



HAL
open science

Développement d'outils et de méthodes de sélection génomique chez le bar et la daurade

Ronan Griot

► **To cite this version:**

Ronan Griot. Développement d'outils et de méthodes de sélection génomique chez le bar et la daurade. Génétique animale. Université Paris-Saclay, 2021. Français. NNT : 2021UPASB010 . tel-03429983

HAL Id: tel-03429983

<https://pastel.hal.science/tel-03429983>

Submitted on 16 Nov 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.

Développement d'outils et de méthodes de sélection génomique chez le bar et la daurade

*Development of tools and methods for genomic selection in the
European seabass and gilthead seabream*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581 Agriculture, Alimentation, Biologie,
Environnement et Santé (ABIES)
Spécialité de doctorat : Génétique Animale
Unité de recherche : Université Paris-Saclay, INRAE, AgroParisTech, GABI,
78350, Jouy-en-Josas
Réfèrent : AgroParisTech

Thèse présentée et soutenue à Paris-Saclay,
le 30/03/2021, par

Ronan GRIOT

Composition du Jury

Etienne VERRIER Professeur, AgroParisTech	Président
Pierre BOUDRY Chercheur, Ifremer	Rapporteur & Examineur
Frédéric FARNIR Professeur, Université de Liège	Rapporteur & Examineur
Pierrick HAFFRAY Chercheur, SYSAAF	Examineur
Pascale LE ROY Directrice de recherche, INRAE - Centre Bretagne-Normandie	Examinatrice

Direction de la thèse

Marc VANDEPUTTE Ingénieur de Recherche, INRAE - Centre IdF-Jouy-en-Josas-Antony	Directeur de thèse
François ALLAL Chercheur, Ifremer	Co-encadrant & Invité
Sophie BRARD-FUDULEA Chercheuse, SYSAAF	Co-encadrante & Invitée
Thierry MORIN Chercheur, ANSES	Invité

REMERCIEMENTS

Ce manuscrit est le fruit de trois ans de travail et de collaboration. Je souhaite donc tous vous remercier d'avoir rendu ce projet possible et cette thèse passionnante.

Mes premiers remerciements vont évidemment à mon directeur de thèse, Marc Vandeputte, pour son implication et son soutien tout au long de ces trois années. Tu auras su me partager ta passion pour le monde de la génétique aquacole, moi qui venais de la génétique des plantes ; et adapter cette thèse pour tirer le meilleur de moi-même tout en comblant mes points faibles. J'espère avoir été à la hauteur car toi, tu l'as été.

Cette thèse n'aurait pas été possible sans le SYSAAF, qui a fourni beaucoup d'efforts tant pour la création du projet que pour sa bonne conduite. Je remercie plus particulièrement Pierrick Haffray, Romain Morvezen et Sophie Brard-Fudulea pour leur implication sans faille dans le suivi de cette thèse et leurs nombreuses idées. Vous avez permis de placer le côté appliqué au cœur de ce projet. Je remercie l'ensemble du SYSAAF pour ces bons moments passés en votre compagnie, qu'ils soient scientifiques ou gastronomiques. Ils m'ont permis de découvrir de nombreux recoins de la France que je ne connaissais pas.

Je remercie l'équipe Geneteam (ou Animalteam) et plus largement l'ensemble de l'Ifremer de Palavas pour m'avoir accueilli. Je souhaite remercier plus particulièrement François Allal, qui a non seulement été un encadrant et un appui scientifique indéniable mais aussi un colloc de bureau, un partenaire de programmation et un exemple de présentation PowerPoint. Je remercie également l'ensemble de l'équipe technique, Frédéric Clota, Marie-Odile Blanc, François Ruelle, Stéphane Lallement et Alain Vergnet pour les moments que l'on a partagés, notamment lors des (nombreuses) biométries qui m'ont permis de participer à un large éventail de projets tous aussi intéressants les uns que les autres.

Je remercie également l'équipe Génétique et Aquaculture de l'INRAE qui, lors des échanges que nous avons eus, a toujours manifesté un grand intérêt dans cette thèse et a soulevé de nombreuses questions qui ont enrichi ce travail. Je souhaite remercier particulièrement Florence Phocas, pour son soutien et son aide précieuse en particulier lors de la période de confinement, et qui a amené une expertise plus que nécessaire.

Je remercie les entreprises Ferme Marine du Douhet et Ecloserie Marine de Gravelines-Ichtus, et spécialement Joseph Brunier, Bruno Peyrou, Sophie Cariou et Jean-Sébastien Bruant pour avoir rendu ce projet possible. De la production des poissons au suivi des résultats, ils se sont impliqués dans ce travail qui, je l'espère, aura un impact sur la sélection française du bar et de la daurade. Merci de m'avoir accueilli sur vos sites et fait découvrir le métier de sélectionneur aquacole.

Je remercie Etienne Verrier et Pierre-Alexandre Gagnaire qui ont suivi l'avancée de ces travaux avec intérêt lors des comités de thèse qui ont ponctué ces trois années.

Je remercie également tous les post-docs et doctorants avec qui j'ai partagé d'agréables moments et discussions : Mathieu, Sara, Charles et Jonathan.

Un merci tout particulier à ma famille qui n'a jamais cessé de me faire confiance et de croire en mes choix.

Enfin, je tiens à remercier celle qui doit me supporter en dehors des horaires de bureau, Marie-Julie. Merci pour ton soutien, ton écoute et surtout ta patience.

Table des matières

CHAPITRE 1 : INTRODUCTION GENERALE	1
1.1. L'essor de l'aquaculture mondiale	2
1.2. L'amélioration génétique comme réponse aux enjeux.....	3
1.3. Le bar et la daurade, clés de l'aquaculture méditerranéenne	4
1.4. Une production sensible aux pathogènes.....	7
1.4.1. La nodaviriose	7
1.4.2. La vibriose.....	8
1.4.3. La pasteurellose.....	8
1.4.4. Vue globale.....	9
1.5. L'apport des outils génomiques en sélection.....	11
1.5.1. La sélection sans génomique.....	11
1.5.2. Le développement de marqueurs génomiques	12
1.5.3. L'utilisation des marqueurs dans l'estimation des valeurs génétiques : la sélection génomique.....	13
1.5.4. L'utilisation des marqueurs dans l'étude de l'architecture génétique : l'identification de QTL	15
1.6. Objectifs de la thèse	16
CHAPITRE 2 : UTILISATION DE MARQUEURS SNPs POUR L'ASSIGNATION DE PARENTE	18
2.1. Développement d'un logiciel d'assignation de parenté.....	20
2.1.1. Résumé	20
2.1.2. Introduction.....	20
2.1.3. Matériel et Méthodes	22
2.1.4. Résultats.....	29
2.1.5. Discussion.....	33
2.2. Discussion complémentaire : Utilisation de l'assignation de parenté dans un contexte aquacole	37
CHAPITRE 3 : ETUDE DE L'ARCHITECTURE GENETIQUE DE LA RESISTANCE A DIFFERENTS PATHOGENES CHEZ LE BAR ET LA DAURADE.....	42
3.1. Détection de QTL de résistance à la nodaviriose dans trois populations de bar européen par l'utilisation d'une puce de 57k SNP.....	44
3.1.1. Résumé	44
3.1.2. Introduction.....	45
3.1.3. Matériel et Méthodes	46
3.1.4. Résultats.....	52
3.1.5. Discussion.....	57

