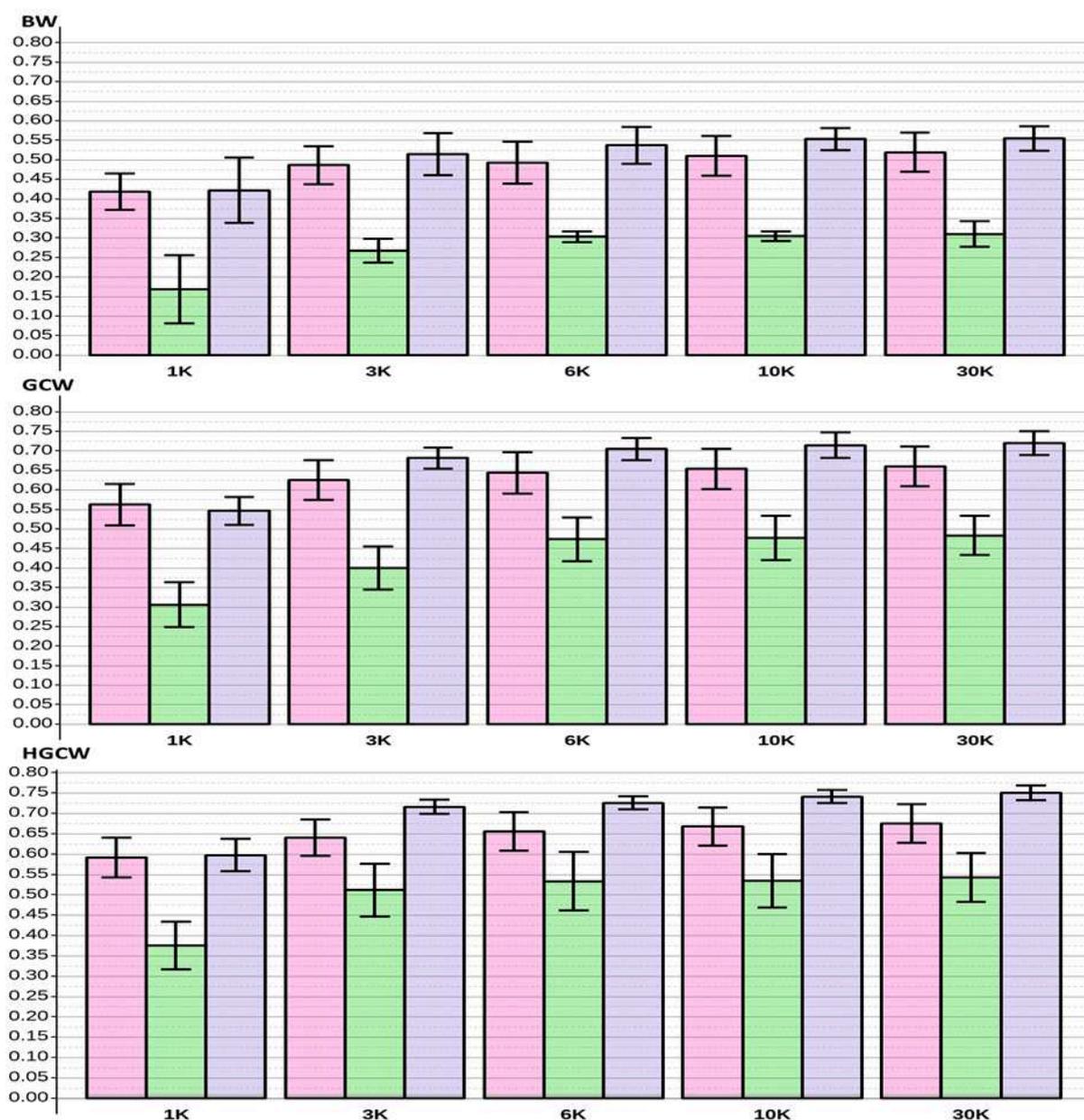
Aussi, nous avons comparé l'efficacité de la SG basée sur des génotypages LD dans des scénarios entre-génération versus intra-génération (SW2) afin d'évaluer si le fait de réduire le nombre de marqueurs pouvait ou non être compensé par un accroissement de la PA en cumulant les informations de deux génération successives.

Comme attendu en comparant le scénario SB1 au scénario SW2, on note que, quelle que soit la puce LD ou MD considérée, l'efficacité de la SG est nettement plus faible quand l'évaluation des candidats se fait à partir des données de la génération précédente (**Figure 30**). Contrairement à ce que nous avons décrit pour le scénario SW2 (cf **3.2.2.**) où nous observons un gain de précision de la SG avec une densité croissante en marqueurs jusqu'à 10 K, nous n'observons plus de gain de précision pour le scénario SB1 au-delà de 6K (**Figure 30**).

La **Figure 30** montre que quelle que soit la densité en marqueurs considérée à partir de 3K, le scénario SB5 permet une meilleure efficacité de la SG que le scénario SW2 à puce donnée. La perte de précision liée à la diminution des marqueurs est relativement similaire entre les

scénarios SW2 et SB5. On note, seulement une baisse de précision un peu plus importante de 6K à 3K dans le scénario SB5 par rapport au scénario SW2.

Ces résultats montrent que le cumul des données de la génération précédente avec la génération en cours d'évaluation ne permet pas de limiter la perte de précision associée à une faible densité en marqueurs.



**Figure 30:** Précision d'une évaluation génomique pour les caractères de production pour trois scénarios : SW2 (couleur rose), SB2 (couleur verte) et SB5 (couleur violette) en fonction du nombre de marqueur (1K à 30K).

Les précisions sont obtenues avec les puces AQ sans tri préalable sur la MAF. BW : Poids à la découpe, GCW : poids à la découpe éviscère, HGCW : poids à la découpe éviscère et étété. Scénario SW2 : 960 individus de la génération 10 en PA et individus de la génération 10 en PV, SB2 : 1 315 individus de la génération 9 en PA, SB5 : 1 315 individus de la génération 9 et 1 068 individus de la génération 10 en PA. Les précisions sont obtenues avec les puces AQ sans tri préalable sur la MAF.

### 3.4. Progrès génétiques attendus sur les caractères de découpe selon la stratégie d'évaluation:

Afin d'évaluer l'intérêt potentiel de la sélection génomique, nous avons calculé les progrès génétiques théoriques attendus pour une évaluation BLUP sur collatéraux ou GBLUP selon la densité de marqueurs considérée (Tableau 25). Ces calculs ont été effectués pour les scénarios de population d'apprentissage constituée intra-génération SW2 ou entre générations SB5. Le cas d'école considéré correspond à un taux de sélection de 10% et un objectif de sélection se réduisant à un seul caractère: soit le poids éviscéré et étêté corrigé du poids à la découpe (HGCW), soit le rendement de découpe (HGCW/BW).

Scénario	HGCW (g)				HGCW/BW (%)			
	SW2		SB5		SW2		SB5	
	$\Delta G$ (g)	$\Delta G/\bar{y}$ (%)	$\Delta G$ (g)	$\Delta G/\bar{y}$	$\Delta G$ (%)	$\Delta G/\bar{y}$ (%)	$\Delta G$ (%)	$\Delta G/\bar{y}$ (%)
<b>BLUP</b>	196	10,26	199	10,41	1,58	2,00	1,60	2,03
<b>1K-05-AQ</b>	216	11,30	216	11,30	1,72	2,19	1,72	2,19
<b>3K-05-AQ</b>	240	12,57	262	13,70	1,90	2,41	2,06	2,61
<b>6K-05-AQ</b>	247	12,93	270	14,16	1,95	2,47	2,12	2,69
<b>10K-05-AQ</b>	251	13,14	274	14,35	1,98	2,51	2,14	2,72
<b>Puce MD</b>	253	13,26	276	14,47	2,00	2,53	2,16	2,74

**Tableau 25 :** Progrès génétique théorique par génération pour un taux de sélection de 10% exprimé en unité physique ( $\Delta G$ ) et en pourcentage de la performance moyenne ( $\Delta G/\bar{y}$ ) pour le poids éviscéré et étêté (HGCW) et le rendement de découpe (HGCW/BW).

Que cela soit en sélection sur pedigree ou en sélection génomique, on peut avoir des gains de production importants de 10 à 13% de la moyenne pour HCWG, correspondant à des gains de 2 à 2,5% du rendement moyen de découpe dans le scénario SW2. La sélection génomique intra-génération permet un gain de plus de 13% de HGCW dès l'utilisation d'une puce 10K alors que le gain n'est que de 10% en sélection BLUP.

La prise en compte des informations de la génération précédente n'améliore guère l'efficacité de la sélection sur pedigree, mais permet un gain supplémentaire de plus de 1% de HGCW en sélection génomique dès une densité minimale 3K du génotypage. Une sélection sur puce 6K permettrait un gain quasi identique à une sélection avec la puce commerciale MD actuelle. Il faut toutefois noter que la puce MD actuelle ne couvre que très imparfaitement le génome (cf. Chapitre 3:) et que l'emploi d'un meilleur outil de génotypage MD permettrait vraisemblablement un surcroît d'efficacité potentielle de la sélection génomique. Si cette

hypothèse se confirmait, il serait alors vraisemblable qu'un outil LD avec plus de marqueurs (8 à 10K peut-être au lieu de 6K) permette de meilleurs gains de production.

## 4. Résumé

Les caractères de rendement (poids du filet) et de qualité de chair (teneur en lipides, couleur de la chair ...) sont importants pour la filière et ne peuvent être mesurés directement sur les candidats à la sélection puisqu'ils nécessitent l'abattage des individus pour réaliser les mesures. Les mesures sur des collatéraux des candidats à la sélection sont alors effectuées et l'utilisation d'un BLUP ou GBLUP – modèle animal permet alors l'évaluation des candidats pour ces caractères dont la mesure est létale.

Les objectifs de cette étude sont d'une part d'estimer les paramètres génétiques des caractères de rendement de découpe et de qualité de chair, d'autre part d'évaluer l'efficacité attendue d'une sélection génomique *versus* une sélection sur pedigree pour ces caractères, en fonction de la structuration de la population de référence et de la nature de la puce disponible pour réaliser le génotypage. Cette étude s'appuie sur les données de 2 739 individus issus de deux générations de la lignée A du sélectionneur Aqualande. Ces individus ont été génotypés avec la puce Axiom™ Trout Genotyping array (57 000 SNPs). Les phénotypes considérés dans l'analyse sont : le poids à la découpe (BW), le poids éviscéré (GCW), le poids éviscéré et étêté (HGCW), la surface de tissus adipeux de la darne (ATA), l'intensité de la couleur rouge de la chair a\* (FCA) et la clarté de la chair L\* (FCL).

Les caractères de découpe ont une héritabilité estimée moyenne à forte (BW : 0,25, GCW : 0,54 et HGCW : 0,51). Pour les caractères mesurés sur la darne, de manière différente entre les deux générations phénotypées, l'héritabilité du caractère ATA est de 0,44 et 0,27 pour les générations 9 et 10, respectivement ; les héritabilités de FCA et FCL sont respectivement de 0,26 et 0,15 en génération 9 et de 0,35 et 0,47 en génération 10.

La sélection génomique (SG) basée sur un génotypage MD permet un gain de précision de 11% à 35% intra-génération, par rapport à une sélection sur pedigree. Pour être efficace, la SG doit considérer une population de référence constituée *a minima* de plusieurs centaines d'individus phénotypés et génotypés qui soient des collatéraux des candidats à la sélection. En effet, une sélection basée sur la seule information phénotypique de la génération précédente n'est pas efficace que cela soit en évaluation BLUP ou GBLUP, même si la perte de précision par rapport

à une évaluation intra-génération est nettement plus faible en GBLUP (-12% à -67%), qu'en BLUP (-78% à -112%). Cumuler l'information de la génération précédente à celle de la génération en cours d'évaluation permet un gain de précision plus important en GBLUP (+7 à +11%) qu'en BLUP (+1 à +3%). Une SG basée sur un génotypage LD avec 6 000 à 10 000 marqueurs est presque aussi efficace qu'une SG basée sur un génotypage MD pour l'ensemble des caractères étudiés, les pertes de précision étant tout au plus de 5%. La mise en place d'une sélection génomique intra-génération sur HGCW permettrait avec 6 000 à 10 000 marqueurs une augmentation du HGCW d'environ 14% par génération avec une intensité de sélection de 10% alors qu'elle est d'environ 10% en BLUP.

## 5. Annexes

Puce LD	SW1						SW2					
	BW		GCW		HGCW		BW		GCW		HGCW	
	r	b	r	b	r	b	r	b	r	b	r	b
BLUP	0,43 (0,07)	0,99 (0,19)	0,51 (0,04)	1,01 (0,13)	0,50 (0,06)	0,98 (0,14)	0,46 (0,07)	1,11 (0,22)	0,54 (0,04)	1,00 (0,09)	0,51 (0,05)	0,98 (0,11)
1K-05-AQ	0,41 (0,08)	0,72 (0,17)	0,53 (0,05)	0,78 (0,09)	0,46 (0,06)	0,70 (0,10)	0,42 (0,05)	0,78 (0,11)	0,59 (0,05)	0,83 (0,07)	0,56 (0,05)	0,82 (0,08)
1K-25-AQ	0,45 (0,08)	0,79 (0,16)	0,52 (0,05)	0,76 (0,09)	0,44 (0,06)	0,66 (0,10)	0,41 (0,06)	0,78 (0,13)	0,59 (0,05)	0,82 (0,07)	0,55 (0,05)	0,79 (0,08)
3K-05-AQ	0,46 (0,08)	0,85 (0,18)	0,59 (0,05)	0,90 (0,09)	0,52 (0,06)	0,80 (0,10)	0,49 (0,05)	0,95 (0,14)	0,64 (0,04)	0,93 (0,07)	0,63 (0,05)	0,94 (0,10)
3K-25-AQ	0,46 (0,08)	0,84 (0,17)	0,61 (0,05)	0,91 (0,10)	0,53 (0,06)	0,81 (0,10)	0,46 (0,05)	0,90 (0,14)	0,65 (0,05)	0,93 (0,07)	0,63 (0,05)	0,92 (0,10)
3K-05-VD	0,44 (0,08)	0,81 (0,17)	0,60 (0,05)	0,90 (0,11)	0,52 (0,05)	0,79 (0,09)	0,51 (0,05)	0,95 (0,14)	0,62 (0,05)	0,88 (0,07)	0,60 (0,05)	0,88 (0,09)
3K-25-VD	0,45 (0,08)	0,84 (0,18)	0,61 (0,05)	0,90 (0,10)	0,53 (0,05)	0,81 (0,10)	0,46 (0,05)	0,87 (0,14)	0,64 (0,05)	0,90 (0,07)	0,62 (0,05)	0,89 (0,09)
6K-05-AQ	0,47 (0,08)	0,87 (0,18)	0,63 (0,05)	0,95 (0,10)	0,56 (0,06)	0,86 (0,10)	0,49 (0,05)	0,97 (0,15)	0,66 (0,05)	0,96 (0,08)	0,64 (0,05)	0,96 (0,10)
6K-25-AQ	0,47 (0,08)	0,84 (0,16)	0,64 (0,04)	0,95 (0,10)	0,54 (0,05)	0,83 (0,09)	0,49 (0,05)	0,95 (0,13)	0,67 (0,05)	0,95 (0,07)	0,65 (0,05)	0,94 (0,10)
6K-05-VD	0,44 (0,08)	0,81 (0,17)	0,62 (0,05)	0,94 (0,11)	0,54 (0,05)	0,84 (0,10)	0,51 (0,05)	0,97 (0,14)	0,64 (0,05)	0,92 (0,07)	0,63 (0,05)	0,93 (0,09)
6K-25-VD	0,45 (0,08)	0,83 (0,17)	0,63 (0,05)	0,94 (0,10)	0,56 (0,05)	0,85 (0,09)	0,49 (0,05)	0,92 (0,13)	0,66 (0,05)	0,94 (0,07)	0,64 (0,05)	0,93 (0,09)
6K-05-co	0,44 (0,08)	0,83 (0,19)	0,63 (0,05)	0,95 (0,10)	0,55 (0,05)	0,85 (0,10)	0,48 (0,05)	0,95 (0,13)	0,65 (0,04)	0,94 (0,06)	0,64 (0,05)	0,95 (0,09)
10K-05-AQ	0,48 (0,08)	0,87 (0,17)	0,64 (0,04)	0,96 (0,10)	0,56 (0,05)	0,87 (0,10)	0,51 (0,05)	0,99 (0,14)	0,67 (0,05)	0,97 (0,07)	0,65 (0,05)	0,97 (0,10)
10K-25-AQ	0,49 (0,08)	0,86 (0,16)	0,63 (0,04)	0,94 (0,10)	0,55 (0,05)	0,84 (0,10)	0,51 (0,05)	0,95 (0,13)	0,68 (0,05)	0,95 (0,07)	0,66 (0,05)	0,94 (0,09)
10K-05-VD	0,47 (0,08)	0,86 (0,17)	0,63 (0,04)	0,95 (0,10)	0,56 (0,05)	0,85 (0,10)	0,51 (0,05)	0,97 (0,14)	0,67 (0,05)	0,95 (0,07)	0,65 (0,05)	0,95 (0,09)
10K-25-VD	0,42 (0,08)	0,79 (0,17)	0,62 (0,05)	0,94 (0,10)	0,56 (0,05)	0,85 (0,09)	0,51 (0,05)	0,96 (0,13)	0,67 (0,05)	0,95 (0,07)	0,65 (0,05)	0,94 (0,09)
Puce MD	0,49 (0,08)	0,87 (0,17)	0,63 (0,04)	0,95 (0,10)	0,56 (0,06)	0,86 (0,10)	0,52 (0,05)	0,98 (0,14)	0,68 (0,05)	0,96 (0,07)	0,66 (0,05)	0,96 (0,09)

**Annexe - tableau 4 :** Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SW1 et SW2. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids éviscéré et étêté.

Scénario SW1 : 960 individus de la génération 9 en PA et individus de la génération 9 en PV, scénario

SW2 : 960 individus de la génération 10 en PA et individus de la génération 10 en PV.

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison. AQ et VD indiquent à partir de quelle population ont été effectués les choix des marqueurs. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD.