3.2.	Validation du QTL de résistance à la nodavirose chez le bar	60
3.2.1.	Matériel et Méthodes	60
3.2.2.	Résultats	61
3.2.3.	Discussion	64
3.3.	Détection de QTL de résistance à la vibriose chez le bar	65
3.3.1.	Matériel et Méthodes	65
3.3.2.	Résultats	66
3.3.3.	Discussion	68
3.4.	Détection de QTL de résistance à la pasteurellose chez la daurade	68
3.4.1.	Matériel et Méthodes	68
3.4.2.	Résultats	69
3.4.3.	Discussion	70
CHAPITRE 4 : EVALUATION DE L'EFFICACITE DE LA SELECTION GENOMIQUE		73
4.1.	Optimisation de la sélection génomique pour améliorer la résistance aux maladies dans deux espèces de poissons marins, le bar européen (<i>Dicentrarchus labrax</i>) et la daurade royale (<i>Sparus aurata</i>)	75
4.1.1.	Résumé	75
4.1.2.	Introduction	76
4.1.3.	Matériel et Méthodes	77
4.1.4.	Résultats	82
4.1.5.	Discussion	88
CHAPITRE 5 : DISCUSSION GENERALE		91
5.1.	Résumé des principaux résultats	92
5.2.	Applications des résultats dans les programmes de sélection	94
5.2.1.	Optimisation de la sélection génomique	94
5.2.2.	Comment mettre en place la sélection génomique ?	104
5.3.	Mise en perspective	108
5.3.1.	Virose et bactériose	108
5.3.2.	Injection et immersion	109
5.4.	Perspectives	109
5.4.1.	Vers une augmentation du nombre de marqueurs	109
5.4.2.	Vers une sélection assistée par marqueurs ?	112
5.4.3.	Vers un raffinement du phénotype	113
5.5.	Conclusion	115
BIBLIOGRAPHIE		117

CHAPITRE 1 :

INTRODUCTION GENERALE

1.1. L'essor de l'aquaculture mondiale

Il y a 12 000 ans, l'homme commençait à domestiquer les espèces animales et végétales qu'il consommait. Ce n'est que 8 000 ans plus tard, environ 2 000 ans avant J.-C, qu'il débutait la domestication d'espèces aquacoles. Les premières traces d'élevage d'espèces aquacoles ont été trouvées en Chine où les hommes élevaient des carpes communes (*Cyprinus carpio*) dans un but principalement alimentaire vers 2 000 avant J.-C (Nash, 2011). A cette époque, l'aquaculture consistait essentiellement en un nourrissage de poissons capturés. Il a fallu attendre le V^{ème} siècle avant notre ère et le traité de la culture des poissons de Fan Li pour voir apparaître les prémices de l'aquaculture moderne (création des étangs d'élevage, sélection des géniteurs, gestion de la reproduction et maîtrise de l'alimentation et des conditions d'élevage). Cependant, contrairement à l'élevage, l'aquaculture ne s'est pas développée aussi rapidement et la pêche est restée, jusque très récemment, la principale source de produits aquatiques, aussi bien animales que végétales. A partir des années 1990 et malgré un effort de pêche qui s'intensifiait, les captures ont commencé à stagner (Bell et al., 2017) (Figure 1.1). La pêche souffre aujourd'hui d'une mauvaise image auprès des consommateurs, en lien avec la chute des stocks de poissons et son impact sur la biodiversité. Entre 2017, seulement 65,8% des stocks étaient considérés comme exploités de façon durable, contre 90% en 1990 (FAO, 2020).

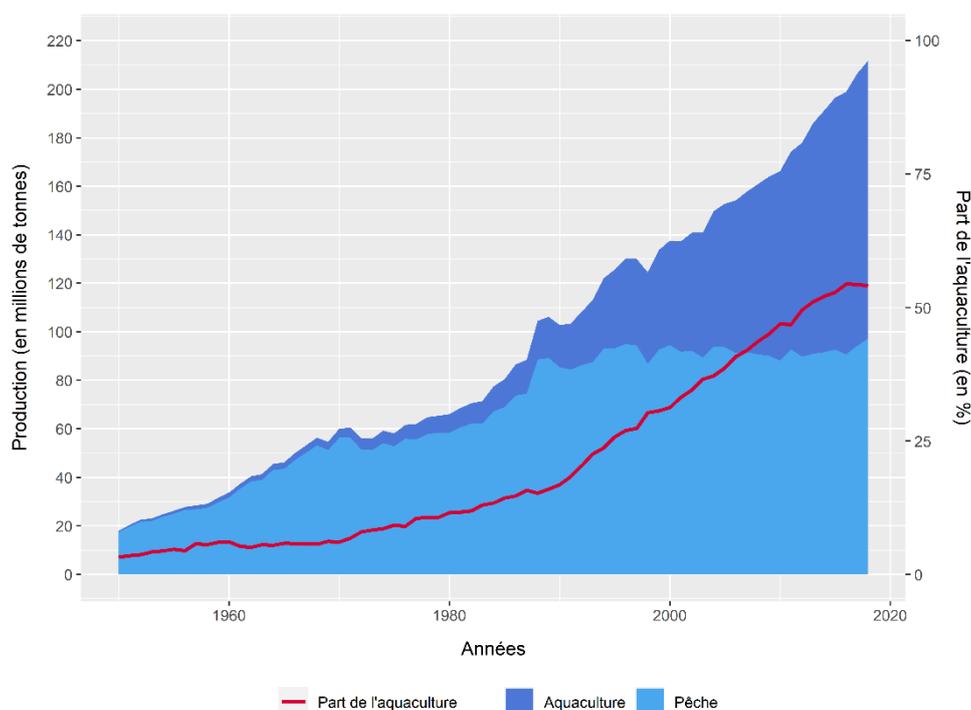


Figure 1.1 : Production de produits aquatiques via l'aquaculture (en bleu foncé) ou la pêche (en bleu clair) dans le monde entre 1950 et 2018. La courbe en rouge représente la part de l'aquaculture dans la production mondiale.

Parallèlement à cette stagnation des captures, la demande mondiale de produits aquatiques n'a cessé de croître, conséquence d'une augmentation de la population, des revenus et de l'urbanisation. La consommation mondiale de poissons est passée de 9,9 kg/an/habitant en 1960 à plus de 20 kg/an/habitant actuellement (FAO, 2020). Dans ce contexte, l'aquaculture s'est rapidement développée depuis les années 1970 pour devenir le secteur agricole à la plus forte croissance à ce jour, avec une croissance de 8,2% par an depuis 1950 (Figure 1.1). En tonnage, l'aquaculture a dépassé la production de viande ovine en 1986 et est aujourd'hui à un niveau comparable à celui de la production bovine (FAO, 1997). Cette croissance n'est cependant pas homogène, tant en localisation qu'en

espèces produites. La majorité de la production mondiale vient d'Asie, en particulier de Chine, qui produit environ 70 millions de tonnes, soit 87% de la production mondiale. C'est également en Chine que la croissance de la production est la plus forte. En Europe, où la croissance de la production est bien plus faible (environ 2,5% par an), la moitié de la production aquacole vient de l'élevage de poissons, qui se concentre autour de 5 espèces majeures : le saumon Atlantique (*Salmo salar*), la truite arc-en-ciel (*Oncorhynchus mykiss*), la carpe (*Cyprinus carpio*), le bar (*Dicentrarchus labrax*) et la daurade royale (*Sparus aurata*) (FAO, 2020). Parmi ces espèces, le saumon a connu une très forte croissance, surpassant de très loin les autres. Seuls le bar et la daurade connaissent aussi une croissance alors que les productions de carpe et de truite arc-en-ciel stagnent, voire diminuent.