Puce LD	SW1						SW2					
	ATA		FCA		FCL		ATA		FCA		FCL	
	r	b	r	b	r	b	r	b	r	b	r	b
BLUP	0,54 (0,05)	1,04 (0,13)	0,38 (0,07)	0,88 (0,17)	0,28 (0,08)	0,70 (0,24)	0,39 (0,04)	0,82 (0,12)	0,44 (0,06)	1,02 (0,18)	0,65 (0,06)	1,20 (0,17)
1K-05-AQ	0,55 (0,07)	0,85 (0,13)	0,40 (0,07)	0,68 (0,13)	0,28 (0,08)	0,51 (0,16)	0,48 (0,04)	0,74 (0,08)	0,40 (0,06)	0,7 (0,13)	0,67 (0,05)	0,97 (0,12)
1K-25-AQ	0,58 (0,06)	0,88 (0,12)	0,42 (0,06)	0,72 (0,11)	0,30 (0,08)	0,54 (0,15)	0,46 (0,04)	0,72 (0,09)	0,42 (0,06)	0,73 (0,14)	0,68 (0,05)	0,98 (0,12)
3K-05-AQ	0,63 (0,06)	0,99 (0,13)	0,40 (0,06)	0,73 (0,12)	0,31 (0,08)	0,59 (0,17)	0,51 (0,04)	0,83 (0,10)	0,48 (0,06)	0,87 (0,15)	0,73 (0,05)	1,10 (0,12)
3K-25-AQ	0,64 (0,06)	0,99 (0,13)	0,42 (0,07)	0,74 (0,13)	0,31 (0,08)	0,59 (0,17)	0,51 (0,05)	0,81 (0,11)	0,48 (0,06)	0,85 (0,15)	0,72 (0,06)	1,08 (0,12)
3K-05-VD	0,61 (0,06)	0,95 (0,12)	0,43 (0,07)	0,76 (0,12)	0,31 (0,07)	0,59 (0,16)	0,050 (0,05)	0,80 (0,12)	0,52 (0,06)	0,91 (0,15)	0,72 (0,05)	1,06 (0,11)
3K-25-VD	0,63 (0,06)	0,97 (0,12)	0,43 (0,06)	0,74 (0,11)	0,29 (0,07)	0,54 (0,15)	0,49 (0,05)	0,78 (0,12)	0,48 (0,06)	0,82 (0,15)	0,70 (0,06)	1,04 (0,12)
6K-05-AQ	0,65 (0,06)	1,02 (0,13)	0,43 (0,07)	0,78 (0,12)	0,31 (0,08)	0,61 (0,18)	0,49 (0,04)	0,82 (0,11)	0,50 (0,06)	0,90 (0,16)	0,73 (0,06)	1,11 (0,12)
6K-25-AQ	0,68 (0,06)	1,03 (0,13)	0,41 (0,07)	0,73 (0,11)	0,31 (0,08)	0,59 (0,16)	0,50 (0,04)	0,81 (0,10)	0,50 (0,06)	0,89 (0,15)	0,73 (0,05)	1,09 (0,11)
6K-05-VD	0,65 (0,06)	1,00 (0,12)	0,44 (0,06)	0,77 (0,11)	0,30 (0,08)	0,57 (0,17)	0,51 (0,05)	0,82 (0,12)	0,50 (0,06)	0,88 (0,16)	0,74 (0,06)	1,09 (0,12)
6K-25-VD	0,67 (0,06)	1,02 (0,12)	0,43 (0,06)	0,75 (0,11)	0,31 (0,07)	0,58 (0,16)	0,49 (0,05)	0,79 (0,11)	0,47 (0,06)	0,84 (0,15)	0,72 (0,05)	1,07 (0,12)
6K-05-co	0,65 (0,06)	1,03 (0,12)	0,42 (0,06)	0,76 (0,11)	0,30 (0,08)	0,58 (0,17)	0,49 (0,05)	0,81 (0,11)	0,49 (0,06)	0,89 (0,16)	0,72 (0,06)	1,09 (0,12)
10K-05-AQ	0,68 (0,06)	1,05 (0,12)	0,43 (0,06)	0,77 (0,11)	0,31 (0,08)	0,61 (0,17)	0,50 (0,04)	0,83 (0,11)	0,50 (0,06)	0,91 (0,16)	0,73 (0,05)	1,12 (0,12)
10K-25-AQ	0,69 (0,06)	1,03 (0,13)	0,42 (0,06)	0,73 (0,11)	0,31 (0,08)	0,58 (0,16)	0,51 (0,04)	0,81 (0,10)	0,49 (0,07)	0,86 (0,15)	0,73 (0,05)	1,08 (0,11)
10K-05-VD	0,67 (0,06)	1,03 (0,12)	0,43 (0,06)	0,76 (0,11)	0,30 (0,08)	0,58 (0,17)	0,50 (0,05)	0,82 (0,12)	0,49 (0,06)	0,88 (0,16)	0,73 (0,06)	1,09 (0,12)
10K-25-VD	0,68 (0,06)	1,03 (0,12)	0,44 (0,06)	0,76 (0,10)	0,31 (0,08)	0,58 (0,16)	0,5 (0,05)	0,81 (0,11)	0,49 (0,06)	0,87 (0,15)	0,72 (0,05)	1,09 (0,12)
Puce MD	0,69 (0,06)	1,04 (0,12)	0,43 (0,06)	0,77 (0,11)	0,32 (0,08)	0,61 (0,17)	0,51 (0,05)	0,82 (0,11)	0,49 (0,06)	0,87 (0,15)	0,73 (0,05)	1,11 (0,11)

**Annexe - tableau 5 :** Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SW1 et SW2 et les différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

ATA : surface adipeuse de la darne, FLA : couleur de la chair A, FCL : couleur de la chair L Scénario SW1 : 960 individus de la génération 9 en PA et individus de la génération 9 en PV, scénario SW2 : 960 individus de la génération 10 en PA et individus de la génération 10 en PV.

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison. AQ et VD indiquent à partir de quelle population ont été effectués les choix des marqueurs. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD.

## BLUP

Trait	SB1		SB2		SB3		SB4		SB5	
	r	b	r	b	r	b	r	b	r	b
BW	0,15 (0,06)	0,39 (0,15)	0,15 (0,06)	0,36 (0,15)	0,33 (0,06)	0,79 (0,13)	0,42 (0,05)	0,94 (0,12)	0,47 (0,05)	1,00 (0,12)
GCW	0,27 (0,10)	0,80 (0,28)	0,28 (0,11)	0,78 (0,30)	0,46 (0,09)	0,97 (0,19)	0,51 (0,07)	0,99 (0,13)	0,54 (0,07)	0,99 (0,10)
HGCW	0,12 (0,08)	0,38 (0,25)	0,12 (0,08)	0,38 (0,25)	0,41 (0,07)	0,98 (0,16)	0,48 (0,05)	1,00 (0,09)	0,52 (0,05)	0,99 (0,03)
ATA	0,04 (0,11)	0,14 (0,35)	-0,05 (0,09)	-0,14 (0,25)	0,25 (0,08)	0,65 (0,23)	0,32 (0,07)	0,75 (0,18)	0,36 (0,06)	0,79 (0,14)
FCA	0,06 (0,07)	0,26 (0,29)	0,03 (0,07)	0,13 (0,28)	0,32 (0,10)	0,99 (0,33)	0,42 (0,1)	1,12 (0,30)	0,48 (0,10)	1,14 (0,30)
FCL	0,12 (0,08)	0,50 (0,36)	0,14 (0,05)	0,57 (0,24)	0,48 (0,06)	1,21 (0,18)	0,58 (0,06)	1,22 (0,15)	0,63 (0,06)	1,19 (0,15)

## GBLUP

Trait	SB1		SB2		SB3		SB4		SB5	
	r	b	r	b	r	b	r	b	r	b
BW	0,27 (0,05)	0,42 (0,08)	0,31 (0,03)	0,44 (0,06)	0,43 (0,04)	0,63 (0,07)	0,50 (0,04)	0,74 (0,07)	0,55 (0,03)	0,80 (0,03)
GCW	0,50 (0,05)	0,8 (0,10)	0,54 (0,06)	0,79 (0,11)	0,66 (0,04)	0,89 (0,09)	0,71 (0,03)	0,93 (0,07)	0,75 (0,02)	0,95 (0,07)
HGCW	0,44 (0,05)	0,61 (0,09)	0,48 (0,05)	0,63 (0,1)	0,62 (0,03)	0,79 (0,09)	0,68 (0,03)	0,86 (0,09)	0,72 (0,03)	0,89 (0,09)
ATA	0,42 (0,04)	0,58 (0,05)	0,45 (0,03)	0,58 (0,03)	0,52 (0,04)	0,66 (0,05)	0,56 (0,04)	0,70 (0,06)	0,58 (0,05)	0,73 (0,06)
FCA	0,18 (0,08)	0,41 (0,17)	0,19 (0,07)	0,41 (0,14)	0,38 (0,07)	0,74 (0,14)	0,47 (0,06)	0,86 (0,12)	0,53 (0,05)	0,91 (0,10)
FCL	0,24 (0,08)	0,55 (0,18)	0,28 (0,06)	0,62 (0,12)	0,52 (0,05)	0,94 (0,11)	0,63 (0,04)	1,02 (0,10)	0,70 (0,04)	1,05 (0,10)

**Annexe - tableau 6 :** Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB1 à SB5. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids éviscéré et étêté. Scénario SB1 : 960 individus de la génération 9 en PA, SB2 : 1 315 individus de la génération 9 en PA, SB3 : 1 315 individus de la génération 9 et 356 individus de la génération 10 en PA, SB4 : 1 315 individus de la génération 9 et 712 individus de la génération 10 en PA, SB5 : 1 315 individus de la génération 9 et 1 068 individus de la génération 10 en PA.

Puce LD	SB2						SB5					
	BW		GCW		HGCW		BW		GCW		HGCW	
	r	b	r	b	r	b	r	b	r	b	r	b
BLUP	0,15 (0,06)	0,36 (0,15)	0,28 (0,11)	0,78 (0,3)	0,12 (0,08)	0,38 (0,25)	0,47 (0,05)	1,00 (0,12)	0,54 (0,07)	0,99 (0,10)	0,52 (0,05)	0,99 (0,03)
1K-05-AQ	0,17 (0,09)	0,24 (0,12)	0,38 (0,06)	0,52 (0,10)	0,31 (0,06)	0,38 (0,09)	0,42 (0,08)	0,62 (0,12)	0,60 (0,04)	0,76 (0,09)	0,55 (0,04)	0,68 (0,07)
1K-25-AQ	0,21 (0,10)	0,29 (0,13)	0,39 (0,06)	0,53 (0,10)	0,33 (0,03)	0,40 (0,07)	0,41 (0,09)	0,58 (0,12)	0,59 (0,03)	0,73 (0,05)	0,56 (0,02)	0,68 (0,05)
3K-05-AQ	0,27 (0,03)	0,41 (0,03)	0,51 (0,07)	0,77 (0,13)	0,40 (0,06)	0,56 (0,11)	0,51 (0,05)	0,76 (0,07)	0,72 (0,02)	0,90 (0,07)	0,68 (0,03)	0,86 (0,05)
3K-25-AQ	0,29 (0,03)	0,43 (0,03)	0,50 (0,08)	0,72 (0,13)	0,45 (0,05)	0,59 (0,11)	0,49 (0,05)	0,72 (0,07)	0,71 (0,03)	0,88 (0,07)	0,69 (0,02)	0,84 (0,06)
3K-05-VD	0,29 (0,00)	0,44 (0,01)	0,48 (0,08)	0,73 (0,11)	0,45 (0,06)	0,59 (0,10)	0,55 (0,01)	0,80 (0,01)	0,69 (0,05)	0,87 (0,09)	0,65 (0,04)	0,79 (0,07)
3K-25-VD	0,26 (0,03)	0,40 (0,05)	0,47 (0,09)	0,68 (0,12)	0,43 (0,05)	0,56 (0,11)	0,48 (0,04)	0,71 (0,05)	0,70 (0,04)	0,89 (0,07)	0,67 (0,02)	0,81 (0,07)
6K-05-AQ	0,3 (0,01)	0,47 (0,02)	0,53 (0,07)	0,79 (0,13)	0,47 (0,06)	0,65 (0,12)	0,54 (0,05)	0,80 (0,07)	0,73 (0,02)	0,91 (0,07)	0,70 (0,03)	0,87 (0,07)
6K-25-AQ	0,30 (0,02)	0,42 (0,02)	0,54 (0,08)	0,76 (0,12)	0,49 (0,05)	0,65 (0,11)	0,52 (0,02)	0,76 (0,03)	0,73 (0,03)	0,90 (0,07)	0,71 (0,04)	0,87 (0,09)
6K-05-VD	0,28 (0,03)	0,43 (0,05)	0,51 (0,08)	0,75 (0,13)	0,47 (0,03)	0,62 (0,09)	0,54 (0,04)	0,80 (0,04)	0,71 (0,04)	0,90 (0,09)	0,68 (0,01)	0,84 (0,06)
6K-25-VD	0,29 (0,01)	0,43 (0,03)	0,51 (0,08)	0,72 (0,11)	0,46 (0,06)	0,58 (0,12)	0,51 (0,03)	0,76 (0,03)	0,73 (0,03)	0,91 (0,07)	0,70 (0,03)	0,85 (0,09)
6K-05-co	0,29 (0,03)	0,45 (0,04)	0,51 (0,08)	0,73 (0,13)	0,46 (0,07)	0,61 (0,12)	0,53 (0,05)	0,80 (0,07)	0,72 (0,03)	0,90 (0,08)	0,71 (0,04)	0,86 (0,09)
10K-05-AQ	0,30 (0,01)	0,46 (0,03)	0,53 (0,07)	0,79 (0,12)	0,48 (0,06)	0,65 (0,12)	0,55 (0,03)	0,82 (0,05)	0,74 (0,02)	0,92 (0,06)	0,71 (0,03)	0,88 (0,08)
10K-25-AQ	0,31 (0,02)	0,43 (0,04)	0,55 (0,06)	0,78 (0,10)	0,50 (0,05)	0,65 (0,11)	0,54 (0,03)	0,77 (0,04)	0,74 (0,02)	0,91 (0,07)	0,71 (0,04)	0,87 (0,09)
10K-05-VD	0,29 (0,01)	0,44 (0,03)	0,53 (0,07)	0,77 (0,11)	0,49 (0,05)	0,64 (0,11)	0,54 (0,02)	0,80 (0,02)	0,75 (0,03)	0,93 (0,08)	0,71 (0,02)	0,87 (0,08)
10K-25-VD	0,29 (0,02)	0,43 (0,04)	0,53 (0,07)	0,77 (0,11)	0,47 (0,06)	0,62 (0,11)	0,54 (0,03)	0,78 (0,04)	0,74 (0,02)	0,92 (0,06)	0,71 (0,03)	0,86 (0,09)
Puce MD	0,31 (0,03)	0,44 (0,06)	0,54 (0,06)	0,79 (0,11)	0,48 (0,05)	0,63 (0,10)	0,55 (0,03)	0,80 (0,03)	0,75 (0,02)	0,95 (0,07)	0,72 (0,03)	0,89 (0,09)

**Annexe - tableau 7 :** Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB2 et SB5. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids éviscéré et étêté.

SB2 : 1 315 individus de la génération 9 en PA, SB5 : 1 315 individus de la génération 9 et 1 068 individus de la génération 10 en PA. 05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison. AQ et VD indiquent à partir de quelle population ont été effectués les choix des marqueurs. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD.

Puce LD	SB2						SB5					
	ATA		FCA		FCL		ATA		FCA		FCL	
	r	b	r	b	r	b	r	b	r	b	r	b
BLUP	-0,05 (0,09)	-0,14 (0,25)	0,03 (0,07)	0,13 (0,28)	0,14 (0,05)	0,57 (0,24)	0,36 (0,06)	0,79 (0,14)	0,48 (0,10)	1,14 (0,3)	0,63 (0,06)	1,19 (0,15)
1K-05-AQ	0,32 (0,08)	0,42 (0,09)	0,08 (0,07)	0,14 (0,12)	0,19 (0,08)	0,37 (0,16)	0,52 (0,08)	0,66 (0,10)	0,41 (0,04)	0,67 (0,07)	0,61 (0,07)	0,88 (0,12)
1K-25-AQ	0,35 (0,06)	0,47 (0,08)	0,12 (0,07)	0,21 (0,12)	0,18 (0,10)	0,36 (0,2)	0,53 (0,05)	0,68 (0,06)	0,44 (0,05)	0,72 (0,10)	0,61 (0,09)	0,87 (0,17)
3K-05-AQ	0,42 (0,07)	0,63 (0,10)	0,11 (0,06)	0,23 (0,12)	0,23 (0,05)	0,52 (0,14)	0,58 (0,05)	0,76 (0,06)	0,51 (0,05)	0,87 (0,10)	0,67 (0,04)	1,00 (0,10)
3K-25-AQ	0,38 (0,06)	0,56 (0,06)	0,13 (0,08)	0,27 (0,15)	0,24 (0,11)	0,53 (0,25)	0,58 (0,06)	0,75 (0,06)	0,51 (0,07)	0,87 (0,15)	0,67 (0,08)	0,99 (0,17)
3K-05-VD	0,39 (0,07)	0,56 (0,10)	0,17 (0,08)	0,34 (0,15)	0,24 (0,08)	0,53 (0,20)	0,54 (0,05)	0,71 (0,07)	0,55 (0,08)	0,90 (0,15)	0,67 (0,03)	0,98 (0,10)
3K-25-VD	0,39 (0,07)	0,53 (0,09)	0,15 (0,06)	0,30 (0,11)	0,27 (0,08)	0,58 (0,17)	0,55 (0,05)	0,71 (0,07)	0,50 (0,07)	0,83 (0,15)	0,68 (0,04)	0,99 (0,11)
6K-05-AQ	0,40 (0,07)	0,61 (0,09)	0,15 (0,06)	0,33 (0,12)	0,26 (0,07)	0,60 (0,17)	0,57 (0,06)	0,77 (0,09)	0,54 (0,05)	0,93 (0,12)	0,68 (0,04)	1,02 (0,12)
6K-25-AQ	0,40 (0,06)	0,55 (0,07)	0,15 (0,07)	0,31 (0,14)	0,27 (0,08)	0,60 (0,18)	0,58 (0,05)	0,75 (0,06)	0,54 (0,07)	0,94 (0,16)	0,69 (0,05)	1,02 (0,13)
6K-05-VD	0,40 (0,05)	0,57 (0,06)	0,19 (0,05)	0,39 (0,10)	0,25 (0,08)	0,56 (0,19)	0,57 (0,04)	0,75 (0,07)	0,54 (0,06)	0,90 (0,13)	0,69 (0,03)	1,02 (0,10)
6K-25-VD	0,41 (0,05)	0,57 (0,06)	0,17 (0,06)	0,35 (0,11)	0,25 (0,08)	0,53 (0,17)	0,57 (0,03)	0,73 (0,04)	0,51 (0,07)	0,86 (0,14)	0,68 (0,03)	1,01 (0,08)
6K-05-co	0,37 (0,07)	0,55 (0,09)	0,16 (0,07)	0,33 (0,15)	0,28 (0,09)	0,63 (0,19)	0,56 (0,04)	0,76 (0,07)	0,54 (0,07)	0,93 (0,14)	0,69 (0,04)	1,02 (0,11)
10K-05-AQ	0,42 (0,06)	0,61 (0,06)	0,17 (0,07)	0,36 (0,13)	0,27 (0,07)	0,62 (0,17)	0,58 (0,05)	0,77 (0,07)	0,56 (0,06)	0,96 (0,14)	0,70 (0,04)	1,04 (0,11)
10K-25-AQ	0,44 (0,05)	0,57 (0,05)	0,18 (0,06)	0,37 (0,11)	0,26 (0,07)	0,55 (0,14)	0,59 (0,05)	0,74 (0,06)	0,53 (0,05)	0,91 (0,12)	0,69 (0,04)	1,03 (0,11)
10K-05-VD	0,42 (0,04)	0,59 (0,04)	0,18 (0,06)	0,38 (0,11)	0,26 (0,08)	0,57 (0,17)	0,57 (0,04)	0,75 (0,06)	0,53 (0,06)	0,90 (0,13)	0,69 (0,03)	1,03 (0,10)
10K-25-VD	0,42 (0,04)	0,56 (0,03)	0,16 (0,05)	0,34 (0,10)	0,26 (0,09)	0,56 (0,2)	0,58 (0,04)	0,74 (0,05)	0,52 (0,05)	0,88 (0,11)	0,69 (0,04)	1,03 (0,10)
Puce MD	0,45 (0,03)	0,58 (0,03)	0,19 (0,07)	0,41 (0,14)	0,28 (0,06)	0,62 (0,12)	0,58 (0,05)	0,73 (0,06)	0,53 (0,05)	0,91 (0,10)	0,70 (0,04)	1,05 (0,10)

**Annexe - tableau 8 :** Précision moyenne (r) et coefficient moyen de dispersion (b) pour les simulations aléatoires des évaluations BLUP et GBLUP pour les scénarios SB2 et SB5 et les différentes puces LD. Les écarts-types d'échantillonnage sont indiqués entre parenthèses.

ATA : surface adipeuse de la darne, FLA : couleur de la chair A, FCL : couleur de la chair L Scénario SB2 : 1 315 individus de la génération 9 en PA, SB5 : 1 315 individus de la génération 9 et 1 068 individus de la génération 10 en PA.

05 et 25 dans les noms des puces indiquent la MAF considérée avant le choix des SNPs selon le déséquilibre de liaison. AQ et VD indiquent à partir de quelle population ont été effectués les choix des marqueurs. 6K-05-co est la puce LD conçue à partir des génotypes des populations AQ et VD.

## 6. References

- Aguilar I, Tsuruta S, Legarra A. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. 10th WCGALP. 2014;Vancouver(Canada). <https://doi.org/10.13140/2.1.4801.5045>
- Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M. Genome-wide identification of loci associated with growth in rainbow trout. BMC Genomics. 2020;21(1):209. <https://doi.org/10.1186/s12864-020-6617-x>
- Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM. Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). BMC Genomics. 2017a;18(1):121. <https://doi.org/10.1186/s12864-017-3487-y>
- Bangera, Thorland, Hillestad, Vela, Johansen, Refstie, et al. Accuracy of genomic predictions for yield traits , flesh colour and fillet fat in Atlantic salmon ( *Salmo salar* ). In: Proceedings of the World Congresson Genetics Applied to Livestock Production. 2017b. p. 11.627.
- Blay C., Haffray P., Bugeon J., D'Ambrosio J. Dechamp N., Collewet G., et al. Genetic Parameters and Genome wide association studies of Quality traits based on Imaging phenotyping technologies in rainbow trout *Oncorhynchus mykiss*. 2020. (en préparation)
- Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50(3):362–7. <https://doi.org/10.1038/s41588-018-0056-5>
- Bugeon J, Lefevre F, Cardinal M, Uyanik A, Davenel A, Haffray P. Flesh quality in large rainbow trout with high or low fillet yield. J Muscle Foods. 2010;21(4):702–21. <https://doi.org/10.1111/j.1745-4573.2010.00214.x>
- Collewet G, Bugeon J, Idier J, Quelled S, Quittet B, Cambert M, et al. Rapid quantification of muscle fat content and subcutaneous adipose tissue in fish using MRI. Food Chem. 2013;138(2–3):2008–15. <https://doi.org/10.1016/j.foodchem.2012.09.131>
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, et al. Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). Front Genet. 2016;7(NOV). <https://doi.org/10.3389/fgene.2016.00203>
- Haffray P, Bugeon J, Pincet C, Chapuis H, Mazeiraud E, Rossignol MN, et al. Negative genetic correlations between production traits and head or bony tissues in large all-female rainbow trout (*Oncorhynchus mykiss*). Aquaculture. 2012;368–369:145–52. <https://doi.org/10.1016/j.aquaculture.2012.09.023>
- Haffray P, Enez F, Bugeon J, Chapuis H, Dupont-Nivet M, Chatain B, et al. Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. Aquaculture. 2018;490(March):350–4. <https://doi.org/10.1016/j.aquaculture.2018.03.003>
- Kause A, Mäntysaari E, Ritola O, Paananen T, Eskelinen U. Coupling body weight and its composition: A quantitative genetic analysis in rainbow trout. Aquaculture. 2002;211(1–4):65–79. [https://doi.org/10.1016/S0044-8486\(01\)00884-5](https://doi.org/10.1016/S0044-8486(01)00884-5)
- Kause A, Paananen T, Ritola O, Koskinen H. Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. J Anim Sci. 2007;95(12):3218–27. <https://doi.org/10.2527/jas.2007-0332>

- Kause A, Stien LH, Rungruangsak-Torrissen K, Ritola O, Ruohonen K, Kiessling A. Image analysis as a tool to facilitate selective breeding of quality traits in rainbow trout. *Livest Sci.* 2008;114(2–3):315–24. <https://doi.org/10.1016/j.livsci.2007.05.016>
- Leeds TD, Vallejo RL, Weber GM, Gonzalez-Pena D, Silverstein JT. Response to five generations of selection for growth performance traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture.* 2016;465:341–51. <https://doi.org/10.1016/j.aquaculture.2016.08.036>
- Marty-Mahé P, Loisel P, Fauconneau B, Haffray P, Brossard D, Davenel A. Quality traits of brown trouts (*Salmo trutta*) cutlets described by automated color image analysis. *Aquaculture.* 2004;232(1–4):225–40. [https://doi.org/10.1016/S0044-8486\(03\)00458-7](https://doi.org/10.1016/S0044-8486(03)00458-7)
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Univ Georg Athens, USA. 2018;125.
- Misztal I, Tsuruta S, Strabel T, Druet T, Lee D. BLUPF90 and related programs (BGF90). *Proc 7th World Congr Genet Appl to Livest Prod.* 2002;2.
- Moser G, Khatkar MS, Hayes BJ, Raadsma HW. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet Sel Evol.* 2010;42(1):1–15. <https://doi.org/10.1186/1297-9686-42-37>
- Neira R, Lhorente JP, Araneda C, Díaz N, Bustos E, Alert A. Studies on carcass quality traits in two populations of coho salmon (*Oncorhynchus kisutch*): Phenotypic and genetic parameters. *Aquaculture.* 2004;241(1–4):117–31 <https://doi.org/10.1016/j.aquaculture.2004.08.009>
- Neto RVR, Yoshida GM, Lhorente JP, Yáñez JM. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Mol Genet Genomics.* 2019;294(3):563–71. <https://doi.org/10.1007/s00438-018-1518-2>
- Norris AT, Cunningham EP. Estimates of phenotypic and genetic parameters for flesh colour traits in farmed Atlantic salmon based on multiple trait animal model. *Livest Prod Sci.* 2004;89(2–3):209–22. <https://doi.org/10.1016/j.livprodsci.2004.02.010>
- Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwisse THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front Genet.* 2014;5(NOV):1–8. <https://doi.org/10.3389/fgene.2014.00402>
- Powell J, White I, Guy D, Brotherstone S. Genetic parameters of production traits in Atlantic salmon (*Salmo salar*). *Aquaculture.* 2008;274(2–4):225–31. <https://doi.org/10.1016/j.aquaculture.2007.11.036>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007;81(3):559–75. <https://doi.org/10.1086/519795>
- Qiao R, Gao J, Zhang Z, Li L, Xie X, Fan Y, et al. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. *Genet Sel Evol.* 2015;47(1):1–13. <https://doi.org/10.1186/s12711-015-0089-5>
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 2014;15(1). <https://doi.org/10.1186/1471-2164-15-478>
- Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD. The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*). *BMC Genet.* 2015a;16(1). <https://doi.org/10.1186/s12863-015-0215-y>
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and

- genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015b;16(1):1–9. <https://doi.org/10.1186/s12864-015-2117-9>
- Tsai HY, Matika O, Edwards SMK, Antolín-Sánchez R, Hamilton A, Guy DR, et al. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 Genes, Genomes, Genet*. 2017;7(4):1377–83. <https://doi.org/10.1534/g3.117.040717>
- Tsairidou S, Hamilton A, Robledo D, Bron JE, Houston RD. Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. *G3 Genes, Genomes, Genet*. 2020;10(2):581–90. <https://doi.org/10.1534/g3.119.400800>
- Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, et al. 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*. 2017;49(1):1–13. <https://doi.org/10.1186/s12711-017-0293-6>
- Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Sixin L, Parsons JE, et al. 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*. 2018;(February):1–12. <https://doi.org/10.1111/jbg.12335>
- Vandeputte M, Bugeon J, Bestin A, Desgranges A, Allamellou JM, Tyran AS, et al. First Evidence of Realized Selection Response on Fillet Yield in Rainbow Trout *Oncorhynchus mykiss*, Using Sib Selection or Based on Correlated Ultrasound Measurements. *Front Genet*. 2019;10(December):1–13. <https://doi.org/10.3389/fgene.2019.01225>
- Vandeputte M, Haffray P. Parentage assignment with genomic markers: A major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front Genet*. 2014;5(DEC):1–8. <https://doi.org/10.3389/fgene.2014.00432>
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb)*. 2012;94(2):73–83. <https://doi.org/10.1017/S0016672312000274>
- Weng Z, Wolc A, Shen X, Fernando RL, Dekkers JCM, Arango J, et al. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genet Sel Evol*. 2016;48(1):1–10. <https://doi.org/10.1186/s12711-016-0198-9>
- Wolc A, Arango J, Settar P, Fulton JE, O’Sullivan NP, Preisinger R, et al. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet Sel Evol*. 2011;43(1):1–8. <https://doi.org/10.1186/1297-9686-43-23>
- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46(11):1173–86. <https://doi.org/10.1038/ng.3097>
- Yoshida GM, Carvalheiro R, Lhorente JP, Correa K, Figueroa R, Houston RD, et al. Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture*. 2018;491(March 2017):147–54. <https://doi.org/10.1016/j.aquaculture.2018.03.004>



## Chapitre 5: Architecture génétique et sélection pour la résistance à la nécrose pancréatique infectieuse

### 1. Introduction

En aquaculture, la résistance aux maladies est l'un des caractères les plus sélectionnés après les performances de croissance, découpe et qualité de la chair (Chavanne et al. 2016; Abdelrahman et al. 2017). Les maladies peuvent avoir un impact très important sur les performances économiques des fermes aquacoles, mais aussi dans l'économie d'une entreprise de sélection en fonction des stades auxquelles elles impactent les cheptels, avec parfois l'obligation d'abattage total des animaux d'une pisciculture pour limiter la propagation des agents pathogènes de maladies réputées légalement contagieuses (Bishop et Woolliams 2014). L'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail) place la nécrose pancréatique infectieuse à la 6<sup>ème</sup> position dans la hiérarchisation des dangers sanitaires présents en France pour les poissons d'élevage (ANSES 2015). L'impact économique de cette maladie est jugé moyen à fort (note de 6/10, ANSES 2015) en Europe, avec une mortalité importante des alevins (jusqu'à 90%) et l'exigence de garanties complémentaires pour l'envoi d'œufs et d'alevins dans des pays au statut sanitaire équivalent au pays d'envoi au sein de l'union Européenne (Décision 2010/221/UE).

La nécrose pancréatique infectieuse ou IPN (Infectious Pancreatic Necrosis) est une maladie très contagieuse, mais qui ne nécessite pas l'abattage des animaux dans les piscicultures touchées. Elle est causée par un virus de la famille des Birnaviridae et du genre *Aquabirnavirus* (Dopazo 2020). Ce virus touche essentiellement les salmonidés, mais il a été aussi isolé chez d'autres espèces de poissons en milieu naturel (Dopazo 2020). L'IPN est connue dans le monde entier avec des épisodes initiaux décrits en Amérique du Nord et en Europe, mais également, depuis les années 80, dans les pays asiatiques (Dopazo 2020). Le virus se transmet horizontalement (entre individus par la contamination de l'eau) et verticalement (d'un reproducteur à sa descendance), potentiellement *in ovo*. Les survivants d'une infection deviennent des porteurs sains durant plusieurs années (Dopazo 2020). Les *Aquabirnavirus* peuvent être très résistants et survivre pendant de longues périodes dans de multiples conditions environnementales (Dopazo 2020). L'IPN touche essentiellement les jeunes alevins après l'absorption du sac vitellin et la première alimentation (Dopazo 2020). Les symptômes

principaux sont un comportement inhabituel (anorexie, nage anormale), le noircissement de la peau et des hémorragies (Dopazo 2020). Sans intervention humaine, l'IPN a un fort potentiel de diffusion horizontal, de persistance et d'évolution avec un très large spectre d'hôtes et des différences de virulence importantes entre souches virales (ANSES 2015).

Storset et al. (2007) ont montré une forte réponse à la sélection chez le saumon Atlantique pour la résistance à l'IPN. Ils ont comparé dans un même challenge expérimental et au sein d'une même population les taux de mortalité de familles sélectionnées sur une génération soit pour une forte résistance, soit pour une faible résistance à l'IPN. Pour les familles hautement résistantes à l'IPN, les taux de mortalité ont été de 29% et 30% respectivement. Pour les familles à faible résistance pour l'IPN, ils ont été de 67% et 79% respectivement en eau de rivière et en mer.

Chez le saumon Atlantique, un QTL majeur a été trouvé sur le groupe de liaison 21 (chromosome 26) expliquant 29% et de la variance phénotypique (soit 89% de la variance génétique) en condition d'épisode infectieux naturel dans une population écossaise (Houston et al. 2008), ainsi qu'en challenge expérimental dans une population norvégienne (Moen et al. 2009).

Différentes études chez la truite arc-en-ciel ont montré une variabilité génétique pour la résistance à l'IPN avec des héritabilités de 0,35 et 0,39 en GBLUP pour la durée de vie post-challenge et la survie à la fin du challenge, respectivement (Yoshida et al. 2019) et 0,53 et 0,82 en modèle BayesC $\pi$  (Rodríguez et al. 2019). Mais il n'existe que peu de travaux visant à caractériser précisément l'architecture de la résistance à l'IPN chez la truite arc-en-ciel (Ozaki et al. 2001, 2007; Rodríguez et al. 2019). Dans une famille de backcross expérimental entre lignées japonaises résistante et sensible, Ozaki et al. (2001) ont identifiés 2 QTLs qui expliquaient 27 et 34% de la variance phénotypique, montrant que des QTLs majeurs de résistance à l'IPN pouvaient être détectés dans ces populations de truite arc-en-ciel. D'après Rodriguez et al. (2019), ces QTLs sont situés sur les chromosomes Omy14 et Omy16 et ne correspondent à aucun des QTLs détectés dans leur population chilienne (Yoshiad et al. 2019). Aucun des QTLs identifiés chez la truite arc-en-ciel ne correspond au gène majeur identifié chez le saumon Atlantique (Moen et al. 2015).

Les avancées des concurrents, la diversité des situations rapportées et la potentialité d'identifier des QTL d'intérêt dans les lignées des sélectionneurs français justifient l'étude de l'architecture de la résistance à l'IPN dans les populations françaises de truite arc-en-ciel afin de pouvoir mettre en œuvre soit une sélection assistée par marqueurs (SAM) pour fixer des QTLs majeurs

qui seraient identifiés dans ces populations, soit une sélection génomique (SG) si la résistance à l'IPN est très polygénique. Les deux objectifs de cette étude sont donc de (i) étudier l'architecture de la résistance à l'IPN dans la population du sélectionneur Bretagne-Truite et (ii) estimer l'efficacité de la SG par rapport à une sélection sur pedigree.

## 2. Matériels et méthodes

### 2.1. Caractéristiques de la population étudiée et du challenge expérimental

Les animaux proviennent de l'entreprise Bretagne-Truite regroupant différents sites piscicoles en Bretagne et en Normandie (<http://www.bretagne-truite.fr/>). La cohorte étudiée provient de la 7<sup>ème</sup> génération d'une sélection multicaractères combinant sélection massale de type PROSPER (Chevassus et al. 2004) sur la croissance et amélioration du rendement à l'éviscération par sélection sur apparentés assistée par ultrasons (Haffray et al. 2013).

Pour tester la résistance des animaux à l'IPN, un challenge expérimental a été mis en place. Pour réaliser ce challenge, 3 lots de 3 000 œufs ont été fournis par Bretagne Truite à la plateforme SYSAAF-ANSES Fortior Genetics (Technopôle Brest F29280 PLOUZANE). Les œufs ont été incubés en eau de rivière à température régulée à 11°C ( $\pm 1^\circ\text{C}$ ). Les alevins ont été répartis au hasard en 3 lots A, B (challengés en duplicat) et témoin (non challengé) de 2 000 alevins d'un poids moyen de 0,17g dans des bacs de 50L.

Les poissons du bac témoin ont été contrôlés négativement pour la présence du virus de l'IPN et aucune mortalité n'a été observée à 25 jours.

Les deux bacs des lots A et B ont été contaminés par balnéation à J21 post-nourrissage avec la souche virale réactivée NN193 (souche isolée du virus IPN à Bretagne Truite suite à un épisode de mortalité) à une dose d'environ  $10^{E+5}$  DICT<sub>50</sub>/mL (DICT<sub>50</sub> : dose infectieuse sur 50% de culture tissulaire). La préparation virale du lot A était congelée lors de la contamination et non congelée pour le lot B. La mortalité a été suivie quotidiennement durant un mois (du 8 janvier au 3 février 2016). Les prélèvements de nageoires ont été conservés dans des pots à prélèvement remplis d'alcool (>95°), identifiés par date et conservés à  $+5^\circ\text{C} \pm 3^\circ\text{C}$ . L'observation des symptômes cliniques caractéristiques de l'IPN a été faite régulièrement et des analyses

virologiques ont été réalisées jusqu'à 25 jours post-challenge. Au terme de l'épreuve infectieuse, les poissons survivants ont été euthanasiés.

Les premières mortalités ont été observées 6 jours après la contamination expérimentale. Les analyses virologiques réalisées sur les organes de poissons morts à 6 jours après la contamination expérimentale ont mis en évidence la présence du virus de l'IPN dans les tissus des poissons morts. La mortalité finale (évaluée à 23 jours) est de 91,8% pour le lot B et 41,0% pour le lot A. Cette forte différence de mortalité entre les 2 lots pourrait être expliquée par la congélation de la préparation virale du lot A associée à une diminution de la virulence.

L'analyse des données de challenge est donc réalisée seulement pour le lot B. Les phénotypes étudiés sont la durée de vie post-challenge (TD, en jours) et la survie à 6 jours (D6, statut 0/1).

## 2.2. Génotypages

1 503 individus ont été génotypés avec la puce Axiom® Trout Genotyping array (57 501 SNP). L'ensemble des individus (1 310 individus challengés et 193 parents : 96 pères, 91 mères et 6 reproducteurs sans descendant challengé) passe le contrôle de qualité avec au moins 95% des SNPs génotypés. 4 individus n'ont ni père et ni mère connus, 1 individu a le père inconnu, 1 individu a sa mère inconnue et 22 individus ont un père identifié dans le pedigree mais non génotypé. La structuration familiale est similaire aux populations étudiées précédemment (Tableau 26).

Effectifs challengés	moyenne $\pm$ SD	minimum	maximum
Descendants par couple	2 $\pm$ 1	1	6
Descendants par père	13 $\pm$ 4	6	20
Descendants par mère	14 $\pm$ 4	4	21
Demi-sœurs par individu challengé	25 $\pm$ 5	2	37

**Tableau 26 :** Structuration familiale des 679 familles identifiées de la cohorte challengée et génotypée.

Un contrôle qualité a été appliqué aux 38 350 SNPs exploitables de la puce 57K pour ne retenir que les SNPs dont le call-rate est supérieur à 97%, la MAF supérieure à 5% et le test d'écart à l'équilibre d'Hardy-Weinberg a une p-value  $>$  0,0001 sur les 1 503 individus génotypés. Cela conduit à conserver 30 437 SNPs. Les contrôles qualité ont été effectués avec Plink 1.9 (Purcell et al. 2007).

Le génotype complet pour les 30 437 marqueurs a été obtenu par imputation à l'aide du logiciel Fimpute v3 (Sargolzaei et al. 2014). En comparant les génotypes des parents et des descendants, 4 SNPs avec plus de 100 erreurs mendéliennes ont été écartés de l'analyse pour retenir au final 30 433 SNPs. Le test des incompatibilités mendéliennes réalisé par Fimpute sur les 30 433 marqueurs a permis de contrôler la qualité du pedigree fourni. Des corrections sur le pedigree ont été apportées pour 8 individus et une inversion entre 2 mères pour 31 individus sur la plaque de génotypage a été identifiée.