Comme toutes les productions agricoles industrialisées, la pisciculture doit faire face à de nombreux enjeux (Ahmed et al., 2019). Le premier est l'utilisation de petits poissons pélagiques issus de captures de pêche pour la production d'huiles et de farines de poissons qui rentrent dans la composition de l'alimentation des animaux omnivores et carnivores. Leur usage pour l'alimentation animale rentre en compétition avec leur consommation par les humains et questionne sur la durabilité des stocks sauvages (Tacon and Metian, 2009). Cette production a augmenté entre 1976 et 1994 pour diminuer depuis (Tacon and Metian, 2009). Cette diminution est principalement due aux innovations dans la composition des aliments destinés aux poissons d'élevage, en particulier la substitution des farines et huiles de poissons par des farines et huiles d'origine végétale. Ainsi, l'alimentation des animaux d'aquaculture n'est plus une menace pour la durabilité de l'aquaculture (Tacon and Metian, 2008). Ensuite, à l'instar des autres productions animales, les questions de santé et de bien-être animal, de sécurité sanitaire et de qualité des produits sont centrales pour la durabilité et le développement de l'aquaculture. Dans le contexte de changement climatique actuel, la hausse des températures, l'acidification des océans ainsi que la diminution du taux d'oxygène dans l'eau sont autant d'enjeux pour l'avenir de l'aquaculture. Ces éléments joueront en particulier sur la survenue et la gravité des épidémies dans les élevages, qu'elles soient favorisées par l'augmentation de la température ou par l'émergence de nouveaux agents pathogènes (Reverter et al., 2020).

1.2. L'amélioration génétique comme réponse aux enjeux

La rapide croissance de la pisciculture a été permise par le perfectionnement des méthodes d'élevage et de la nutrition, ainsi que par la maîtrise de la reproduction et par la sélection génétique. Chez les autres espèces d'élevage, la sélection a permis une augmentation importante de la productivité. Avant même la théorisation de la sélection, les caractères de production comme la croissance, la quantité de lait ou d'œufs ont été les premiers à être sélectionnés. Chez le poulet de chair (*Gallus gallus domesticus*), le taux de croissance a été multiplié par 4,9 entre 1957 et 2001 (Havenstein et al., 2003) tandis que chez le porc (*Sus scrofa domesticus*), entre 1960 et 1996, la croissance a augmenté de 52% (Rauw et al., 1998).

En plus des caractères de productivité, la sélection a permis une amélioration de l'utilisation des ressources, en particulier en ce qui concerne l'alimentation des animaux mais aussi une diminution de l'impact environnemental. Ces deux aspects peuvent être combinés dans l'amélioration de l'efficacité alimentaire (Arthur and Herd, 2005; Besson et al., 2016). Chez le poulet de chair, l'efficacité alimentaire, mesurée par l'indice de conversion, est passée de 2,5 à 1,83 et chez le porc, de 3,24 à 2,26 entre 1960 et 1996 (Rauw et al., 1998).

L'amélioration de la résistance aux agents pathogènes, quant à elle permet à la fois une augmentation de la productivité par une baisse de la mortalité et aussi une diminution de l'utilisation de traitements, que ce soit des antibiotiques ou des vaccins. En aquaculture, la première mise en place documentée d'une sélection génétique a eu lieu en 1919 (Embry and Hayford, 1925). Elle a été réalisée par sélection massale pour la résistance à la furunculose chez l'omble de fontaine (*Salvelinus fontinalis*). Après quatre générations de sélection, la mortalité avait chuté de 67%.

Cependant, la rationalisation de la sélection ne s'est vraiment mise en place que dans les années 1970 et s'est développée d'abord sur le saumon avec la mise en place de la sélection familiale et sur apparentés, bien plus efficace que la sélection massale (Gjedrem, 1985). Elle coïncide avec l'essor de l'aquaculture dans le monde. En 2010, il est estimé qu'environ 8,2% de la production aquacole mondiale provient de cheptels sélectionnés (Gjedrem et al., 2012), contre plus de 80% en Europe en 2016 (Janssen et al., 2017). Ceci montre à la fois une place particulière de l'Europe dans le monde de la sélection aquacole, et une forte dynamique sur son développement.

La sélection génétique peut permettre d'augmenter le chiffre d'affaire des fermes aquacoles par la réduction du temps d'élevage ou bien par la création de nouveaux marchés, comme cela été le cas avec la vente de truite fumée concurrençant le saumon fumé grâce à l'augmentation de la taille des truites arc-en-ciel obtenue par sélection. Elle peut également améliorer la rentabilité des fermes par la réduction des intrants, en particulier l'alimentation ou les produits de traitements, grâce à l'amélioration de l'efficacité alimentaire et de la résistance aux maladies. La sélection peut également jouer un rôle dans l'acceptation sociétale de l'aquaculture, car elle diminue l'impact écologique par la réduction des émissions de gaz à effet de serre, la réduction des rejets comme le phosphate et le nitrate provenant des déjections ou de l'aliment non ingéré (Besson et al., 2016), mais aussi par la réduction de l'usage des antibiotiques.

1.3. Le bar et la daurade, clés de l'aquaculture méditerranéenne

Le bar et la daurade sont les deux espèces emblématiques de l'aquaculture méditerranéenne. Elles représentaient 91% de la production en 2016, 48% provenant de la production de bar et 43% de celle de daurade, avec une production respective de 200 000 tonnes et de 160 000 tonnes en 2018. Ces deux espèces sont majoritairement élevées en Turquie, en Grèce, en Egypte et en Espagne, qui produisent à elles seules environ 90% de la production (FAOSTAT, 2020). Chez ces deux espèces, l'aquaculture représente une large majorité de la production : depuis l'industrialisation de la production dans les années 1990, elle a rapidement pris le pas sur les captures de pêches jusqu'à représenter 97% de la production totale (Figure 1.2).