### 2.3. Modèle animal et évaluations BLUP et GBLUP

Pour prédire les valeurs génétiques et estimer les paramètres génétiques, un modèle animal est considéré pour décrire les performances.

$$\mathbf{y} = \mu + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

où  $\mathbf{y}$  est le vecteur des performances pour le caractère (TD ou D6),  $\mu$  la moyenne du caractère,  $\mathbf{Z}$  la matrice d'incidence pour le vecteur  $\mathbf{u}$  des valeurs génétiques des animaux et  $\mathbf{e}$  le vecteur des résidus du modèle.

En évaluation génétique classique, la résolution des équations de ce modèle est basée sur la méthodologie BLUP et l'information du pedigree permettant de construire la matrice de parenté  $\mathbf{A}$ . Le pedigree est composé de 31488 individus remontant les 7 générations de sélection de la lignée étudiée.

En évaluation génomique GBLUP, la matrice de parenté sur pedigree ( $\mathbf{A}$ ) est simplement remplacée par une matrice de parenté génomique ( $\mathbf{G}$ ).

Les (G)EBV et les composantes de la variance sont obtenues à l'aide des packages BLUPf90 et AIREMLf90 (Miszta et al. 2002, 2018). Les GEBV permettent de calculer les effets estimés des SNPs grâce au package postGSF10 (Aguilar et al. 2014).

Afin d'identifier des QTLs sur l'ensemble des caractères étudiés, une analyse d'association pangénomique basée sur le GBLUP a été réalisée. Une région du génome a été considérée comme un QTL quand un SNP de la région avait un effet significatif à un seuil de 5% à l'échelle du génome (après correction de Bonferroni), correspondant à un  $-\log(p\text{-value})$  supérieur à 5,78. Le QTL a été considéré comme simplement suggestif si  $-\log(p\text{-value})$  est compris entre 4,30 et

5,78, ce qui correspond à un seuil de significativité à 5% à l'échelle du chromosome après correction de Bonferroni. L'intervalle de confiance d'un QTL a été estimé par la méthode du drop-off (Li 2011) en considérant une limite inférieure du  $-\log(p\text{-value})$  correspondant à une valeur réduite de -1,5 à partir de la valeur maximale atteinte au SNP pic du QTL. Tous les SNPs dans le voisinage du pic SNP (avec une fenêtre coulissante de 1 Mb) et un  $-\log(p\text{-value})$  au-dessus de la limite inférieure de drop-off sont inclus dans l'intervalle de confiance du QTL.

## 2.4. Modèle BayesC $\pi$ et sélection de variables

Une approche bayésienne de sélection de variables a été utilisée pour mieux estimer les effets des SNPs et leurs variances en considérant le problème statistique du nombre très élevé de paramètres (les effets de SNPs) à estimer par rapport au nombre de données disponibles (Habier et al. 2011). L'approche BayesC $\pi$  consiste à considérer un modèle de mélange avec une proportion  $\pi$  de SNPs ayant des effets non nuls (distribués selon une loi normale de moyenne nulle et variance suivant une loi de Chi2-inverse) et une proportion  $(1-\pi)$  des SNPs sans effet à chaque cycle d'estimation des paramètres par méthode de MCMC. Le modèle d'analyse des performances est le suivant :

$$y_i = \mu + u_i + \sum_{j=1}^n \delta_j z_{ij} a_j + e_i$$

avec  $y_i$  la performance de l'animal  $i$ ,  $\mu$  la moyenne du caractère,  $u_i$  la valeur de l'effet polygénique,

$n$  le nombre total de SNPs,  $z_{ij}$  le génotype (codé 0 et 2 pour les homozygotes et 1 pour l'hétérozygote) au locus  $j$  pour animal  $i$ ,  $a_j$  l'effet du SNP  $j$ ,  $e_i$  le résidu,  $\delta_j$  est une indicatrice : à l'itération  $k$ ,  $\delta_j=1$  si le SNP  $j$  est inclus dans le modèle (probabilité  $\pi$ ) et  $\delta_j=0$  s'il n'est pas inclus (probabilité  $1-\pi$ ).

L'estimation du paramètre  $\pi$  se fait sous l'hypothèse d'une distribution bêta  $B(\alpha, \beta)$  de paramètres  $\alpha$  et  $\beta=30\ 433$ .

Dans un premier modèle sans effet polygénique  $u_i$  considéré, on a fixé  $\alpha$  à 300 pour qu'en espérance 300 SNPs soient retenus à chaque cycle comme ayant un effet non-nul parmi les 30 433 SNPs inclus dans l'analyse. La valeur moyenne de  $\pi$  a ainsi été considérée proche de 1%. Pour établir ce choix, différents tests ont été effectués afin d'effectuer un compromis

raisonnable entre le nombre limité de données disponibles (environ 1 300) et le fait que les caractères semblaient très polygéniques.

Dans un second modèle, nous avons considéré un effet additif polygénique outre les effets des SNPs pour rendre compte du fait que les marqueurs ne semblaient pas pouvoir capter toute la variabilité génétique. On a alors considéré  $\alpha=30$  pour essayer de préciser au mieux les effets et la localisation des principaux QTLs.

Pour chaque analyse, l'algorithme MCMC effectue 600 000 cycles avec une phase de chauffe de 5 000 cycles ; l'algorithme est lancé deux fois avec une racine différente afin d'étudier la qualité de la convergence. Les résultats sont sauvegardés tous les 40 cycles pour estimer les paramètres. Les analyses ont été effectuées avec le logiciel BESSiE (Boerner and Tier 2016).

L'importance de l'association entre un SNP et le phénotype est appréciée par le Bayes Factor (BF), calculé avec la formule suivante (Schurink et al. 2012) :

$$BF = \left(\frac{p_i}{1-p_i}\right) / \left(\frac{\pi}{1-\pi}\right)$$

avec  $p_i$  la probabilité d'inclusion du SNP  $i$  dans le modèle.

Selon l'approche proposée par Kass et Raftery (1995), le BF est transformé en  $\log BF = 2 * \ln(BF)$  pour avoir des valeurs dans l'échelle usuelle de valeurs observées dans les tests de déviance et ratio de vraisemblance. Un seuil de  $\log BF \geq 5$  est considéré pour définir la réelle présence d'un QTL et son intervalle de crédibilité est calculé autour du SNP pic ( $\log BF$  maximal) en intégrant dans la région QTL tous les SNPs avec un  $\log BF \geq 2$  dans une fenêtre coulissante de 1Mb à partir du pic SNP. Nous avons également considéré comme QTLs suggestifs les zones du génome où un pic SNP avait un  $\log BF \geq 4$ .

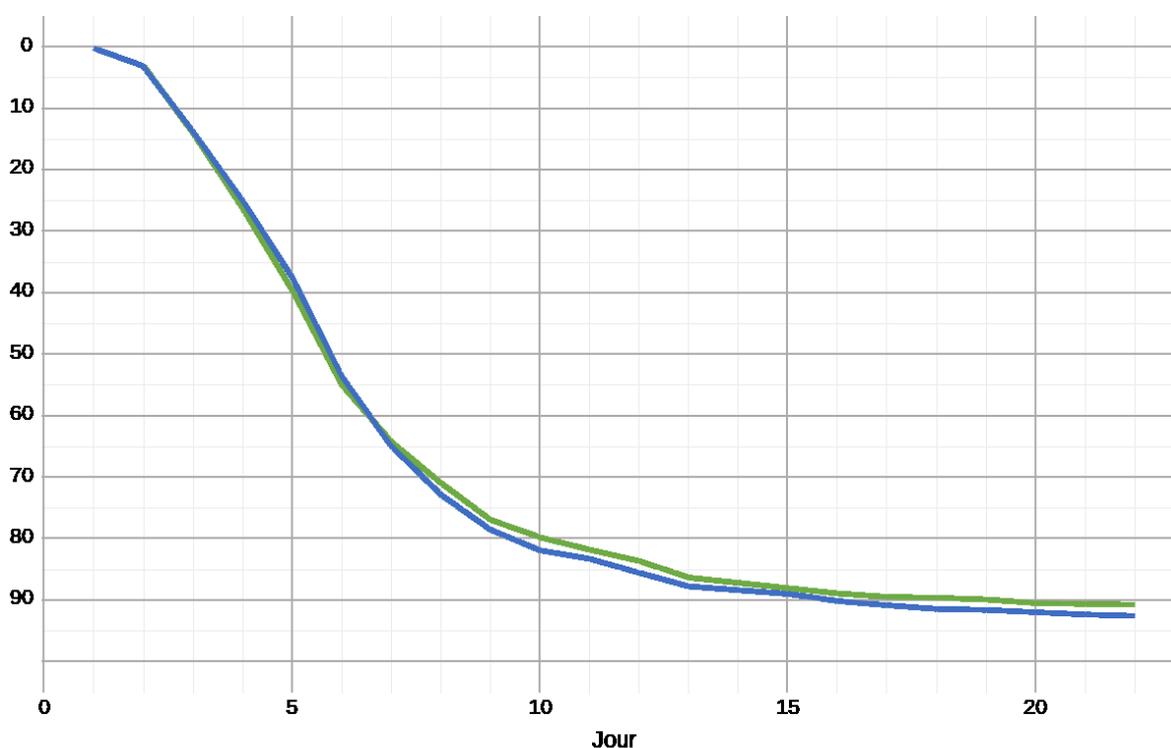
### 3. Résultats & discussion

#### 3.1. Challenge expérimental

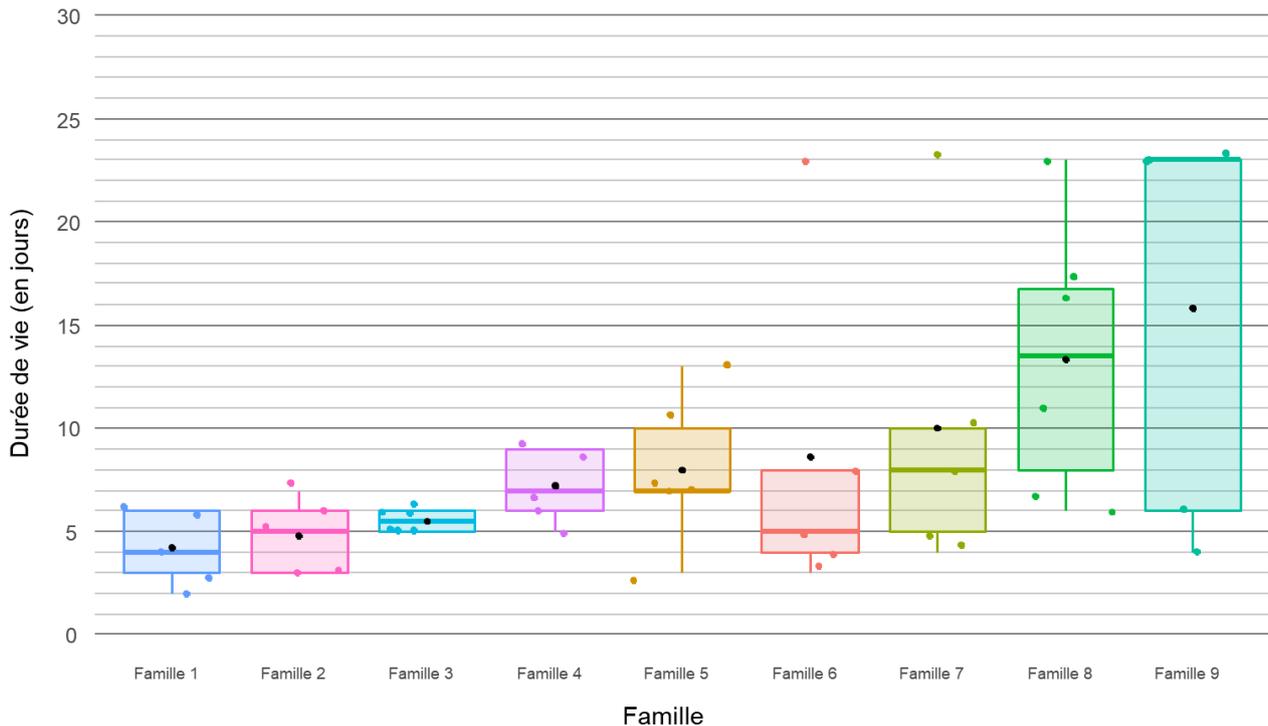
Le taux de mortalité dans notre étude est de 91,8% à 22 jours et de 54% à 6 jours post-infection, (Figure 31). Le taux de mortalité d'un challenge est lié à différents paramètres comme la sensibilité des populations et les conditions expérimentales telles que les modalités de

l'infection (baignade ou injection), la souche de virus utilisée et le poids des poissons au challenge.

On observe une forte hétérogénéité de survie entre les familles (Figure 32). Dans certaines familles, la majorité des individus sont morts au 6eme jour et dans d'autres familles la majorité des individus sont encore vivants après 6 jours (Figure 32). Qui plus est, au sein d'une même famille, des individus sont morts très rapidement (<6 jours) et des individus survivent à l'expérimentation (Figure 32). La sélection génomique permettrait donc de choisir les individus les plus résistants dans chaque famille, ce qui ne peut être fait par sélection sur pedigree.



**Figure 31 :** Mortalité cumulée en fonction des jours post-infection pour les individus génotypés (courbe bleue) et non génotypés (courbe verte).



**Figure 32 :** Boîte à moustache de la durée de vie post-infection (en jours) pour les 9 familles ayant au moins 5 descendants challengés et génotypés .

### 3.2. Paramètres génétiques

Les héritabilités estimées sont faibles et similaires pour les deux caractères, mais variables selon le modèle d'analyse génétique considéré (de 0,03 à 0,09) (Tableau 2). On observe pour les deux caractères une  $h^2$  plus faible en GBLUP ou BayesC $\pi$  strictement génomique par rapport au BLUP. Cela s'explique sans doute en partie par le fait que ce n'est que la cohorte récente qui est génotypée, mais aussi par le fait que la couverture du génome en marqueurs est insuffisante pour capter toute la variabilité génétique de ces caractères. Aussi, le modèle bayésien combinant information des marqueurs et effet polygénique permet d'obtenir des estimations d'héritabilité se rapprochant de l'évaluation BLUP sur pedigree. Dans le modèle BayesC $\pi$  avec une composante polygénique sur pedigree, la variance polygénique représente environ 80% de la variance génétique. Ce résultat conforte l'hypothèse que nous n'avons pas assez de marqueurs dans les zones d'intérêt pour capturer la variabilité génétique avec les marqueurs, mais et surtout à mettre en relation avec le fait que les caractères sont très polygéniques et à faible héritabilité, et que le dispositif de taille insuffisante induit alors une sous-estimation sans doute forte des effets des SNPs.

## TD

Modèle génétique	Var. poly.	Var. SNP	Var. génétique	Var. résiduelle	h <sup>2</sup>
BLUP	--	--	2,40 (0,93)	25,10 (1,16)	0,09 (0,03)
GBLUP (SNP)	--	--	1,55 (0,76)	28,20 (1,27)	0,05 (0,03)
GBLUP (SNP+pedigree)	--	--	1,53 (0,74)	28,24 (1,26)	0,05 (0,02)
BAYESCπ sans effet polygénique	--	1,20 (0,55)	1,20 (0,55)	28,58 (1,21)	0,04
BAYESCπ avec effet polygénique	1,66 (0,76)	0,41 (0,31)	2,07 (0,78)	27,84 (1,30)	0,07

## D6

Modèle génétique	Var. poly.	Var. SNP	Var. génétique	Var. résiduelle	h <sup>2</sup>
BLUP	--	--	204 (80)	2284 (104)	0,08 (0,03)
GBLUP (SNP)	--	--	154 (63)	2329 (104)	0,03 (0,03)
GBLUP (SNP+pedigree)	--	--	151 (61)	2334 (103)	0,06 (0,02)
BAYESCπ sans effet polygénique	--	127 (49)	127 (49)	2357 (101)	0,05
BAYESCπ avec effet polygénique	151 (64)	42 (29)	193 (65)	2303 (107)	0,08

**Tableau 27 :** Composantes de la variance pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) selon différents modèles génétiques (ecart-type d'erreur entre parenthèses) var. : variance, var. poly. : variance polygénique, h<sup>2</sup> : héritabilité du caractère.

De précédentes études chez la truite arc-en-ciel de la la résistance à l'IPN rapportent des héritabilités beaucoup plus élevées que dans notre études, de l'ordre de 0,55 (Wetten et al. 2011) sur des poissons de 0,15 g, 0,39 (Flores-Mara et al. 2017) sur des poissons d'environ 2,24g une héritabilité à 0,53 pour TD et 0,82 pour la survie à la fin du challenge (63 jours) en modèle BayesCπ (Rodríguez et al. 2019). Sur la même population, des estimations précédentes en BLUP et GBLUP respectivement avaient abouti a des estimations de 0,25 et 0,40 pour TD et de 0,24 et 0,35 pour la survie (Yoshida et al. 2019).

Le décalage est très important entre les estimations d'héritabilité rapportées par les auteurs précédents et cette étude. Plusieurs hypothèses peuvent être évoquées pour tenter de comprendre ce résultat et d'éventuellement y remédier :

- Le taux de mortalité élevé et rapide dans notre étude est une première piste d'explication car cela peut écraser les différences observables entre parents en ne permettant pas une expression suffisante de la variabilité génétique. Wetten et al. (2011) rapportent par exemple une mortalité de seulement 26,3% à 35 jours après le début du challenge alors que notre mortalité est déjà de 54% 6 jours après challenge et de 90% 20 jours post-challenge. Flores-Mara et al. (2017) et Rodríguez et al. (2019) rapportent une mortalité très basse à 13,8% à 63 jours post-challenge.

- L'origine de la souche virale ne semble pas à remettre en cause, celle-ci ayant été isolée à partir de l'élevage de sélection. Par contre sa virulence exceptionnelle pourrait ne pas représenter la virulence moyenne en élevage en 2020.
- Le protocole de challenge constitue une troisième piste à explorer. Dans nos travaux tous les alevins ont été infectés par balnéation, comme dans toutes les études précédemment citées. Une approche différente pourrait consister à introduire dans le lot à challenger des poissons moribonds préalablement infectés avec cette souche virale. Cette approche pourrait peut-être permettre un développement moins brutal de la maladie et une mise en place des systèmes de défense immunitaire plus en phase avec les différences de potentiel génétique des individus.
- Un autre aspect pourrait aussi consister à challenger des poissons un peu plus vieux (ou plus lourds) bien que le poids au challenge corresponde ici à la mise en place des défenses immunitaires essentiellement innées.
- Le fond génétique de la lignée pourrait aussi être évoqué. La lignée pourrait être particulièrement sensible. Cependant, cette hypothèse est surprenante, car les truites élevées en France ont subi de très fortes mortalités à l'IPN dans les années 80 et des variants résistants ont dû être fixés à cette époque.

L'optimisation du challenge avec l'utilisation d'une souche virale moins virulente constitue toutefois la piste la plus simple à privilégier pour améliorer l'efficacité de la sélection.

### 3.3. QTL

L'architecture génétique des deux caractères de résistance à l'IPN s'avère être très polygénique (BayesC $\pi$  : [Figure 33](#), GBLUP : [Annexe - figure 2](#)). Quelques QTLs sont identifiés, mais restent avant tout seulement suggestifs et expliquent moins de 0,5% de la variance génétique. Les principaux pics identifiés sont résumés dans le [Tableau 28](#). Aucun QTL n'a une évidence forte :  $\log_{BF} < 8$  en modèle bayésien et  $-\log(p\text{-value})$  non significative en GBLUP.

Pour les 2 caractères TD et D6, on détecte 2 QTLs communs sur les chromosomes 4 et 17. L'intervalle de crédibilité sur le chromosome 17 est d'un seul SNP pour D6 et de 0,82 Mb pour TD. Le SNP peak se situe dans un gène (17,058– 17,117 Mb) codant pour une protéine

« ubiquitin carboxyl-terminal hydrolase 34-like ». Pour le QTL localisé sur le chromosome 4, l'intervalle de crédibilité est trop important (1,67 Mb) pour réellement proposer une liste de gènes candidats. Le SNP peak se situe à proximité du gène pour le récepteur Eph A7 (Epha7 : 56,534 – 56,400 Mb) dont des mutations ont un effet sur le système nerveux de la souris (ID MGI:3583808, MGI:3575006 d'après la base de donnée : Mouse Genome Database (Bult et al. 2019)).

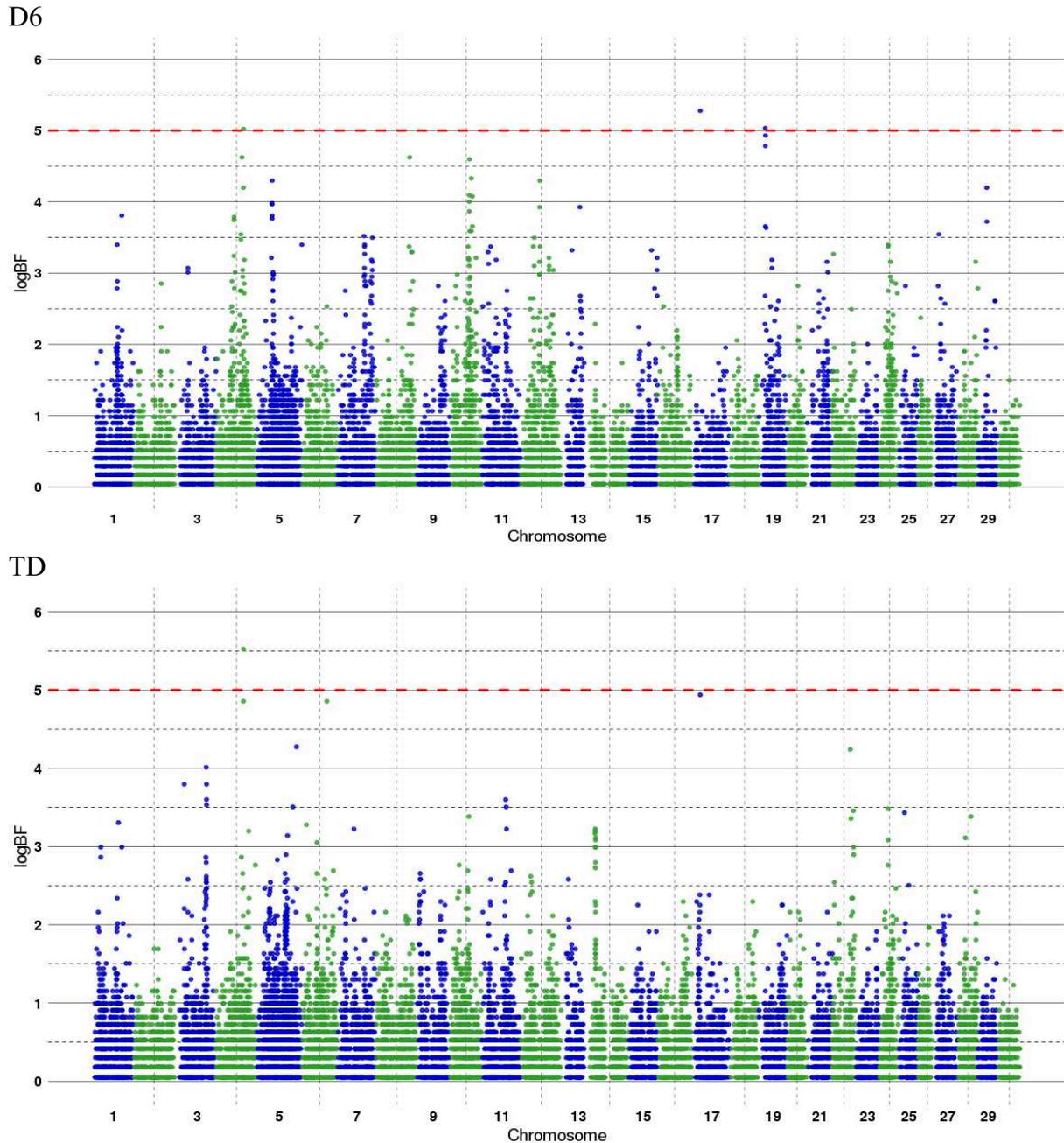
Pour le caractère D6, nous avons également une zone d'intérêt sur le chromosome 10 avec un QTL putatif localisé entre 36,4 Mb et 41,5 Mb. Notons que cette région correspond à une zone très homozygote dans de nombreuses populations françaises (D'Ambrosio et al. 2019) et qu'un QTL pour la résistance à *Flavobacterium psychrophilum* a été identifié entre 41,4 et 49,9 Mb dans la même lignée du sélectionneur Bretagne-Truite (Fraslin et al. 2019) ainsi que dans un lignée américaine (Vallejo et al. 2017b). Le dernier QTL identifié pour D6 est situé dans une région (12,15 – 13, Mb) du chromosome 19 où beaucoup de gènes non identifiés sont présents. Notons toutefois qu'un QTL pour la résistance à la flavobactériose est localisé entre 10 et 20 Mb sur le chromosome 19 (Wiens et al. 2013). Peut-être des mécanismes génériques et précoces de résistance aux infections virales comme bactériennes sont donc à rechercher au niveau des chromosomes 10 et 19 du génome de la truite.

Trait	SNP peak			GBLUP	BayesCpi		
	Omy	Name SNP	Position (pb)	-Log (p-value)	LogBF	Intervalle de crédibilité (Mb)	Var* (%)
<b>TD</b>	4	Affx-88955500	56 407 317	2,89	5,52	54,74-56,41	0,13
<b>TD</b>	6	Affx-88908261	51 326 069	2,83	4,86	50,38-51,33	0,05
<b>TD</b>	17	Affx-88923189	17 080 168	4,15	4,94	16,20-17,08	0,04
<b>TD</b>	24	Affx-88933593	33 276 040	4,71	7,17	33,28-33,50	0,12
<b>D6</b>	4	Affx-88955500	56 407 317	3,01	5,02	54,74-58,07	0,17
<b>D6</b>	10	Affx-8895573	40 244 556	3,40	4,60	36,44-41,52	0,33
<b>D6</b>	17	Affx-88923189	17 080 168	4,19	5,28	17,08-17,08	0,03
<b>D6</b>	19	Affx-88937347	12 951 375	3,78	5,03	12,15-13,36	0,11

**Tableau 28** : Statistiques descriptives des QTL détectés pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) par approche GBLUP et BayesCπ.

Var\* : proportion de la variance génétique expliquée par l'ensemble des SNPs inclus dans l'intervalle de crédibilité du QTL ; SNP speak : SNP avec le log(p-value) ou LogBF le plus élevés.

Pour le caractère TD, le QTL avec la plus forte évidence est identifié sur le chromosome 24 dans une région comprenant un seul gène annoté dans l'intervalle de crédibilité codant pour la cholinesterase-like (33,415 – 33,43 Mb).



**Figure 33 :** Manhattan plot des QTLs détectés en analyse BayesC $\pi$  (incluant un effet polygénique) pour la résistance à l'IPN. La ligne pointillée rouge correspond au seuil de  $\log\text{BF} > 5$  pour définir la présence d'un QTL. TD : durée de vie post-infection, D6 : survie à 6 jours.

Concernant le QTL situé sur le chromosome 6, le SNP peak est situé entre deux gènes : NAV2 (neuron navigator 2-like : 51,129 – 51,260 Mb) et E2F8 (transcription factor E2F8-like : 51,463 – 51,474 Mb). Le gène NAV2 appartient à la famille des gènes « neurone navigateur » jouant un rôle dans la croissance et la migration cellulaire. Des souris MGI mutantes pour NAV2 ont des anomalies d'odorat, de vision, des pertes de poids, un comportement et une structure du squelette anormaux (ID MGI:6156559 ; MGI : 3054375 ; MGI:2387958). E2F8 est un facteur

de transcription régulant l'expression de gènes requis pour la progression du cycle cellulaire pour que le noyau de la cellule se divise au bon moment. Des souris MGI pour E2F8 peuvent ne présenter aucune anomalie, ou bien des anomalies respiratoires, une mortalité et des anomalies embryonnaires, une croissance retardée et des anomalies du système cardiovasculaire (ID MGI:3773723).

Pour conclure, aucun gène candidat évident ne peut être proposé suite à notre étude et les zones d'intérêt que nous avons mises en évidence ne sont pas retrouvées dans les travaux antérieurs concernant la résistance à l'IPN chez la truite arc-en-ciel. Les QTLs identifiés dans la littérature se retrouvent sur les chromosomes 14, 16 (Ozaki et al. 2001, 2007) et les chromosomes 5 et 23 (Rodríguez et al. 2019) et un brevet a été déposé sur des zones du chromosome 1 (Santi et al. 2019).

### 3.4. Efficacité de la sélection sur pedigree et génomique

Pour étudier l'efficacité de la SG (GBLUP) *versus* la sélection sur pedigree (BLUP), nous avons effectué de la cross-validation avec une population d'apprentissage de 1 000 individus et une population de validation de 310 individus, en échantillonnant aléatoirement 40 fois la population totale de 1 310 individus. Les moyennes et écart-types d'échantillonnage des précisions et coefficients de dispersion obtenues pour les (G)EBV sont résumés dans le [Tableau 29](#).

Trait	BLUP		GBLUP	
	r	b	r	b
TD	0,24 (0,18)	0,74 (0,59)	0,27 (0,18)	0,63 (0,44)
D6	0,33 (0,18)	0,93 (0,53)	0,36 (0,15)	0,76 (0,33)

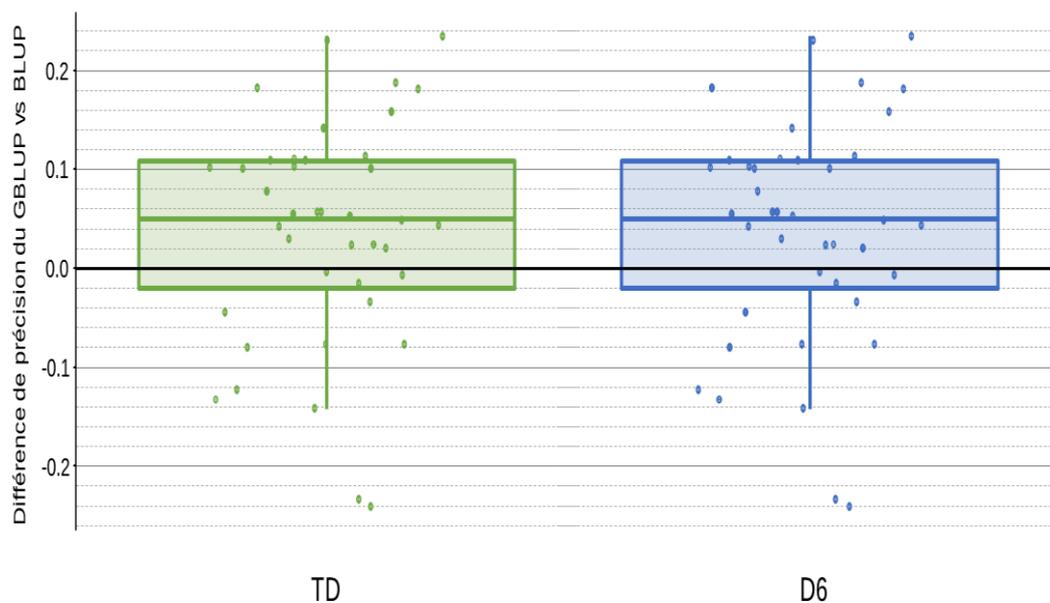
**Tableau 29** : Précision moyenne (r) et coefficient moyen de dispersion (b) d'évaluations BLUP et GBLUP pour la résistance à l'IPN :

moyennes calculées sur 40 simulations avec 1 000 individus en population d'apprentissage et 310 individus dans la population de validation (écarts types d'échantillonnage entre parenthèses). D6 : survie à 6 jours et TD : durée de vie post-infection

Sur l'ensemble des évaluations, les précisions sont plus faibles et les (G)EBV plus surestimées ( $b \ll 1$ ) pour TD que pour D6. Nous avons un gain de précision du GBLUP *versus* BLUP

d'environ 9% et 13% respectivement pour D6 et TD. La précision du GBLUP est supérieure dans 27 des échantillonnages réalisés pour les deux caractères (**Figure 34**). L'ensemble des précisions obtenues sur les 40 échantillonnages réalisés pour le BLUP et GBLUP sont résumées en **Annexe - figure 3**.

Yoshida et al. (2019) estiment une précision de 0,49 en BLUP et 0,53 en ssGBLUP pour TD et 0,50 en BLUP et 0,56 en ssGBLUP pour une variable mort/vivant (63 jours post-infection) avec une population d'apprentissage comprenant 577 individus génotypés et phénotypés et 1 557 individus seulement phénotypés, soit un gain de précision en ssGBLUP de 7 et 11% par rapport à une évaluation BLUP. Dans notre étude, la sélection génomique permettrait également un gain de précision d'environ 10% pour la résistance à l'IPN, mais nos précisions sont beaucoup plus faibles que celles obtenues par Yoshida et al. (2019) sans doute en raison des faibles héritabilités observées dans notre dispositif.



**Figure 34 :** Boîtes à moustaches du gain de précision du GBLUP versus BLUP pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) pour les 40 simulations avec une population d'apprentissage de 1 000 individus et une population de validation de 310 individus.

#### 4. Résumé

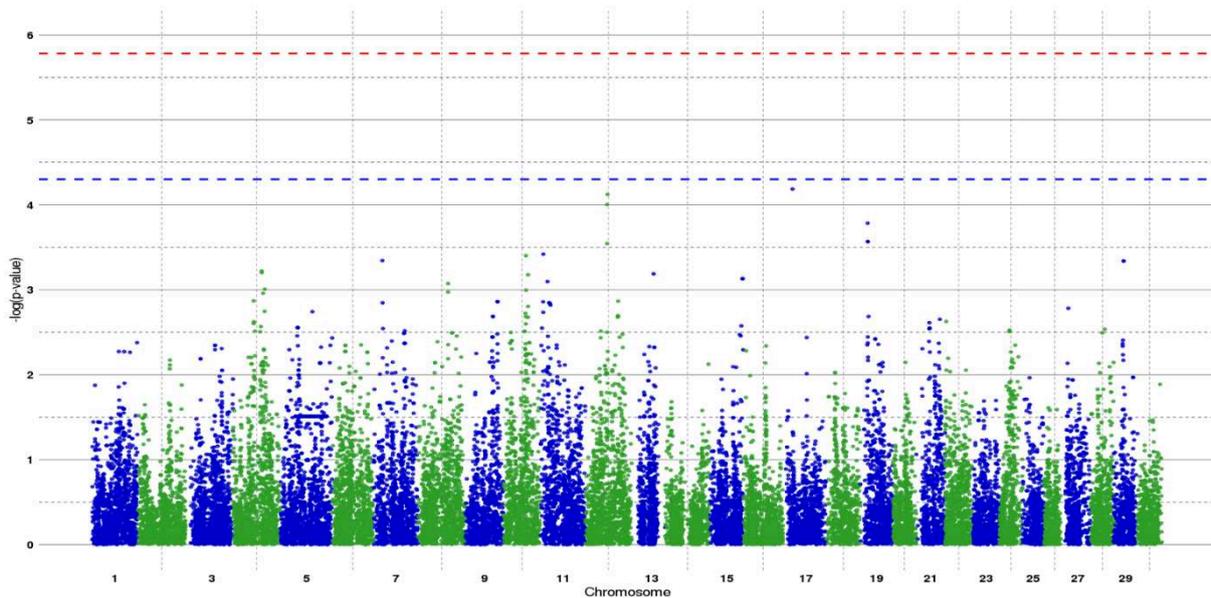
Les maladies ont un impact très important dans l'économie des entreprises aquacoles et la présence de certaines pathologies peut amener au sacrifice total d'un élevage infecté. Aussi, la résistance aux maladies est l'un des objectifs principaux des programmes de sélection chez

la truite arc-en-ciel, et la résistance à l'IPN est particulièrement importante. En effet, l'ANSES place l'IPN dans les dangers sanitaires les plus importants présents en France pour les poissons d'élevage.

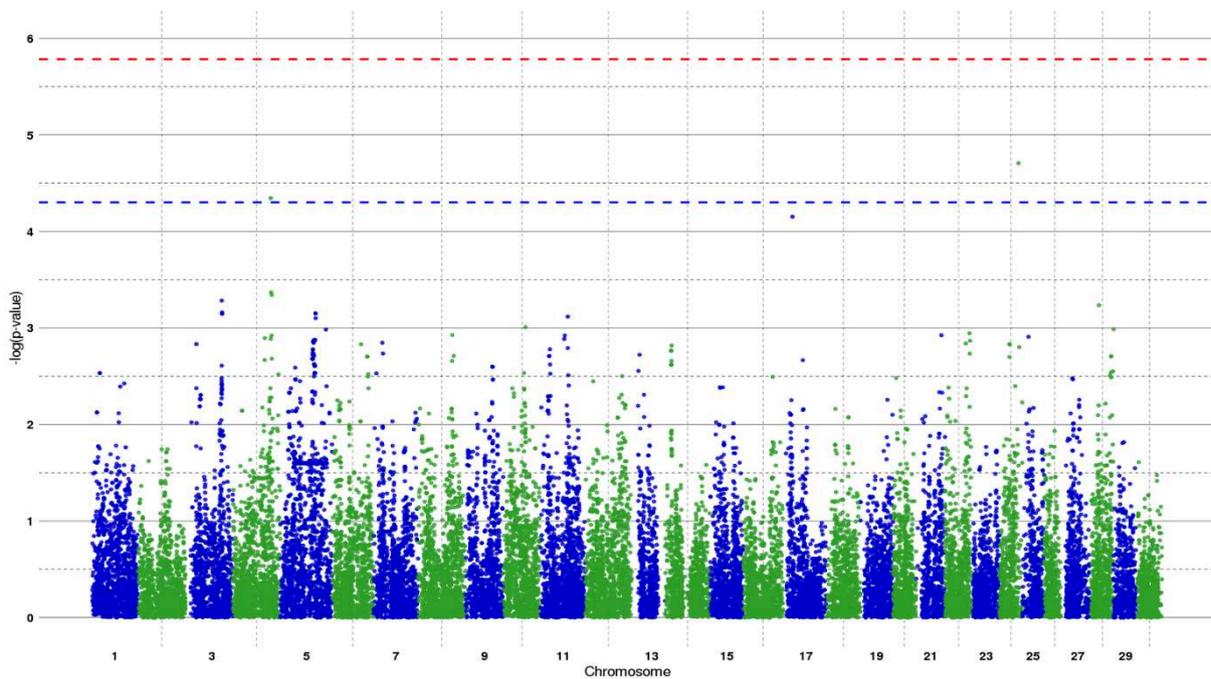
Pour estimer les paramètres génétiques et préciser l'architecture génétique de la résistance à l'IPN ainsi que pour évaluer l'intérêt potentiel de la SG, 1310 truites de la lignée du sélectionneur Bretagne-Truite ont été génotypées avec la puce Axiom™ Trout Genotyping array (57K SNP) et phénotypés pour leur survie à 6 jours et leur durée de vie après infection expérimentale avec le virus de l'IPN. L'estimation des paramètres génétiques et les analyses d'association pan-génomique pour détecter des QTLs ont été réalisées par approches GBLUP et BayesC $\pi$ . L'estimation des précisions des évaluations GBLUP et BLUP sur pedigree a été réalisée par validation croisée en considérant 40 échantillons tirés au hasard d'une population d'apprentissage de 1 000 individus et une population de validation de 310 individus.