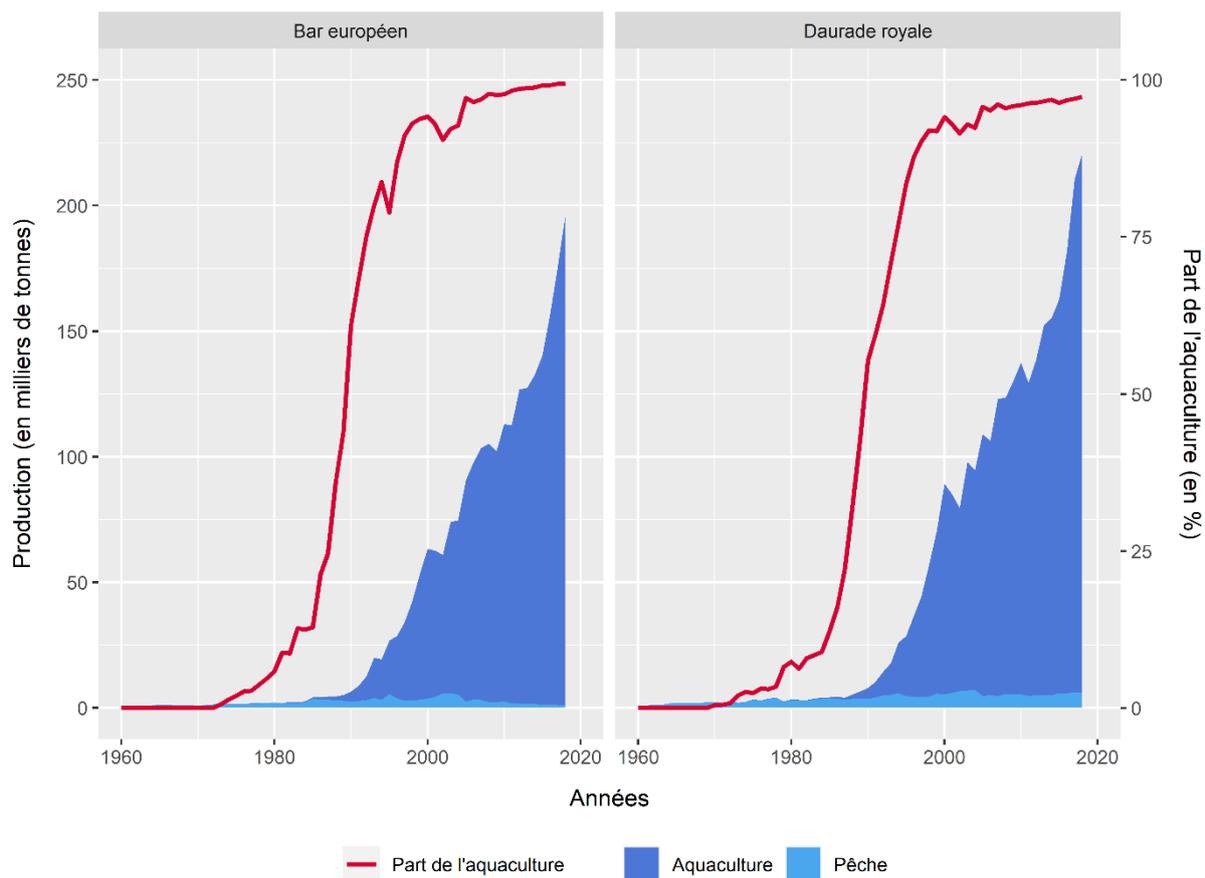


Figure 1.2 : Production de bar européen (panneau de gauche) et daurade royale (panneau de droite) issue de l'aquaculture (en bleu foncé) et de la pêche (en bleu clair) entre 1960 et 2018. La courbe en rouge représente la part de l'aquaculture dans la production.

Dans la nature, ces deux espèces se retrouvent principalement le long des côtes méditerranéennes et atlantiques, mais également en mer du Nord (Figure 1.3). Elles sont euryhalines et peuvent s'adapter aussi bien aux étangs qu'aux estuaires. Leur cycle de reproduction est proche. Le bar se reproduit de décembre à mars en mer Méditerranée et de mars à juin dans l'océan Atlantique (Vandeputte et al., 2019). La daurade se reproduit quant à elle d'octobre à décembre. Elles sont particulièrement prolifiques, produisant environ 200 000 œufs/kg/ponte pour le bar (Vandeputte et al., 2019) et entre 20 000 et 30 000 œufs/kg/jour pour la daurade dont la ponte s'étale sur plusieurs jours (Pousao-Ferreira et al., 1999).

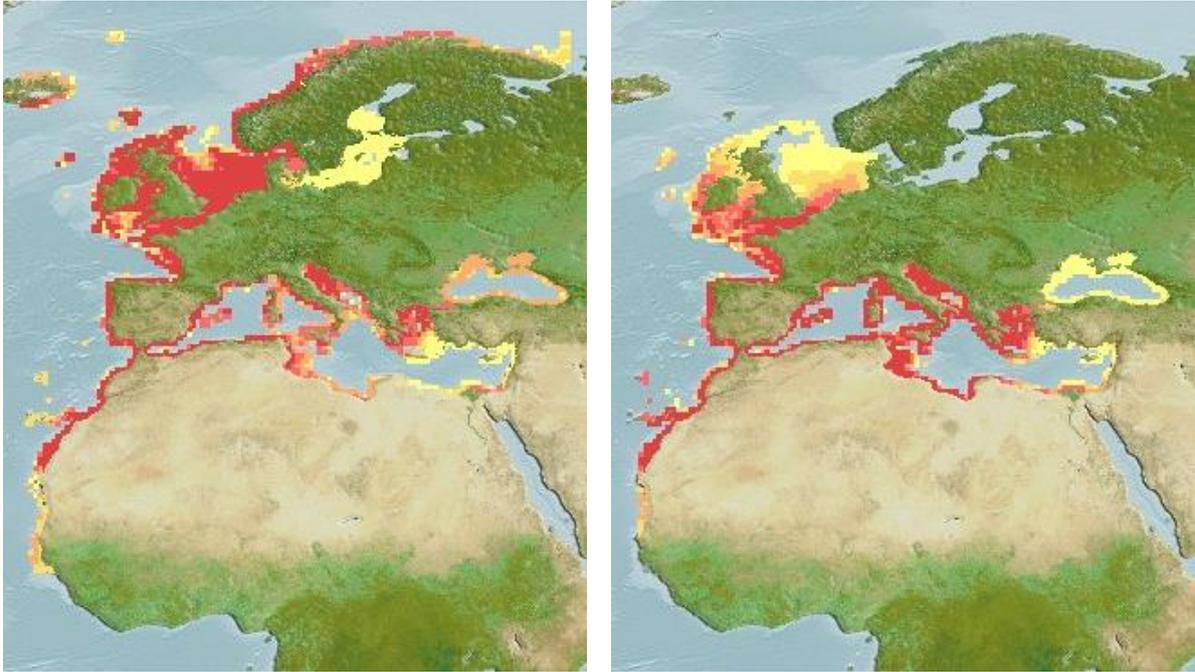


Figure 1.3 : Carte de répartition des populations sauvages de bar européen (panneau de gauche) et de daurade royale (panneau de droite). Source : aquamaps

Comme pour tout élevage, la maîtrise du cycle de vie est un impératif, en particulier celle de la reproduction en captivité, qui permet de choisir des géniteurs, et ainsi de domestiquer les espèces. Chez le bar et la daurade, les élevages ont commencé par la capture d'alevins sauvages dans les années 1970. Les poissons étaient ensuite élevés majoritairement dans des systèmes extensifs jusqu'à taille commerciale puis vendus. Dans les années 1980, chez le bar comme chez la daurade, la reproduction artificielle ainsi que l'élevage des larves et alevins a été maîtrisé, permettant une intensification de la production et le début de l'amélioration génétique par la domestication (Bagni, 2005; Colloca and Cerasi, 2005).

Malgré une part négligeable dans la production de ces deux espèces, la France est un leader dans la sélection génétique et la production d'alevins, qui est exportée à l'étranger à 90%. Deux entreprises réalisent la majeure partie de la production d'alevins en France : le groupe Aqualande par le biais de ses filiales Ferme Marine du Douhet et les Poissons du Soleil et le groupe Gloria Maris avec sa filiale Ecloserie Marine de Gravelines-Ichtus. Ces deux entreprises produisent respectivement 20 millions et 30 millions d'alevins sur les 60 millions produits en France. Elles sont également des leaders au niveau européen. Bien que concurrentes, ces deux entreprises sont adhérentes du SYSAAF (SYndicat des Sélectionneurs Avicoles et Aquacoles Français). Le SYSAAF offre un appui technique à la sélection, en optimisant les plans de croisement selon les objectifs du sélectionneur, en réalisant les estimations de valeurs génétiques sur plusieurs caractères sélectionnés et en favorisant le développement R&D des entreprises par la mutualisation des ressources, en particulier génétiques (performances, pedigree, génotypes).