L'héritabilité estimée dans notre étude pour la résistance à l'IPN est faible (environ 0,08). La résistance à l'IPN semble très polygénique et aucun QTL à effet fort n'a été détecté. Les quelques QTLs suggestifs identifiés expliquent moins de 0,5% de la variance génétique et ne correspondent à aucun des QTLs décrits dans la littérature. Bien que questions fortes se posent sur l'optimisation du challenge réalisé, le dispositif de notre étude n'apparaît pas assez puissant pour identifier des QTLs d'intérêt pour la résistance à l'IPN en raison non seulement de la faible héritabilité du caractère, mais aussi peut-être d'une couverture insuffisante du génome par les marqueurs de la puce 57K. Une sélection assistée sur marqueurs ne peut donc être efficace, mais une sélection génomique avec 1 000 individus permettrait un gain de précision d'environ 10% par rapport à une sélection sur pedigree.

## 5. Annexes



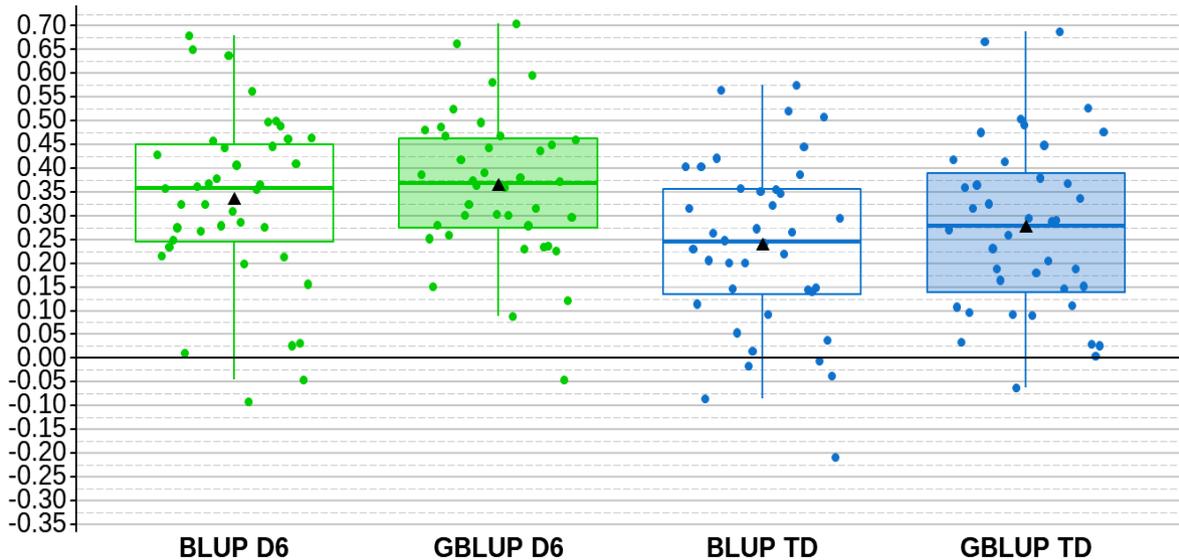
D6



TD

**Annexe - figure 2 :** Manhattan plot des QTLs détectés en analyse GBLUP pour la résistance à l'IPN.

La ligne pointillée rouge correspond au seuil significatif de 5% à l'échelle du génome après correction de Bonferroni ( $-\log(p\text{-value}) > 5,78$ ) et la ligne bleue pointillée correspond au seuil significatif de 5% à l'échelle du chromosome ( $-\log(p\text{-value}) > 4,30$ ). TD : durée de vie post-infection D6 : survie à 6 jours.



**Annexe - figure 3 :** Boîtes à moustaches de la précision de l'évaluation génétique BLUP (boîtes à moustaches transparentes) et GBLUP (boîtes à moustaches pleines) pour la durée de vie post-infection (TD) et la survie à 6 jours (D6) avec une population d'apprentissage de 1 000 individus.

Chaque point représente les 40 simulations et le triangle noir représente la moyenne.

## 6. Références

- Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, et al. Aquaculture genomics, genetics and breeding in the United States: Current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18(1):1–23. <https://doi.org/10.1186/s12864-017-3557-1>
- Aguilar I, Tsuruta S, Legarra A. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. 10th WCGALP. 2014;Vancouver(Canada). <https://doi.org/10.13140/2.1.4801.5045>
- ANSES. AVIS de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à la « hiérarchisation des dangers sanitaires exotiques ou présents en France métropolitaine chez les poissons d'élevage » [internet : <https://www.anses.fr/fr/content/avis-de-lanses-relatif-a-la-hiérarchisation-des-dangers-sanitaires-exotiques-ou-présents-3> ]. 2015.
- Bishop SC, Woolliams JA. Genomics and disease resistance studies in livestock. *Livest Sci*. 2014;166(1):190–8. <https://doi.org/10.1016/j.livsci.2014.04.034>
- Boerner V, Tier B. BESSiE: A software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. *Genet Sel Evol*. 2016;48(1):1–5. <https://doi.org/10.1186/s12711-016-0241-x>
- Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019;47(D1):D801–6. <https://doi.org/10.1093/nar/gky1056>

- Chavanne H, Janssen K, Hofherr J, Contini F, Haffray P, Aquatrace Consortium, et al. A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquac Int.* 2016;24(5):1287–307. <https://doi.org/10.1007/s10499-016-9985-0>
- Chevassus B, Quillet E, Krieg F, Hollebecq M-G, Mambrini M, Fauré A, et al. Enhanced individual selection for selecting fast growing fish: the “PROSPER” method, with application on brown trout (*Salmo trutta fario*). *Genet Sel Evol.* 2004;36(6):643–61. <https://doi.org/10.1051/gse:2004022>
- Dopazo CP. The infectious pancreatic necrosis virus (IPNV) and its virulence determinants: What is known and what should be known. *Pathogens.* 2020;9(2). <https://doi.org/10.3390/pathogens9020094>
- Flores-Mara R, Rodríguez FH, Bangerla R, Lhorente JP, Neira R, Newman S, et al. Resistance against infectious pancreatic necrosis exhibits significant genetic variation and is not genetically correlated with harvest weight in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture.* 2017;479(May):155–60. <https://doi.org/10.1016/j.aquaculture.2017.05.042>
- Fraslin C, Brard-Fudulea S, D’Ambrosio J, Bestin A, Charles M, Haffray P, et al. 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.* 2019;50(3):293–7. <https://doi.org/10.1111/age.12777>
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics.* 2011;12. <https://doi.org/10.1186/1471-2105-12-186>
- Haffray P, Bugeon J, Rivard Q, Quittet B, Puyo S, Allamelou JM, et al. Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture.* 2013;410–411:236–44. <https://doi.org/10.1016/j.aquaculture.2013.06.016>
- Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB, et al. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics.* 2008;178(2):1109–15. <https://doi.org/10.1534/genetics.107.082974>
- Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995;90(430):773–95. <https://doi.org/10.1080/01621459.1995.10476572>
- Li H. A quick method to calculate QTL confidence interval. *J Genet.* 2011;90. <https://doi.org/10.1007/s12041-011-0077-7>
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. Univ Georg Athens, USA. 2018;125.
- Misztal I, Tsuruta S, Strabel T, Druet T, Lee D. BLUPF90 and related programs (BGF90). *Proc 7th World Congr Genet Appl to Livest Prod.* 2002;2.
- Moen T, Baranski M, Sonesson AK, Kjøglum S. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): Population-level associations between markers and trait. *BMC Genomics.* 2009;10:1–14. <https://doi.org/10.1186/1471-2164-10-368>
- Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Ødegård J, et al. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics.* 2015;200(4):1313–26. <https://doi.org/10.1534/genetics.115.175406>
- Ozaki A, Khoo SK, Yoshiura Y, Ototake M, Sakamoto T, Dijkstra JM, et al. Identification of additional quantitative trait loci (QTL) responsible for susceptibility to infectious pancreatic necrosis virus in rainbow trout. *Fish Pathol.* 2007;42(3):131–40.

- <https://doi.org/10.3147/jsfp.42.131>
- Ozaki A, Sakamoto T, Khoo S, Nakamura K, Coimbra MRM, Akutsu T, et al. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol Gen Genet*. 2001;265(1):23–31. <https://doi.org/10.1007/s004380000392>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795>
- Rodríguez FH, Flores-Mara R, Yoshida GM, Barría A, Jedlicki AM, Lhorente JP, et al. 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*. 2019;9(9):2897–904. <https://doi.org/10.1534/g3.119.400463>
- Santi N, Moen T, Ødegard J. US20190241980A1. 2019. p. 107.
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15(1). <https://doi.org/10.1186/1471-2164-15-478>
- Schurink A, Janss LL, Heuven HC. Bayesian Variable Selection to identify QTL affecting a simulated quantitative trait. *BMC Proc*. 2012;6(SUPPL. 2):S8. <https://doi.org/10.1186/1753-6561-6-S2-S8>
- Storset A, Strand C, Wetten M, Kjøglum S, Ramstad A. Response to selection for resistance against infectious pancreatic necrosis in Atlantic salmon (*Salmo salar* L.). *Aquaculture*. 2007;272(SUPPL. 1):62–8. <https://doi.org/10.1016/j.aquaculture.2007.08.011>
- Vallejo RL, Liu S, Gao G, Fragomeni BO, Hernandez AG, Leeds TD, et al. Similar Genetic Architecture with Shared and Unique Quantitative Trait Loci for Bacterial Cold Water Disease Resistance in Two Rainbow Trout Breeding Populations. *Front Genet*. 2017;8(OCT):156. <https://doi.org/10.3389/fgene.2017.00156>
- Wetten M, Kjøglum S, Fjalestad KT, Skjærvik O, Storset A. Genetic variation in resistance to infectious pancreatic necrosis in rainbow trout (*Oncorhynchus mykiss*) after a challenge test. *Aquac Res*. 2011;42(12):1745–51. <https://doi.org/10.1111/j.1365-2109.2010.02771.x>
- Wiens GD, Vallejo RL, Leeds TD, Palti Y, Hadidi S, Liu S. Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: identification of QTL on chromosome Omy19. *PLoS One*. 2013; <https://doi.org/10.1371/journal.pone.0075749>
- Yoshida GM, Carvalheiro R, Rodríguez FH, Lhorente JP, Yáñez JM. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics*. 2019;111(2):127–32. <https://doi.org/10.1016/j.ygeno.2018.01.008>

## Chapitre 6: Discussion générale : Optimisation des schémas de sélection génomique des truites arc-en-ciel

Nous avons montré que grâce à la sélection génomique (SG) des gains de précision de 11% à 37% selon les caractères et la structure de la population de référence sont attendus par rapport à une sélection sur pedigree (Tableau 30). Ce gain de précision devrait donc permettre un accroissement de l'efficacité des investissements en sélection. Ils sont cependant moins importants que ceux décrits dans la littérature (Tableau 1) qui peuvent aller jusqu'à 300% chez les salmonidés. Ceci s'explique au moins pour partie par de plus fortes précisions de la sélection BLUP sur pedigree dans les schémas de sélection français où les familles sont produites selon des schémas de croisement de type semi-factoriel qui permettent l'exploitation d'informations d'un plus grand nombre d'individus plus ou moins apparentés que dans des schémas de croisement de type hiérarchique (Haffray et al. 2018). De plus, dans nos schémas, les familles sont élevées en commun dès le stade oeillé au lieu d'être élevées dans des bassins séparés jusqu'à l'âge du marquage individuel, ce qui limite les biais d'évaluation liés aux effets bassins.

	Caractères de reproduction						Caractères de production					
	SD	FW	SW	EN	EW	ED	BW	GCW	HGCW	ATA	FCA	FCL
puce 6K	16%	20%	23%	22%	22%	27%	7%	22%	26%	27%	13%	11%
puce 10K	16%	20%	28%	24%	24%	29%	10%	25%	28%	28%	14%	12%
puce MD	20%	16%	32%	31%	20%	29%	12%	26%	29%	32%	11%	12%

**Tableau 30 :** Gain de précision par rapport au BLUP d'une sélection intra-cohorte basée sur un GBLUP et des génotypes LD\* et MD pour les caractères de reproduction et production évalués dans une population d'apprentissage d'environ 1 000 individus.

\*LD : puces 6K-05-VD et 10K-05-VD pour les caractères de reproduction et puces 6K-05-AQ et 10K-05-AQ pour les caractères de production.

BW : Poids à la découpe, GCW : poids à la découpe éviscéré, HGCW : poids à la découpe éviscéré et étêté, ATA : surface adipeuse de la darne, FLA : paramètre a\* de couleur de la chair, FCL : clarté de la chair L. SD : semaine de ponte, FW : poids de la femelle pleine, SW : poids de la ponte, EN : nombre d'œufs de la ponte, EW : poids moyen d'un œuf, ED : diamètre moyen d'un œuf.

Nous avons aussi vu que tous les caractères étudiés dans cette thèse sont très polygéniques, ce qui est le cas de la grande majorité des caractères d'intérêt en élevage dans toutes les espèces (Bovin : Bouwman et al. 2018 ; Porc : Qiao et al. 2015 ; Truite arc-en-ciel : Gonzalez-Pena et al. 2016; Neto et al. 2019; Ali et al. 2020). Seuls quelques QTLs expliquant plus de 1% de la variance génétique ont été détectés pour chaque caractère et aucun QTL à effet majeur n'a été identifié. La sélection assistée par marqueurs (SAM) n'est donc pas une voie à privilégier pour

utiliser l'information génomique afin d'accroître l'efficacité de la sélection dans les lignées étudiées car seule la SG, exploitant plusieurs milliers de marqueurs, permet de capter la variance génétique le long du génome.

La mise en œuvre de la SG semble toutefois coûteuse pour de petites et moyennes entreprises de sélection au regard des gains relatifs limités de précision attendus si le coût des outils de génotypage commerciaux n'est pas réduit et la structure du schéma de sélection repensée. À partir des résultats obtenus dans la thèse, nous allons donc maintenant proposer des pistes pour mettre en œuvre une SG qui soit non seulement techniquement efficace, mais aussi économiquement viable.

## 1. Le choix d' outils de génotypage pertinents

### 1.1. Utilisation de puces LD

Tout d'abord, notons que les coûts de génotypage peuvent être significativement réduits par l'utilisation de puces LD ( $\leq 10K$  SNPs). Cette réduction de coût peut être d'autant plus forte que de gros volumes de puces sont achetés par un même consortium et/ou que des puces LD de basse densité ( $\leq 3K$  SNPs) ou à très basse densité ( $\leq 300$  SNPs) peuvent être exploitées. Par ailleurs, rappelons que des génotypages LD de 6 à 10K SNPs permettent d'obtenir des précisions très proches de celles obtenues en génotypage MD et donc des gains de précision par rapport à une évaluation sur pedigree encore utiles : selon les caractères, les gains attendus sont de 7 à 27% pour les puces 6K et de 10 à 29% pour les puces 10K considérées dans nos études (Tableau 30).

Plusieurs pistes sont envisageables pour optimiser le design de ces puces LD. Souvent, une stratégie visant à choisir des marqueurs uniformément distribués le long du génome (en distance physique) et à MAF élevée est considérée dans le choix des marqueurs des puces LD. Il a toutefois été montré qu'une stratégie basée sur le DL pouvait permettre une imputation plus précise (Herry et al. 2018) et que la densification en marqueurs au niveau des points chauds de recombinaison et des télomères améliorerait la précision (Wellmann et al. 2013; Bolormaa et al. 2015). Dans la thèse, nous avons privilégié le fait de limiter le DL local entre marqueurs dans des fenêtres de quelques centaines de kb afin d'éliminer les marqueurs trop redondants, tout en

respectant une répartition relativement homogène le long du génome. Nous avons aussi testé l'intérêt de ne conserver que des marqueurs à MAF moyennes ou fortes ( $MAF > 25\%$ ) avant de pratiquer un tri sur le DL pour améliorer l'informativité des marqueurs retenus. Mais, à nombre de marqueurs retenus constants, ce tri sur la MAF ne s'est généralement pas avéré plus intéressant que de faire un choix plus sévère pour éliminer les marqueurs en fort DL. Toutefois, un tri moins sévère sur la MAF (par exemple à un seuil de 10 ou 15%) aurait peut-être pu se révéler intéressant. Dans la littérature, il a été également proposé de choisir les marqueurs pour leurs effets sur les caractères importants à améliorer (Moser et al. 2010; Vallejo et al. 2017), mais ceci limite l'intérêt de l'outil de génotypage ainsi construit à une seule population et à un nombre restreint de caractères. Ce choix ne peut réellement se justifier que lorsque des QTLs à effets forts (et si possible partagés entre populations) sont identifiés.

Au-delà du choix des marqueurs les plus informatifs pour une population, se pose la question du choix des marqueurs pour une utilisation dans les diverses populations trutticoles. En effet, des outils génériques sont nécessaires pour réduire significativement les coûts du génotypage par des achats en gros volumes d'un même design.

Dans le chapitre 4 de cette thèse, nous avons montré qu'établir une puce générique par simple mise en commun des SNPs partagés entre outils de génotypage dédiés à des populations spécifiques n'a pas permis d'améliorer la précision comparée à l'utilisation d'une puce basée uniquement sur le LD et les MAF d'une population autre que la population d'application. Des travaux supplémentaires sont donc nécessaires pour proposer une stratégie optimale pour bien couvrir le génome en SNPs informatifs pour toutes les populations. Au lieu d'utiliser des puces spécifiques à plus haute densité de marqueurs ( $\gg N$ ) pour définir une puce générique à  $N$  SNPs, une stratégie à tester serait de constituer cette puce générique à  $N$  SNPs par ajout des SNPs de puces spécifiques à moindre densité ( $\ll N$ ). Une autre possibilité serait de choisir les marqueurs avec la MAF la plus importante sur l'ensemble des populations puis de compléter par des SNPs d'intérêt pour chacune des populations en vérifiant qu'une couverture homogène du génome est respectée.

Enfin, grâce à l'imputation des génotypes manquants, des puces LD à très basse densité ( $< 1K$  SNPs) peuvent également être utilisées en complément de puces MD pour réduire les coûts de la SG tout en assurant une efficacité de la sélection quasi-identique à celle d'une SG utilisant seulement des génotypes MD. Ce mode de sélection génomique « allégée » a été initialement proposé en sélection avicole (Wang et al. 2011; Wolc et al. 2016) et porcine (Huang et al. 2012;

Wellmann et al. 2013) en utilisant comme population de référence les parents des générations précédentes (ce qui réduit le coût de génotypage MD par génération à quelques centaines d'individus reproducteurs) et des panels de marqueurs à très basse densité (~400 SNPs) pour limiter le coût de génotypage. Plus récemment, l'intérêt de cette approche a été montré chez le saumon Atlantique (Tsai et al. 2017; Tsairidou et al. 2020) en considérant que les candidats à la sélection et les collatéraux phénotypés sont génotypés à très faible densité (200 à 250 marqueurs). Leurs génotypes LD sont alors imputés en MD grâce aux génotypes MD de leurs parents. Ceci permet de réduire les coûts de génotypages sans perte de précision importante. Dès lors que le nombre de marqueurs du génotypage LD est de 600 SNPs, l'imputation en MD permet d'obtenir des précisions quasi-maximales (Tsairidou et al. 2020). Notons aussi qu'un génotypage des parents avec 5K SNPs permet d'obtenir dans ces travaux des précisions très proches de celles obtenues avec plusieurs dizaines de milliers de marqueurs. Le génotypage pour une très faible densité de marqueurs étant nettement moins cher que le génotypage MD, cette approche d'imputation du génotype MD peut être une solution intéressante pour mettre en œuvre la SG dans les programmes d'élevage piscicole classiques où de grandes familles de frères et sœurs sont disponibles et partagent de longs segments génomiques, ce qui se traduit par une précision de prédiction génomique relativement élevée avec une faible densité de marqueurs. Il convient toutefois de noter que les candidats à la sélection retenus comme reproducteurs doivent être re-génotypés avec la puce MD pour assurer la qualité de l'imputation de la génération suivante.