1.4. Une production sensible aux pathogènes

Comme toute production agricole, l'aquaculture est touchée par de nombreux pathogènes. Ces épidémies sont principalement dues aux fortes densités d'individus dans les systèmes d'élevage (Reno, 1998). D'autres facteurs environnementaux comme la température, la qualité de l'eau et de l'alimentation jouent aussi un rôle.

Le bar et la daurade sont touchés par différentes maladies virales, bactériennes et parasitaires. Parmi ces nombreuses maladies, la nécrose nerveuse virale (NNV), la vibriose et la pasteurellose sont considérées comme ayant le plus d'impact sur la production, autant au stade alevin qu'à taille commerciale (Vendramin et al., 2016).

1.4.1. La nodavirose

La nodavirose est une maladie d'origine virale causée par le virus de la nécrose virale (VNN), appelé aussi virus de l'encéphalopathie et rétinopathie (VER). C'est un virus à ARN appartenant à la famille des betanodavirus qui provoque des nécroses des tissus nerveux (Mori et al., 1992). Il a été rapporté pour la première fois en 1985 chez le poisson perroquet au Japon (Yoshikoshi and Inoue, 1990) et ce n'est qu'en 1988 que les premières épidémies sont apparues en Méditerranée (Breuil et al., 1991). Actuellement, il se retrouve dans toutes les mers du globe, à l'exception de l'Amérique du sud où aucun cas n'a été rapporté à ce jour. Il est capable d'infecter plus de 70 espèces aquatiques, aussi bien d'eau douce que marines. En aquaculture, il touche principalement le bar, la daurade et le turbot en Europe mais aussi le barramundi, l'ombrine et le cabillaud. En Europe, les productions méditerranéennes sont particulièrement affectées car le VNN prolifère dans des eaux dont la température est supérieure à 20°C (Doan et al., 2017b).

Le virus se transmet par l'eau, dans laquelle il peut survivre un long moment sans hôte, mais également par contact d'un individu sain avec un individu porteur, qu'il ait des symptômes ou non. Les symptômes se caractérisent par une nage en rond ou en spirale et une dépigmentation de la peau (Miccoli et al., 2019). Ce comportement erratique est causé par la nécrose des tissus nerveux qui, si elle devient trop importante, provoque la mort de l'individu (Doan et al., 2017b). Dans la majorité des cas, une détection par PCR ou anticorps spécifique ainsi qu'un prélèvement de tissu nerveux pour vérifier la présence de cellules nécrosées permettent de confirmer le diagnostic. En plus de la transmission horizontale, le VNN peut se transmettre par voie verticale, d'un individu à sa descendance par le biais des œufs après infection des gonades (Costa and Thompson, 2016).

Face à ce virus, aucun mode de traitement n'est efficace à ce jour. Seuls des moyens préventifs permettent de réduire la fréquence des épidémies dans les fermes. Les principales méthodes résident dans le fait de ne pas importer le virus à l'intérieur de la structure, en procédant à des contrôles antigéniques ou PCR des poissons provenant de l'extérieur et à une mise en quarantaine. Dans le cas d'œufs, une désinfection par ultra-violet ou ozonation est le moyen le plus efficace (Doan et al., 2017b). A ce jour, un seul vaccin est disponible et offre une protection contre le virus jusqu'à 12 mois (Miccoli et al., 2019).

1.4.2. La vibriose

La vibriose est une maladie causée par les bactéries du genre *Vibrio*. Parmi les 123 espèces décrites appartenant à ce genre, les plus répandues en pisciculture sont *Vibrio parahaemolyticus*, *Vibrio alginolyticus*, *Vibrio harveyi*, *Vibrio owensii* et *Vibrio campbellii*. Principalement répandues dans les estuaires et les zones côtières (Thompson et al., 2004), elles jouent un rôle dans la dégradation de la matière organique (Damir et al., 2013). La première épidémie de vibriose a été rapportée en 1973 chez l'anguille en Norvège et au Royaume-Uni (McCarthy, 1976). Du fait de la grande diversité d'espèces bactériennes provoquant la vibriose, elle est capable de toucher aussi bien les poissons que les mollusques et les crustacés (Ina-Salwany et al., 2019).

La vibriose est une maladie opportuniste, souvent déclenchée par un stress ou lorsque les défenses immunitaires de l'hôte sont amoindries. Elle se développe préférentiellement à une température supérieure à 17°C et se transmet aussi bien par l'eau que par contact. Les individus contaminés présentent une extrême maigreur, une décoloration de la peau et un état léthargique. Dans un second temps, des ulcères et nécroses apparaissent sur la peau. Sans traitement, l'infection se généralise et la septicémie est responsable de la mort. La vibriose est également capable de se transmettre à la descendance en contaminant les œufs (Marco-Noales et al., 2001).

Au sein des fermes piscicoles, une bonne qualité de l'eau, en particulier une faible salinité est une mesure préventive pour éviter les épidémies de vibriose (Prayitno and Latchford, 1995). La réduction du stress, par la réduction des manipulations et de la densité peut également aider dans la prévention. Une dernière mesure préventive est la vaccination du cheptel grâce à des vaccins développés à partir de bactéries inactivées, qui offrent une protection allant de six semaines à une année (Miccoli et al., 2019). Une fois les poissons infectés, un traitement par antibiotique est la solution la plus couramment employée (Ina-Salwany et al., 2019).

1.4.3. La pasteurellose

La pasteurellose, ou photobactériose, est une maladie bactérienne causée par *Photobacterium damsela subsp. piscicida*. Appartenant à la famille des Vibrionacées, la pasteurellose est une forme de vibriose. A cause de sa large distribution géographique, de ses nombreuses antibiorésistances et de sa capacité à infecter différentes espèces marines, elle est vue comme une menace émergente pour la pisciculture méditerranéenne (Andreoni and Magnani, 2014). Elle a été pour la première fois rapportée dans des populations sauvages de perches blanches aux Etats-Unis en 1963 (Snieszko et al., 1964). C'est en 1970 au Japon que les premières espèces d'aquaculture, en l'occurrence des juvéniles de sérioles (*Seriola quinqueradiata*), sont touchées par la pasteurellose (Kubota et al., 1970). Il faudra attendre 1990 pour observer la première épidémie en Europe chez les juvéniles de daurade en Espagne (Toranzo et al., 1991).

Tout comme la vibriose, la pasteurellose se développe préférentiellement à une température supérieure de 15°C avec un optimum aux alentours des 22°C (Romalde, 2002). Les symptômes sont très proches de ceux de la vibriose (dépigmentation de la peau, nécrose - Toranzo et al., 1991).

Les méthodes de prévention contre la vibriose s'appliquent à la pasteurellose. Des vaccins ont été élaborés de la même manière, et consistent en l'injection de bactéries inactivées (Miccoli et al., 2019). Les antibiotiques s'avèrent n'être que très peu efficaces à cause des antibiorésistances de la bactérie, mais restent l'unique méthode de traitement à ce jour.

1.4.4. Vue globale

Bien qu'appartenant à des genres et même des règnes différents, ces pathogènes ont de nombreux points communs. Tout d'abord, ils sont tous capables d'infecter aussi bien le bar que la daurade et causent de fortes mortalités dans les élevages aquacoles (Doan et al., 2017b; Ina-Salwany et al., 2019; Romalde, 2002). Ils sont également plus virulents lorsque la température de l'eau est élevée et sont dits saisonniers puisqu'ils s'observent plus fréquemment pendant l'été. La fréquence d'apparition risque ainsi d'augmenter dans le contexte de changement climatique. Enfin ces pathogènes touchent particulièrement les juvéniles, causant des dégâts importants dans les éclosiers, et sont donc une menace importante pour les sélectionneurs (Doan et al., 2017b; Toranzo et al., 1991). L'ensemble des caractéristiques des agents pathogènes décrits précédemment sont résumés dans le Tableau 1.1.

Tableau 1.1 : Tableau récapitulatif des trois maladies décrites en 1.4.

	NODAVIROSE	VIBRIOSE	PASTEURELLOSE
Organisme pathogène	Virus de l'encéphalopathie et de la rétinopathie (VER) / Virus de la Nécrose Nerveuse (VNN)	<i>Vibrio parahaemolyticus</i> , <i>Vibrio alginolyticus</i> , <i>Vibrio harveyi</i> , <i>Vibrio owensii</i> et <i>Vibrio campbelli</i>	<i>Photobacterium damsela</i> subsp. <i>piscicida</i>
Règne	Virus à ARN	Bactérie gram -	Bactérie gram -
Famille	<i>Nodaviridae</i>	<i>Vibrionaceae</i>	<i>Vibrionaceae</i>
Genre	<i>Betanodavirus</i>	<i>Vibrio</i>	<i>Photobacterium</i>
Distribution géographique	Monde sauf Amérique du sud	Mondiale	Mondiale
Première découverte	Japon en 1985 chez le poisson perroquet	Norvège en 1973 chez l'anguille	USA en 1963 chez la perche blanche et le bar rayé
Symptômes visibles	Nage anormale/erratique (nage en spirale ou en rond), léthargie et assombrissement de la peau	Ulcères et nécroses cutanés, léthargie, maigreur excessive, nage erratique, assombrissement de la peau	Ulcères et nécroses cutanés, maigreur excessive, assombrissement de la peau
Conditions d'apparition	Eau > 15 - 20°C	Eau > 17°C et salinité entre 30 et 35 ppt, stress	Eau > 18 - 20°C
Transmission	<ul style="list-style-type: none"> • Horizontale : contact et eau • Verticale : gonades et œufs 	<ul style="list-style-type: none"> • Horizontale : contact et eau • Verticale : œufs 	<ul style="list-style-type: none"> • Horizontale : contact et eau • Verticale : œufs
Moyens de prévention	Quarantaine, traitement des œufs, contrôle de la qualité de l'eau, vaccination	Quarantaine, contrôle de la qualité de l'eau, vaccination	Quarantaine, contrôle de la qualité de l'eau, vaccination
Moyens de traitement	Aucun	Antibiotiques	Antibiotiques (mais nombreuses antibio-résistances)

1.5. L'apport des outils génomiques en sélection pour améliorer la résistance aux maladies

1.5.1. La sélection sans génomique

L'intérêt pour l'amélioration de la résistance aux maladies par sélection est apparu dès la mise en place de programmes de sélection en aquaculture (Embry and Hayford, 1925). A l'origine, c'est une sélection massale qui a été appliquée pour sélectionner les individus les plus résistants. Avec la sélection massale, les candidats à la sélection sont infectés par un pathogène, volontairement ou non. Les survivants, qui sont qualifiés de résistants, poursuivent les étapes de sélection avant de devenir des géniteurs et de produire la génération future. La sélection massale, bien que très simple à mettre en place, offre de nombreux inconvénients. Premièrement, les candidats à la sélection sont directement infectés par l'agent pathogène, ce qui peut provoquer des épidémies qui se propagent à l'ensemble l'écloserie, aussi bien à cause d'une transmission horizontale entre les différents cheptels présents, mais aussi à cause d'une transmission verticale qui se propagera dans la descendance. Ensuite, la pression de sélection ne peut pas être contrôlée puisqu'elle est égale au taux de survie, qui dépend donc de la sévérité de l'épisode de mortalité.

La sélection sur apparentés est l'alternative la plus couramment employée. Dans ce schéma, les candidats à la sélection sont séparés en deux groupes (Ødegård et al., 2011). L'un des groupes sera volontairement infecté par le pathogène pour lequel le sélectionneur souhaite améliorer la résistance de son cheptel. Cette contamination volontaire en milieu contrôlé est appelée challenge. Lors du challenge, la mortalité est enregistrée tous les jours. A la fin du challenge, les candidats à la sélection qui n'ont pas été challengés peuvent être indexés. Pour cela, leurs valeurs génétiques sont estimées grâce à un modèle BLUP (Best Linear Unbiased Prediction) à partir des phénotypes enregistrés dans le groupe challengé et des relations de parenté entre les individus des deux groupes qui sont enregistrées dans un pedigree. Ainsi, les candidats à la sélection, qui n'ont jamais été infectés par le pathogène, ont une valeur génétique qui retranscrit leur aptitude à résister à ce pathogène, ce qui permet de les classer et de les sélectionner pour produire de futurs géniteurs (Ødegård et al., 2011). La sélection sur apparentés limite le risque d'épidémie au sein de l'écloserie. Elle permet également de contrôler la pression de sélection, et donc de maîtriser l'évolution du gain génétique et de la consanguinité dans les générations futures. Cependant, la connaissance exacte du pedigree est une condition nécessaire à l'obtention de valeurs génétiques par cette méthode.

Pour cela, l'information familiale doit être conservée tout au long de la vie de l'individu. Chez les espèces d'élevage terrestre, cette information est gardée grâce à l'identification de l'individu à sa naissance par un système de marquage physique individuel. Grâce à la traçabilité de l'origine du père et de la mère, le pedigree d'un individu peut facilement être connu. Dans l'élevage de poissons, l'identification individuelle se fait généralement par l'injection d'une puce RFID. Or, à la naissance, il est impossible d'injecter de telles puces car les larves sont trop petites. En attendant une taille suffisante pour les identifier, il faut tout de même conserver l'information familiale. Pour cela, la première méthode a été d'élever chaque famille séparément. Cette méthode a l'inconvénient majeur de produire une confusion entre l'effet de l'environnement d'élevage (le bac) et l'effet de la famille (Vandeputte et Haffray, 2014). Le moyen d'éviter cette confusion est de mélanger plusieurs familles dans un même bac, mais alors l'information familiale est perdue. Cependant, grâce aux marqueurs génétiques, elle peut maintenant être retracée *a posteriori* (Vandeputte et Haffray, 2014). L'apport de l'assignation de parenté par marqueurs génétiques a été considérable dans les programmes de sélection aquacole. En plus de la confusion de l'effet de l'environnement d'élevage avec l'effet génétique, l'élevage en familles séparées ajoute une contrainte sur le nombre de familles à élever, qui

est limité par le nombre de bassins disponibles, et nécessite des infrastructures conséquentes (Vandeputte et al., 2009). Ainsi, l'assignation de parenté *a posteriori* par marqueurs génétiques a permis de lever ces contraintes, permettant un élevage des larves en familles mélangées et une augmentation du nombre de familles. Elle a donc été largement adoptée par les sélectionneurs en aquaculture.