Par la réduction des coûts de génotypage induite par l'imputation, cette stratégie peut aussi permettre d'augmenter le nombre d'individus génotypés et donc d'améliorer la précision de la SG si ce sont des individus phénotypés ou d'augmenter l'intensité de la sélection si ce sont des candidats à la sélection. Toutefois, la précision de l'imputation étant dépendante de l'existence d'un déséquilibre de liaison suffisamment fort entre marqueurs adjacents et de MAF suffisantes (Huang et al. 2012 ; Wellman et al. 2013), il ne paraît pas forcément évident qu'une même puce très basse densité puisse permettre une imputation précise pour des populations génétiquement diverses. S'il faut développer autant de puces très basse densité que de populations différentes à évaluer dans une espèce, les commandes ne pourront alors pas être mutualisées pour réduire les coûts. Une alternative pourrait consister à combiner le génotypage par séquençage (GBS) sur les candidats et les populations de référence et le génotypage MD des parents. Des travaux analysant d'éventuelles réductions des prix de génotypage associées à des pertes de la précision de la SG seraient donc encore à mener pour préciser les conditions

optimales tant sur le plan technique qu'économique de l'utilisation de puces LD ou de panels de marqueurs ciblés par GBS.

## 1.2. Intérêt du séquençage ou du génotypage à haute densité

Par ailleurs, nous avons aussi vu que la qualité de la couverture du génome de la truite arc-en-ciel était hétérogène non seulement avec les puces LD, mais aussi avec la puce commerciale MD dont seulement 30 à 35 K SNPs sont utiles pour la SG et où de nombreuses zones de plus de 1Mb ne sont absolument pas couvertes. Pour faire évoluer cette puce MD puis développer des puces LD plus performantes, il serait utile de développer et utiliser des séquences et/ou des génotypes à très haute densité ( $HD > 100K$  SNPs) sur les populations d'intérêt en SG. Cela permettrait d'identifier de nouveaux marqueurs informatifs dans de nombreuses populations et localisés dans les régions peu couvertes par les outils de génotypage actuels.

S'il semble évident que des puces HD ne permettront pas de réels gains de précision de la SG par rapport à des puces MD dans les espèces de salmonidés (Ødegård et al. 2014; Tsai et al. 2015b), tout comme cela a aussi été montré chez les bovins avec des puces très haute densité (777K) (Gunia et al. 2014) ou des séquences (Van Binsbergen et al. 2015), on peut toutefois espérer les utiliser pour identifier les polymorphismes causaux. S'il ne semble pas pertinent d'utiliser l'ensemble des variants de la séquence pour accroître la précision de la SG, le choix de variants causaux ou de marqueurs très proches des variants causaux peut en effet améliorer la précision de la SG en la rendant plus robuste au manque de relations étroites entre population de référence et candidats à la sélection (MacLeod et al. 2016), et cela d'autant plus que les MAF des variants causaux sont faibles (Druet et al. 2014). De tels variants peuvent être inclus dans de nouvelles versions de puces LD ou MD permettant une meilleure efficacité de la SG.

## 1.3. Utilisation de la spectroscopie vibrationnelle pour construire la matrice de parenté

Rincet et al (2018) ont proposé de remplacer la SG par la « sélection phénotypique » en montrant chez le blé et le peuplier qu'il est possible de quantifier indirectement la ressemblance

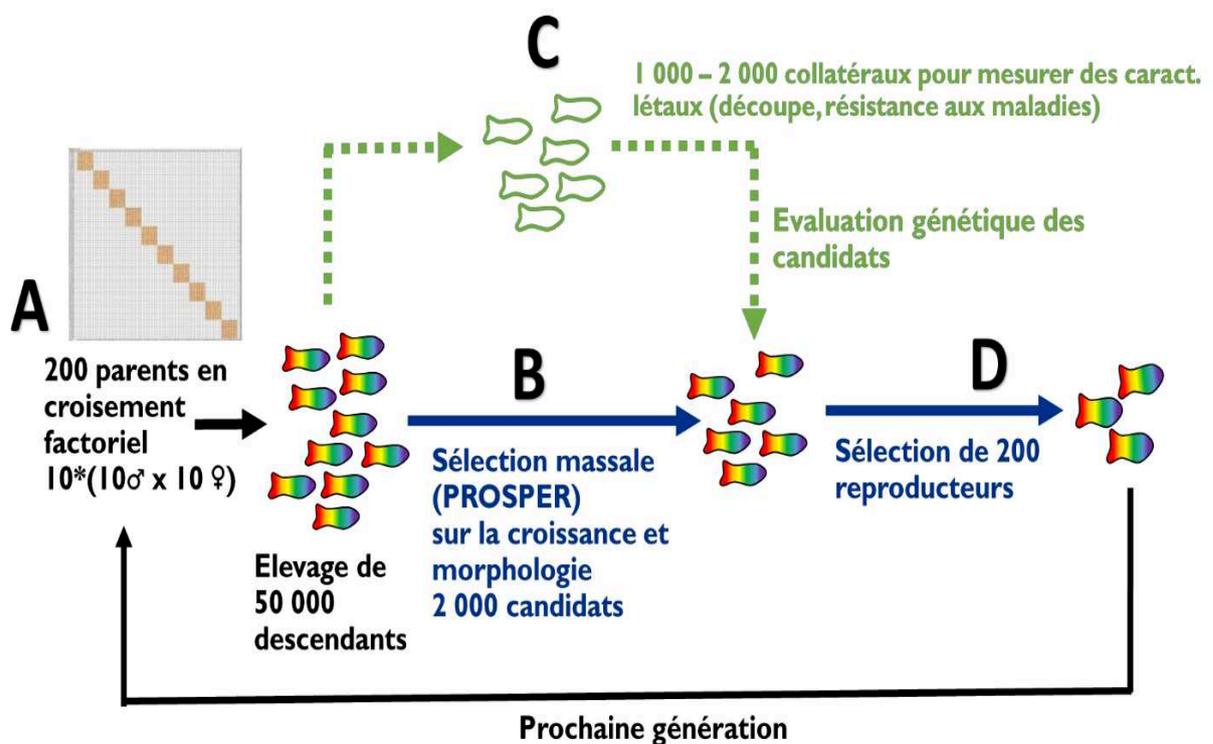
entre individus à partir de données de spectroscopie vibrationnelle et d'utiliser la matrice de ressemblance ainsi construite dans un GLUP à la place de la matrice de parenté génomique. L'avantage premier de cette approche est son moindre coût par rapport au coût du génotypage MD qui reste élevé pour les espèces aquacoles (> 20 € par individu, hors traitement statistique). De façon générale, la spectroscopie vibrationnelle permet de caractériser les bandes d'absorption fondamentales des groupes fonctionnels de substances bio-chimiques constitutives de l'échantillon étudié et donc propres à un individu. Ces bandes d'absorption fondamentales se situent dans la région du moyen-infra rouge (MIR, mid-infrared) alors que les bandes d'absorption issues de leurs combinaisons et de leurs harmoniques se trouvent dans le domaine du proche infra-rouge (NIR, near-infrared) du spectre vibrationnel. Les spectres vibrationnels constituent donc une empreinte chimique d'un individu, empreinte d'autant plus riche et spécifique que l'on se déplace du proche infrarouge vers le moyen infrarouge. Les travaux réalisés chez le blé et chez le peuplier montrent qu'il est possible d'estimer des valeurs génétiques aussi précises en sélection phénotypique qu'en sélection génomique (Rincant *et al.* 2018). Toutefois, le transfert de ce principe aux espèces animales présente plusieurs challenges scientifiques et techniques. En particulier, les résultats ont été acquis à partir de spectres NIR de tissus lignifiés alors que les échantillons biologiques d'espèces aquacoles sont quant à eux très riches en eau. Une saturation de l'information en eau apportée par ces spectres est donc à craindre, ce qui pourrait constituer un biais limitant la puissance de différenciation spectrale entre individus. Une alternative à la spectroscopie NIR serait l'utilisation de la spectrométrie de diffusion Raman, car elle permet plus aisément de s'affranchir de la signature de l'eau présente en spectroscopie NIR et MIR. L'objectif du projet PHENOMIX porté par le SYSAAF et financé par le FEAMP pour 2020-2022 est donc d'évaluer la faisabilité du remplacement de la sélection génomique par la sélection phénotypique en utilisant des technologies spectrales NIR et Raman chez les trois espèces majeures de la filière aquacole française que sont la truite, la daurade et l'huître creuse. Outre le fait que les résultats peuvent dépendre du tissu et de la technologie utilisée, il s'agit aussi de prouver que cette approche peut permettre de quantifier la transmission de ressemblance d'une génération à une autre, point fondamental pour accroître la précision des EBVs, prédire le progrès génétique et assurer le maintien de la variabilité génétique dans les populations aquacoles.

Pour conclure cette première partie de la discussion, je rappelle que l'utilisation de puces LD est pour l'heure la solution à privilégier pour mettre en œuvre une sélection efficace à la fois

techniquement et économiquement dans le cadre de la structure des schémas de sélection trutticole actuelle. Dans la seconde partie de la discussion, je vais proposer des pistes de réflexion visant à modifier les schémas de sélection pour les rendre plus efficaces en SG et/ou en réduire les coûts.

## 2. Mise en oeuvre de la sélection génomique dans le schéma actuel

Rappelons le schéma actuel de sélection appliqué par les entreprises de sélection française (Figure 35), et plus amplement décrit dans le chapitre 1 de la thèse :



**Figure 35** : Description du schéma de sélection actuellement appliqué par les entreprises de sélection trutticole françaises (d'après Haffray et al. 2018).

- A- Plan de croisement en plusieurs petits factoriels pour créer plusieurs centaines à un millier de familles de plein-frères-soeurs
- B- Sélection massive selon la Procédure Optimisée de Sélection par Épreuves Répétées (PROSPER) avec des étapes de tris phénotypiques successifs sur la croissance et la longueur suivis par une sélection massive sur la morphologie externe ou interne (ultrasons) ou sur la teneur en lipides (micro-ondes) des candidats présélectionnés sur la croissance

- C- Évaluation des candidats à la sélection pour des caractères létaux (résistance à une pathologie, caractères de découpe) mesurés sur des collatéraux issus d'un lot dérivé avant le 1<sup>er</sup> tri de sélection massale des candidats
- D- Choix des reproducteurs et des accouplements par une procédure d'optimisation des contributions pour limiter l'augmentation de la consanguinité à 1% par génération

Dans un tel schéma de sélection, la SG permet un gain de précision de 11% à 32% (Tableau 30) par rapport à une évaluation sur pedigree en considérant une population de référence constituée d'environ 1 100 collatéraux des candidats à la sélection (Figure 35-C), mais à un coût plus élevé directement proportionnel au coût du génotypage. Toutefois, nous avons montré que, pour les caractères de reproduction (Chapitre 3), ce coût supplémentaire de génotypage est compensé par le gain de progrès génétique réalisé en seulement 2 à 3 générations de sélection. Nous avons aussi montré qu'il était possible de réduire la taille de la population de référence à 700 individus pour réduire les coûts de la SG tout en gardant une meilleure précision que par une sélection sur pedigree. Les précisions élevées obtenues avec de petites tailles de population de référence s'expliquent par les plans de croisement factoriels, mais aussi par des effectifs efficaces assez faibles dans les populations en sélection, variant de 50 à 70, ainsi que par un DL entre marqueurs maintenu à longue distance (Chapitre 2:). La précision de la SG dépend en effet non seulement de l'apparentement moyen entre les individus, mais aussi du DL entre les marqueurs et les gènes responsables de la variabilité observée (Habier et al. 2011).

Pour accroître la population de référence et ainsi améliorer la précision de la SG, on peut cumuler l'information des collatéraux des générations successives comme nous l'avons montré au chapitre 4. En revanche, génotyper et phénotyper uniquement une génération sur deux n'est pas une voie à recommander pour réduire le coût de mise en oeuvre de la SG, car les pertes de précision associées sont très conséquentes. Il faut envisager de systématiquement phénotyper quelques centaines de collatéraux pour renouveler la population de référence et maintenir la précision de la SG.

Une autre piste est de réduire non pas le nombre d'individus phénotypés, mais le nombre d'individus génotypés de la population de référence, et de maintenir une bonne précision de la SG grâce à une évaluation single-step GBLUP combinant l'information des individus seulement phénotypés à celle des individus phénotypés et génotypés. L'intérêt d'une telle approche mériterait d'être évalué, mais elle nécessite toutefois la connaissance du pedigree de tous les individus phénotypés. Or, dans les programmes de sélection de truite arc-en-ciel mis

## Chapitre 6: La conception d'un nouveau schéma de sélection optimisé pour l'utilisation de l'information génomique

en oeuvre en France, toute la cohorte née de l'ensemble des familles créées à chaque génération est élevée dans le même bassin avant l'identification possible des individus. La connaissance du pedigree a donc lieu *a posteriori* par assignation de parenté qui nécessite un génotypage soit d'une dizaine de marqueurs microsatellites, soit d'une centaine de SNPs (Liu et al. 2016; Griot et al. 2020). Une telle approche pourrait être intéressante si les coûts pour l'assignation deviennent très bas ( $\ll 10$  € / individu) grâce à l'évolution des technologies et à des volumes de marché conséquents, permettant alors d'augmenter la population de référence avec des individus phénotypés et assignés pour augmenter la précision de l'évaluation tout en limitant les coûts de génotypage.

### 3. La conception d'un nouveau schéma de sélection optimisé pour l'utilisation de l'information génomique

Dans cette partie, je vais proposer des pistes de réflexion dont beaucoup devraient être traitées par simulation afin d'évaluer leur intérêt pour optimiser les programmes de sélection à l'heure de la génomique.

#### 3.1. De l'intérêt d'une mutualisation des populations de référence entre sélectionneurs

Avant d'évoquer les éventuelles modifications concernant le programme de sélection d'une seule lignée, abordons une question fréquemment traitée chez les ruminants au démarrage de leurs schémas de SG, à savoir l'intérêt potentiel de cumuler dans une seule évaluation génomique multi-races les informations génomiques et phénotypiques recueillies dans chaque population pour obtenir des prédictions génomiques avec une précision élevée lorsque les populations de référence intra-race sont trop petites. En effet, cette solution pourrait être attractive chez une espèce aquacole où plusieurs lignées sont améliorées pour les mêmes caractères et que seules des petites populations de référence peuvent être constituées par lignée en raison de coûts prohibitifs d'élevage et/ou de phénotypage. Il a été montré en races bovines que l'intérêt de cette approche est souvent très limité, et cela d'autant plus que les candidats ont leurs apparentés proches dans la population de référence (Hozé et al. 2014a). Elle est plus

avantageuse pour les petites populations que pour les grandes populations auxquelles elles ont pu être associées (Hayes et al. 2009; Erbe et al. 2012; Hozé et al. 2014a). Toutefois les gains de précision restent très limités à moins de combiner l'information de populations génétiquement très proches (Brøndum et al. 2011; Hozé et al. 2014b). En effet, il s'agit que suffisamment de QTLs ségrégent dans les diverses populations et que le DL soit bien conservé à travers les populations pour maintenir l'association entre un SNP et un QTL. Sinon l'ajout d'informations de populations distinctes ne permet pas d'améliorer l'estimation des effets des SNPs pour chacune des populations. Ces éléments ne donnent pas d'arguments forts pour envisager que la combinaison des populations de truite arc-en-ciel des différents sélectionneurs français soit *a priori* intéressante pour accroître l'efficacité de la SG. En effet, nous avons montré dans le chapitre 2 que ces populations sont génétiquement différentes et qu'elles ont un pattern de répartition des ROHs le long du génome très hétérogène (D'Ambrosio et al. 2019). Il serait donc très surprenant qu'une population de référence mutualisée permette aux sélectionneurs d'accroître l'efficacité de la SG à coût constant (ou de réduire les coûts de la SG à efficacité constante) pour chacun d'entre eux. De plus, une mutualisation des populations de référence demanderait un phénotypage précis et identique pour l'ensemble des populations. Nous avons montré des pertes nettes de précision dès lors que le phénotypage n'était pas conduit de manière strictement identique d'une cohorte à une autre au sein d'une même population (chapitres 3 et 4).

### 3.2. De l'intérêt d'une revisite du plan d'accouplement

Le plan d'accouplement actuel est un factoriel partiel constitué de très nombreuses familles (600 à 1 000) issus de petits factoriels complets croisant chacun environ 10 mâles et 6 à 10 femelles. L'intérêt des plans factoriels réside non seulement dans une meilleure efficacité de la sélection, mais aussi dans une meilleure préservation de la variabilité génétique dans le cadre d'une sélection massale (Dupont-Nivet et al. 2006) ou sur BLUP (Nirea et al. 2012). Nielsen et al. (2009) montrent que l'efficacité d'une sélection BLUP ou GBLUP sur collatéraux est améliorée quand on considère plutôt 40 que 20 descendants par père, et cela d'autant plus que la densité en marqueurs est importante. Nielsen et al. (2011) montrent également que l'optimisation du plan de sélection en GBLUP versus BLUP pour limiter l'accroissement de consanguinité amène à sélectionner environ deux fois moins de parents en SG qu'en sélection

BLUP considérant l'information des seuls collatéraux ou combinée aux performances propres des candidats dans des plans d'accouplement classique en aquaculture où chaque mâle est accouplé à deux femelles. A l'heure de la mise en route de la SG en aquaculture, il semble donc intéressant de revisiter les plans d'accouplement des schémas de sélection trutticoles. La SG permet, en l'absence de performances individuelles, de capturer l'aléa de méiose grâce aux seuls marqueurs et donc d'avoir une valeur génétique propre à chaque individu d'une même famille en complément de l'évaluation sur pedigree. Il pourrait donc être intéressant dans nos dispositifs en factoriels partiels de considérer un moins grand nombre de familles à nombre total de candidats constant, afin de choisir le meilleur candidat intra-famille et d'ainsi accroître significativement l'intensité de la sélection. Ainsi Sonesson et Ødegård (2016) comparent par simulation l'impact de diverses stratégies de croisement factoriel sur la précision de sélection GBLUP versus BLUP sur pedigree, les taux de consanguinité et les gains génétiques pour des caractères non corrélés, mesurés sur des collatéraux ou directement sur les candidats à la sélection. Ces auteurs comparent des ratios d'accouplement 1:1 et 10:10 en maintenant le nombre de parents constants (100 pères et 100 mères). Ainsi, dans le premier cas, il y aura 30 candidats par couple contre 3 dans le second cas. Ils montrent qu'en SG, changer le ratio d'accouplement de 1:1 à 10:10 a un effet limité sur l'accroissement du gain génétique (<5%) pour tous les caractères. Au contraire, ils montrent qu'en sélection sur BLUP pedigree, le gain génétique sur les caractères mesurés chez les collatéraux augmente de plus de 40% dû à une augmentation forte de l'intensité de sélection des 3 000 candidats sur ces caractères. Pour les deux types d'évaluation, les taux de consanguinité ont diminué de 30% lorsque le taux d'accouplement est passé de 1:1 à 10:10. Il est donc important de conduire les travaux d'optimisation des schémas de sélection génomique en intégrant la gestion de la variabilité génétique dans la réflexion.

### 3.3. De l'intérêt d'augmenter les effectifs de candidats à la sélection et/ou de collatéraux en population de référence

Un facteur limitant de nos dispositifs de sélection est le nombre total de candidats (Figure 35- B) et collatéraux (Figure 35-C) qui peuvent être phénotypés et génotypés. Ce nombre étant supposé fixé pour une lignée donnée, il s'agit de trouver le meilleur compromis entre nombre de candidats à la sélection et nombre de collatéraux en considérant non seulement le nombre de

familles dont ils sont issus, mais aussi le degré d'apparentement entre les deux groupes. À l'heure actuelle, une pré-sélection massale des candidats à la sélection est conduite sur la croissance et la morphologie passant environ de 50 000 descendants produits par le plan d'accouplement à 2 000 candidats génotypés pour l'évaluation BLUP ou GBLUP. Cette sélection entraîne des biais d'estimation des valeurs génétiques qui peuvent être potentiellement forts ainsi qu'un déséquilibre certain dans les familles représentées parmi les candidats retenus à l'issue de cette première étape de sélection. Par ailleurs, le choix des collatéraux à mesurer s'effectuant lors du 1<sup>er</sup> tri de cette sélection PROSPER, il n'est pas évident que tous les candidats aient un nombre de collatéraux constant pour leur évaluation. Les progrès obtenus sur la croissance lors des premières générations de sélection étant déjà conséquents, une alternative possible serait de limiter l'intensité de cette première étape de sélection massale, et de génotyper *a priori* (à très basse densité) le nombre total ciblé de collatéraux et candidats choisis alors quasiment au hasard dans la population des 50 000 descendants. La répartition des individus entre candidats et collatéraux à mesurer s'effectuerait alors pour optimiser, à bénéfice constant, à la fois l'intensité de sélection pour l'ensemble des caractères de l'objectif de sélection et la précision de la sélection en assurant une taille de population de référence et son apparentement avec les candidats à la sélection. A une époque où de nouveaux caractères (résistance aux maladies, efficacité alimentaire...) prennent de l'importance dans les objectifs de sélection, cette revisite de la stratégie de sélection nous paraît utile à envisager. Mais cela nécessite d'en évaluer les impacts précisément par des approches de simulation.

### 3.4. De l'intégration de nouveaux caractères dans les schémas de sélection

La SG offre la possibilité de sélectionner sur de nouveaux caractères difficiles et/ou coûteux à mesurer, mais importants à considérer dans les objectifs de sélection. Elle permet d'obtenir avec le phénotypage d'une population de taille limitée une précision d'évaluation intéressante pour le choix des reproducteurs qui n'était pas atteignable en sélection sur pedigree sans contrôle de plusieurs milliers de collatéraux.

L'un des défis majeurs pour la durabilité des filières aquacoles est l'amélioration des caractères de résistance aux maladies, pour lesquels le potentiel de sélection génétique est important (Gjedrem 2015) et la sélection génomique peut apporter un gain majeur d'efficacité, ces caractères étant difficiles à sélectionner dans les schémas généalogiques (Ødegård et al. 2014).

## Chapitre 6: La conception d'un nouveau schéma de sélection optimisé pour l'utilisation de l'information génomique

Différentes études sur la résistance à différentes maladies montrent une augmentation importante de la précision grâce à la SG (Tableau 1) variant de 10% à 220% quand la précision sur pedigree est très faible. Dédiée au phénotypage de poissons pour la résistance aux maladies, la plateforme de recherche Fortior Genetics a vu le jour en 2017 avec la collaboration du SYSAAF et l'ANSES sur le site de Plouzané (29). Cette plateforme est un outil puissant permettant de mutualiser les coûts, la recherche et l'expertise pour les différentes entreprises de sélection aquacoles. Elle procure aux sélectionneurs français les moyens d'inclure une sélection sur les principaux agents pathogènes dans leurs schémas de sélection. Ces moyens combinés à la mise en œuvre de la SG doivent permettre une sélection efficace pour la résistance aux maladies et ainsi contribuer significativement à l'amélioration de la santé des élevages piscicoles.

Un poisson robuste doit être capable de faire face aux variations de l'environnement, mais doit aussi assurer un niveau minimal de production. Pour évoluer vers un système non seulement économiquement, mais aussi écologiquement efficient, l'efficacité alimentaire est un caractère indicateur de production beaucoup plus pertinent que la classique croissance jusqu'ici plébiscitée en aquaculture. Améliorer l'efficacité alimentaire permet en effet d'améliorer la production tout en valorisant au mieux les intrants (en quantité et qualité, question d'actualité dans le cadre du remplacement des matières premières d'origine marine dans les aliments pour poisson) et en diminuant les rejets (point clé dans un milieu aquatique) (Besson 2017; Besson et al. 2017). Or jusqu'ici, l'efficacité alimentaire n'est pas prise en compte dans les schémas de sélection, car l'ingéré individuel est particulièrement difficile à mesurer chez le poisson qui vit en groupe dans l'eau. Besson et al. (2019) établissent un protocole permettant de mesurer l'efficacité alimentaire individuelle des poissons élevés dans des aquariums individuels. Ils estiment des précisions de la sélection de 0,41 et 0,60 en BLUP et GBLUP respectivement avec seulement 375 individus en population de référence chez le bar. En parallèle de la plateforme de phénotypage des résistances aux maladies, une seconde plateforme dédiée au phénotypage de l'efficacité alimentaire pourrait être un nouvel outil proposé aux sélectionneurs français qui, combiné à la SG, leur procurerait les moyens d'une sélection efficace sur ces composantes majeures de l'efficacité économique et environnementale des systèmes de production que sont la résistance aux maladies et l'efficacité alimentaire.

Pour conclure, si cette thèse a permis de mettre en évidence et de quantifier l'intérêt potentiel de la SG dans le schéma actuel de sélection trutticole français, il s'avère maintenant important d'optimiser les schémas de sélection génomique pour la filière trutticole, et plus largement pour les filières aquacoles. Ces optimisations doivent se faire en revisitant non seulement la structure complète des schémas de sélection actuels et leurs objectifs de sélection pour permettre une meilleure efficacité à court terme des programmes de sélection, mais aussi en préservant leur efficacité à plus long terme. Nous avons montré que la mise en place d'une sélection sur pedigree optimisant les contributions (OCS) des reproducteurs à la génération suivante par le SYSAAF a sans doute permis de limiter l'augmentation de la consanguinité à moins de 1% dans les dernières générations de sélection. Pour être efficace en SG, l'OCS doit être basée sur des relations de parenté établie par ressemblance génomique et non pas sur pedigree (Sonesson et al. 2012). Très récemment, Meuwissen et al. (2020) montrent que la meilleure approche pour cela est de considérer les segments IBD afin d'obtenir le meilleur gain génétique possible sur 20 générations pour les objectifs de maintien de la diversité génétique ciblés. Il s'agit donc maintenant d'intégrer toutes ces nouvelles considérations dans l'optimisation des schémas de sélection aquacoles.

#### 4. Références

- Besson M. Matching breeding goals with farming systems to enhance the sustainability of fish farming. 2017. <https://doi.org/10.18174/395219>
- Besson M, Allal F, Chatain B, Vergnet A, Clota F. Combining Individual Phenotypes of Feed Intake With Genomic Data to Improve Feed Efficiency in Sea Bass. 2019;10:1–14. <https://doi.org/10.3389/fgene.2019.00219>
- Besson M, De Boer IJM, Vandeputte M, Van Arendonk JAM, Quillet E, Komen H, et al. Effect of production quotas on economic and environmental values of growth rate and feed efficiency in sea cage fish farming. PLoS One. 2017;12(3):1–15. <https://doi.org/10.1371/journal.pone.0173131>
- Van Binsbergen R, Calus MPL, Bink MCAM, Van Eeuwijk FA, Schrooten C, Veerkamp RF. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. Genet Sel Evol. 2015;47(1):1–13. <https://doi.org/10.1186/s12711-015-0149-x>
- Bolormaa S, Gore K, Van Der Werf JHJ, Hayes BJ, Daetwyler HD. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Anim Genet. 2015;46(5):544–56. <https://doi.org/10.1111/age.12340>
- Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50(3):362–7. <https://doi.org/10.1038/s41588-018-0056-5>
- Brøndum RF, Rius-Vilarrasa E, Strandén I, Su G, Guldbbrandtsen B, Fikse WF, et al.

- Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *J Dairy Sci.* 2011;94(9):4700–7. <https://doi.org/10.3168/jds.2010-3765>
- D'Ambrosio J, Phocas F, Haffray P, Bestin A, Brard-Fudulea S, Poncet C, et al. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genet Sel Evol.* 2019;51(1). <https://doi.org/10.1186/s12711-019-0468-4>
- Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity (Edinb).* 2014;112(1):39–47. <https://doi.org/10.1038/hdy.2013.13>
- Dupont-Nivet M, Vandeputte M, Haffray P, Chevassus B. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture.* 2006;252(2–4):161–70. <https://doi.org/10.1016/j.aquaculture.2005.07.005>
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci.* 2012;95(7):4114–29. <https://doi.org/10.3168/jds.2011-5019>
- Gjedrem T. Disease resistant fish and shellfish are within reach: A review. *J Mar Sci Eng.* 2015;3(1):146–53. <https://doi.org/10.3390/jmse3010146>
- Griot R, Allal F, Brard-Fudulea S, Morvezzen R, Haffray P, Phocas F, et al. APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Mol Ecol Resour.* 2020;20(2):579–90. <https://doi.org/10.1111/1755-0998.13103>
- Gunia M, Saintilan R, Venot E, Hozé C, Fouilloux MN, Phocas F. Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *J Anim Sci.* 2014;92(8):3258–69. <https://doi.org/10.2527/jas.2013-7478>
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics.* 2011;12. <https://doi.org/10.1186/1471-2105-12-186>
- Haffray P, Enez F, Bugeon J, Chapuis H, Dupont-Nivet M, Chatain B, et al. Accuracy of BLUP breeding values in a factorial mating design with mixed families and marker-based parentage assignment in rainbow trout *Oncorhynchus mykiss*. *Aquaculture.* 2018;490(March):350–4. <https://doi.org/10.1016/j.aquaculture.2018.03.003>
- Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet Sel Evol.* 2009;41(1):1–9. <https://doi.org/10.1186/1297-9686-41-51>
- Herry F, Hérault F, Picard Druet D, Varenne A, Burlot T, Le Roy P, et al. Design of low density SNP chips for genotype imputation in layer chicken. *BMC Genet.* 2018;19(1):1–14. <https://doi.org/10.1186/s12863-018-0695-7>
- Hozé C, Fritz S, Phocas F, Boichard D, Ducrocq V, Croiseau P. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *J Dairy Sci.* 2014a;97(6):3918–29. <https://doi.org/10.3168/jds.2013-7761>
- Hoze C, Fritz S, Phocas F, Boichard D., Ducrocq V., Croiseau P. Genomic Evaluations using Combined reference Populations from Montbéliarde and French Simmental breeds. Proceedings, 10th World Congress of Genetics Applied to Livestock Production 2014b, 17-22 August 2014, Vancouver, BC, Canada
- Huang Y, Hickey JM, Cleveland MA, Maltecca C. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet Sel Evol.* 2012;44(1):1. <https://doi.org/10.1186/1297-9686-44-25>
- Liu S, Palti Y, Gao G, Rexroad CE. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture.* 2016;452:178–82.

- <https://doi.org/10.1016/j.aquaculture.2015.11.001>
- MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*. 2016;17(1):1–21. <https://doi.org/10.1186/s12864-016-2443-6>
- Meuwissen THE, Sonesson AK, Gebregiwi G, Woolliams JA. Management of Genetic Diversity in the Era of Genomics. *Front Genet*. 2020;11(August):1–16. <https://doi.org/10.3389/fgene.2020.00880>
- Moser G, Khatkar MS, Hayes BJ, Raadsma HW. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet Sel Evol*. 2010;42(1):1–15. <https://doi.org/10.1186/1297-9686-42-37>
- Nielsen HM, Sonesson AK, Meuwissen THE. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J Anim Sci*. 2011;99(3):630–8. <https://doi.org/10.2527/jas.2009-2731>
- Nielsen HM, Sonesson AK, Yazdi H, Meuwissen THE. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture*. 2009;289(3–4):259–64. <https://doi.org/10.1016/j.aquaculture.2009.01.027>
- Nirea KG, Sonesson AK, Woolliams JA, Meuwissen TH. Strategies for implementing genomic selection in family-based aquaculture breeding schemes: Double haploid sib test populations. *Genet Sel Evol*. 2012;44(1):1–9. <https://doi.org/10.1186/1297-9686-44-30>
- Ødegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwissen THE. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front Genet*. 2014;5(NOV):1–8. <https://doi.org/10.3389/fgene.2014.00402>
- Qiao R, Gao J, Zhang Z, Li L, Xie X, Fan Y, et al. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. *Genet Sel Evol*. 2015;47(1):1–13. <https://doi.org/10.1186/s12711-015-0089-5>
- Rincent R, Charpentier JP, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, et al. Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3 Genes, Genomes, Genet*. 2018;8(12):3961–72. <https://doi.org/10.1534/g3.118.200760>
- Sonesson AK, Ødegård J. Mating structures for genomic selection breeding programs in aquaculture. *Genet Sel Evol*. 2016;48(1):1–7. <https://doi.org/10.1186/s12711-016-0224-y>
- Sonesson AK, Woolliams JA, Meuwissen THE. Genomic selection requires genomic control of inbreeding. *Genet Sel Evol*. 2012;44(1):1–10. <https://doi.org/10.1186/1297-9686-44-27>
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics*. 2015;16(1):1–9. <https://doi.org/10.1186/s12864-015-2117-9>
- Tsai HY, Matika O, Edwards SMK, Antolín-Sánchez R, Hamilton A, Guy DR, et al. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 Genes, Genomes, Genet*. 2017;7(4):1377–83. <https://doi.org/10.1534/g3.117.040717>
- Tsairidou S, Hamilton A, Robledo D, Bron JE, Houston RD. Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. *G3 Genes, Genomes, Genet*. 2020;10(2):581–90. <https://doi.org/10.1534/g3.119.400800>
- Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, et al. 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*. 2017;49(1):1–13. <https://doi.org/10.1186/s12711-017-0293-6>
- Wang C, Habier D, Garrick DDJ, Leaflet a SR, Fernando RL, Lamont SJ, et al. Accuracy of

- Genomic EBV Using an Evenly Spaced , Low-density SNP Panel in Broiler Chickens. In: 7th European Symposium on Poultry Genetics. 2011. p. 21.
- Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J. Genomic selection using low density marker panels with application to a sire line in pigs. *Genet Sel Evol.* 2013;45(1):1–11. <https://doi.org/10.1186/1297-9686-45-28>
- Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O’Sullivan NP, et al. Implementation of genomic selection in the poultry industry. *Anim Front.* 2016;6(1):23–31.

**Titre :** Intérêt et optimisation de la sélection génomique chez la truite arc-en-ciel

**Mots clés :** aquaculture, diversité génétique, GWAS, reproduction, rendements de découpe, résistance aux maladies

**Résumé :** Les programmes de sélection trutticoles français utilisent depuis 2004 une sélection généalogique avec des pedigrees établis par marquage moléculaire et des performances mesurées sur un à deux milliers de collatéraux des candidats à la sélection. La sélection génomique (SG) s'impose comme une évolution évidente si elle peut être mise en œuvre de manière techniquement et économiquement efficace. Un premier outil de génotypage à moyenne densité (MD) incluant 57 000 marqueurs étant disponible depuis 2015, l'objet de la thèse était d'évaluer l'intérêt de la SG pour trois lignées françaises de truite arc-en-ciel. Un préalable était d'évaluer si la densité de la puce MD était suffisante pour que le déséquilibre de liaison entre marqueurs successifs permette une SG efficace au regard de la diversité génétique des lignées françaises. Les tailles efficaces des lignées ont été estimées à des valeurs de 50 à 70 et le déséquilibre de liaison entre marqueurs successifs à des valeurs ( $r^2 \sim 0,30$ ) compatibles avec une bonne efficacité de la SG. Ces résultats ont été confirmés par des études de validation croisée de la précision de la SG pour divers caractères de reproduction, résistance à la nécrose

pancréatique infectieuse (NPI), croissance, rendement de découpe et qualité de la chair. Avec des héritabilités estimées limitées (NPI) à fortes (poids éviscéré et étêté), tous ces caractères sont très polygéniques : peu de régions génomiques expliquent plus de 1% de la variance génétique et aucune région à effet majeur ( $> 10\%$  de la variance) n'a été identifiée. Les résultats obtenus permettent de conclure à une précision de la SG supérieure de 10 à 37% (selon les caractères étudiés) par rapport à la sélection basée sur le pedigree. Ils ouvrent des pistes pour l'optimisation économique de l'investissement en SG par utilisation de puces à plus basse densité permettant d'obtenir une précision de la SG proche de celle obtenue en MD (perte d'efficacité  $< 5\%$  au-delà de 6 000 marqueurs) et par réduction du nombre de collatéraux phénotypés. En effet, 700 individus collatéraux des candidats à la sélection semblent suffisants pour obtenir des précisions de la SG supérieures à celles d'une sélection sur pedigree. Les entreprises de sélection françaises peuvent donc mettre en œuvre une SG techniquement efficace et rapidement rentable en truite arc-en-ciel.

**Title :** Interest and optimization of genomic selection for rainbow trout

**Keywords :** aquaculture, genetic diversity, GWAS, reproduction, cutting yield, disease resistance

**Abstract:** The French rainbow trout breeding programs are using since 2004 a family selection method based on pedigrees established by molecular markers and performance measured on one to two thousand sibs of the candidates for selection. Genomic selection (GS) is an obvious development for these companies if it can be efficiently implemented. A first medium density (MD) genotyping tool, including 57,000 markers, being available since 2015, the aim of the thesis was to assess the interest of GS for three French lines of rainbow trout. A prerequisite was to assess whether the density of the MD chip was sufficient for the linkage disequilibrium between successive markers to allow effective GS with regard to the genetic diversity of the French lines. The effective sizes of the lines were estimated at values ranging from 50 to 70 and the linkage disequilibrium between successive markers at values ( $r^2 \sim 0.30$ ) compatible with a good GS efficiency. These results were confirmed by cross-validation studies of GS for various

traits of reproduction, growth, resistance to infectious pancreatic necrosis (IPN), cutting yield and quality flesh. With estimated heritability at low (IPN) to high values (eviscerated and headed weight), all these traits are very polygenic: few genomic regions explain more than 1% of the genetic variance and no region with major effect ( $> 10\%$  of the variance) has been identified. The results obtained make it possible to conclude on an accuracy of GS superior of 10 to 30% (according to the studied traits) compared to the selection based on the pedigree. They open up avenues for reflection on an economic optimization of GS by using low density chips to obtain a GS accuracy close to that obtained in MD (loss of efficiency  $< 5\%$  above 6,000 markers) and reduction in the number of phenotyped sibs. Indeed, 700 sibs of the candidates for selection seem sufficient to obtain accuracy of GS superior to the pedigree-based one. French breeding companies can therefore implement a technically efficient and quickly profitable GS in rainbow trout.