1.5.2. Le développement de marqueurs génétiques

Les premiers marqueurs génétiques à être utilisés en aquaculture ont été des allozymes dans les années 1970 (Brody et al., 1981). Puis dans les années 1990, avec le développement des marqueurs microsatellites, l'utilisation des marqueurs dans les schémas de sélection s'est répandue (Herbinger et al., 1995). En effet, les microsatellites sont bien plus informatifs (ils présentent de nombreux allèles, jusqu'à plusieurs dizaines) et plus fréquents que les allozymes puisque plus nombreux dans le génome. Par leur efficacité en assignation de parenté, permettant de retracer le pedigree d'individus issus de croisement complexes, les marqueurs microsatellites ont permis la mise en place de l'élevage en familles mélangées (Marc Vandeputte and Haffray, 2014). Ensuite, ce sont les marqueurs SNPs (Single Nucleotide Polymorphism) qui ont été utilisés. Bien que ne présentant que deux allèles, ils sont encore plus nombreux que les marqueurs microsatellites et aussi uniformément repartis sur le génome. Chez la plupart des espèces, des millions de marqueurs SNPs peuvent être identifiés.

Dans les années 2000 avec la publication du génome humain (International Human Genome Sequencing Consortium, 2004), la possibilité du séquençage complet de l'ADN a été démontrée. Dans les années qui suivirent, les technologies n'ont cessé de s'améliorer et le coût du séquençage n'a cessé de baisser. De nos jours, le séquençage du génome d'une espèce ne coûte que quelques milliers d'euros contre cent millions de dollars pour celui du premier séquençage de l'humain. Grâce à l'essor des techniques de séquençage haut débit, le développement de marqueurs génomiques chez toutes les espèces, y compris aquacoles, s'est trouvé grandement facilité.

La séquence du génome du bar a été publiée en 2014 (Tine et al., 2014) et celle du génome de la daurade en 2018 (Pauletto et al., 2018). Grâce aux efforts fournis et à la qualité de l'assemblage, ces deux génomes font figures de référence. A partir de ces dates, l'identification de marqueurs génomiques a été beaucoup plus aisée. Les marqueurs SNPs peuvent être génotypés de nombreuses façons. Parmi elles, le RAD-seq fait partie d'une famille de méthodes appelée « génotypage par séquençage », et consiste en un séquençage de certaines régions du génome obtenues après digestion de la molécule d'ADN par des enzymes de restriction (Miller et al., 2007). Une fois la séquence des dites régions obtenue, elle est comparée à celle d'autres individus afin de détecter les SNPs. Le RAD-seq permet dans le même temps d'identifier et de génotyper des marqueurs SNPs, et cet avantage lui a valu d'être très utilisé chez des espèces avec peu ou pas de ressources génomiques disponibles. Un de ses inconvénients est que les marqueurs sont plus ou moins spécifiques de chaque expérimentation, et que de fait leurs répétabilité est faible entre expérimentations (Robledo et al., 2018).

Une seconde méthode est l'utilisation d'une puce à SNPs. Bien que des marqueurs issus de RAD-seq puissent être compilés sur une puce à SNPs, l'obtention des marqueurs se fait généralement autrement. Tout d'abord, une population de référence est séquencée. Cette population doit être suffisamment diversifiée pour refléter la diversité existante dans les autres populations qui seront génotypées avec la puce (population d'un institut de recherche ou d'entreprise, d'un consortium regroupant plusieurs entreprises et/ou instituts, espèce entière). Une fois les séquences obtenues, elles sont alignées sur le génome de référence puis les SNPs sont détectés. Parmi les millions de

marqueurs identifiés à ce stade, une étape de tri est nécessaire afin de sélectionner les marqueurs à figurer sur la puce. Les critères de choix dépendent de la stratégie de design de la puce à SNPs, mais se basent en général sur la qualité (Phred quality score) et la profondeur du séquençage, sur la variabilité de ces marqueurs dans la population de référence (minor allele frequency), sur la présence d'autres variants dans les régions attenantes proches, ainsi que la composition en bases des séquences flanquantes, positionnées en amont et en aval du polymorphisme (LaFramboise, 2009). Pour chaque marqueur sélectionné, une amorce complémentaire à la séquence flanquante du polymorphisme est générée sur la puce. En routine, la molécule d'ADN d'un individu va s'hybrider aux amorces complémentaires correspondantes et une amplification spécifique de la base polymorphe suivante permet de révéler le génotype aux différents marqueurs présents sur la puce. Chez le bar et la daurade, les premières puces ont été développées en 2017 pour le bar et 2018 pour la daurade dans le cadre du projet FEAMP Gènesea (Fond Européen pour les Affaires Maritimes et la Pêche). La puce à SNPs du bar est nommée ThermoFisher Axiom™ DlabChip 57K et le procédé de sélection des marqueurs a été décrit dans Griot et al. (2021). Le design de la puce à SNPs chez la daurade, nommée ThermoFisher Axiom™ SaurChip 60K, n'est quant à lui pas encore publiquement disponible. Ces deux puces, d'environ 57 000 et 60 000 marqueurs SNPs respectivement, ont pour objectif d'être utilisées aussi bien par des entreprises de sélection que par des instituts de recherche. Contrairement au RAD-seq qui nécessite une étape d'identification des marqueurs avant l'étape d'obtention des génotypes, la puce à SNPs permet de produire directement les génotypes, l'identification et la sélection des marqueurs ayant été réalisées lors de la phase de développement, ce qui facilite son utilisation en routine (Robledo et al., 2018).

1.5.3. L'utilisation des marqueurs dans l'estimation des valeurs génétiques : la sélection génomique

L'utilisation de l'information de l'ADN en amélioration génétique a été une révolution dans le domaine. En effet, on ne peut imaginer mieux que l'ADN pour différencier génétiquement des individus entre eux et ainsi mieux comprendre les mécanismes génétiques amenant à des performances différentes. C'est en 2001 que l'utilisation de l'information génomique en sélection a été théorisée sous le terme « sélection génomique ». Dans Meuwissen et al. (2001), les auteurs ont montré que l'utilisation d'un grand nombre de marqueurs répartis tout le long du génome augmente la précision de sélection et donc, le gain génétique par génération. Ils ont montré qu'il était possible de prédire la valeur génétique d'un individu en ne connaissant que son génotype, si les associations génotype-phénotype avaient été documentées chez d'autres individus. Chez toutes les espèces où elle a été mise en place, la sélection génomique a démontré un avantage net par rapport aux méthodes de sélection qui étaient utilisées jusque-là (Lin et al., 2014; Meuwissen et al., 2013; Weller et al., 2017). En aquaculture, sa mise en application a été proposée depuis plus de 10 ans (Sonesson et Meuwissen, 2009a) mais son utilisation par des entreprises privées date de 2016 avec la première commercialisation de saumons atlantiques améliorés par sélection génomique par la société Aquagen. En France, la mise en place de la sélection génomique dans les programmes de sélection aquacole est en cours de développement chez la truite arc-en-ciel, le bar, la daurade et l'huitre creuse grâce à divers projets FEAMP (SG-Truite, Gènesea et Quality Huitre) qui réunissent des entreprises privées, le SYSAAF et des instituts de recherche publique (Ifremer, INRAE, CNRS).

La mise en pratique de la sélection génomique se divise en plusieurs étapes (Taylor, 2014). La première est la création d'une population dite de calibration ou d'entraînement. Cette population sera phénotypée et génotypée et permettra de calibrer le modèle de prédiction génomique. Cette

population doit avoir certaines caractéristiques, comme être proche génétiquement des individus qui seront prédits, et être de taille conséquente. Dans le cas de la sélection pour la résistance aux maladies en aquaculture, la population d'entraînement est constituée des individus qui seront challengés contre l'agent pathogène.

Ensuite, à partir de la population d'entraînement, le modèle de prédiction, comme tout modèle statistique, est calibré. En général, un modèle de prédiction génomique va estimer les effets des marqueurs qui ont été génotypés. La somme de ces effets donne la GEBV (Genomic Estimated Breeding Value) qui est la valeur génétique estimée à partir de l'information génomique. Ce type de modèle est appelé SNP-BLUP lorsque les effets des marqueurs suivent une loi normale (Meuwissen et al., 2001; Whittaker et al., 2000) ou fait partie de la famille des modèles bayésiens (BayesA, BayesB, BayesC, ...) lorsque la distribution des effets des marqueurs suit une autre loi (Habier et al., 2011). Un autre modèle, appelé GBLUP, utilise une autre propriété de l'information génomique. Dans le modèle BLUP (Best Linear Unbiased Prediction), les valeurs génétiques des individus sont estimées grâce aux relations d'apparentement entre les individus. Dans sa variante utilisant l'information contenue dans le pedigree, classiquement utilisée en sélection sur apparentés, les valeurs d'apparentements sont calculées à partir du pedigree et sont donc théoriques, identiques pour tous les animaux d'une même famille. Grâce à l'apport de l'information génomique, il est possible de calculer les valeurs d'apparentements réalisées pour chaque individu, qui sont plus précises que celles calculées à partir du pedigree (VanRaden, 2008), et donc d'obtenir des valeurs génétiques plus précises.

La dernière étape est la prédiction des valeurs génétiques des individus appartenant à la population des candidats à la sélection, qui ne sont que génotypés. Grâce à la calibration du modèle réalisée lors de l'étape précédente, les effets des marqueurs (dans le cas du SNP-BLUP ou des modèles bayésiens) ou des relations d'apparentement réalisées (dans le cas du GBLUP) sont estimés. Puis, les valeurs génétiques de la population des candidats à la sélection sont prédites, connaissant leur génotype, et donc soit en appliquant les effets des marqueurs, soit en calculant les relations d'apparentement réalisées entre la population d'entraînement et celle des candidats à la sélection. A partir des GEBV prédites, les candidats peuvent être classés et ensuite poursuivre le processus de sélection pour devenir de futurs géniteurs.

La mise en place de la sélection génomique dans les schémas de sélection aquacole en France aurait l'avantage, en plus de l'amélioration théorique du gain génétique, de ne nécessiter que peu de modifications techniques et logistiques. En effet, comme le génotypage des candidats et de leurs apparentés est déjà mis en pratique pour l'assignation de parenté dans les schémas de sélection sur apparentés, le génotypage sur une puce à SNPs ne représenterait qu'un changement d'outil de génotypage, et ne constitue donc pas une modification en profondeur de la gestion sur le terrain du schéma de sélection. Il impliquerait néanmoins, au niveau du traitement de l'information, une modification des méthodes de contrôle qualité des génotypages et d'évaluation des valeurs génétiques. Compte tenu du coût supérieur du génotypage de dizaine de milliers de marqueurs SNPs par rapport au coût actuel de génotypage d'une dizaine de marqueurs microsatellites pour assignation de parenté, il faut également évaluer le surcroît de performance que permettrait la sélection génomique dans les programmes de sélection en place.

1.5.4. L'utilisation des marqueurs dans l'étude de l'architecture génétique : l'identification de QTL

Les marqueurs génomiques ont permis de mieux comprendre l'architecture génétique des caractères, en particulier celle des résistances aux maladies (Fraslin et al., 2020b; Hocking, 2005; Rothschild et al., 2007). Alors que la génétique quantitative classique, se basant sur la connaissance du pedigree, permettait d'obtenir une estimation de l'héritabilité d'un caractère ainsi que des corrélations génétiques entre différents caractères, la génomique permet d'aller jusqu'à l'identification de régions génomiques impliquées dans la variation du caractère, appelées QTL (Quantitative Trait Locus). Ces QTLs peuvent donc être dénombrés, positionnés et leurs effets sur le caractère peuvent être quantifiés. Quelle que soit la méthode employée, la détection de QTL repose sur l'association statistique entre la fréquence allélique observée pour un marqueur donné et le phénotype. Un niveau de significativité est donné par une p-value ou un LOD score. Le Logarithm of Odds (LOD) est une mesure du degré de liaison génétique entre deux loci au sein d'une famille. Le LOD est plus précisément le logarithme du rapport de vraisemblance entre un modèle localisant le QTL aux abords d'un marqueur particulier et un autre modèle supposant ce gène non lié. Cette valeur est ensuite comparée à un seuil de significativité choisi, souvent de 5%, pour déclarer la présence ou non d'un QTL.

En pratique, pour que des QTLs soient détectés, il faut qu'ils ségrégent dans la population étudiée. A l'origine, ce sont des populations expérimentales qui ont été utilisées pour détecter des QTLs, en particulier des populations F2, de back-cross ou recombinantes (RIL). Dans ce type de croisement, où les individus sont fortement apparentés, les méthodes d'interval mapping ont été largement employées. Elles réalisent, à intervalle régulier le long du génome, un test statistique visant à déterminer la présence d'un QTL à la position testée (Haley et Knott, 1992). Ensuite, avec l'augmentation du nombre de marqueurs, il a été possible de tirer profit du déséquilibre de liaison au sein de population diversifiées, où les relations d'apparentement sont plus distantes. Ainsi, des méthodes de GWAS (Genome-Wide Association Study) ont été développées (Visscher et al., 2012). En GWAS, les fréquences alléliques pour chacun des marqueurs génotypés au sein d'une population sont mis en relation avec le phénotype d'intérêt. En d'autres termes, on teste si la fréquence allélique explique les différences phénotypiques observées dans la population. Ce test donne une valeur de significativité, exprimée en p-value le plus souvent, pour chaque marqueur génotypé. Il existe plusieurs méthodes pour effectuer ces tests. Le plus couramment utilisé est le modèle GBLUP, capable de tester simultanément tous les marqueurs (Aguilar et al., 2019). Les modèles bayésiens sont également largement utilisés car capables de tester des hypothèses d'architectures génétiques différentes (Legarra et al., 2015). Ces méthodes peuvent également être appliquées à des données de séquence.

En pisciculture, il existe deux exemples majeurs de découverte de QTL ayant un effet fort sur la résistance à une maladie (Fraslin et al., 2020b). Le premier est celui de la découverte d'un QTL impliqué dans la résistance à la NPI (Nécrose Pancréatique Infectieuse) chez le saumon Atlantique qui explique 83% de la variance génétique (Moen et al., 2009). En pratique, les candidats à la sélection sont génotypés au marqueur associé au QTL et, selon leur génotype, peuvent être déclarés comme résistants ou sensibles (Moen et al., 2015). La détection de ce QTL a également permis une approche fonctionnelle, allant jusqu'à la découverte de la mutation dans le gène impliqué dans la résistance (Moen et al., 2015). L'utilisation de ce QTL en sélection a permis de diminuer de 75% le nombre d'épidémies dans la production de saumon (Moen et al., 2015). Le deuxième est celui de la détection d'un QTL expliquant 65% de la variance phénotypique, impliqué dans la résistance à la SHV (Septicémie Hémorragique Virale) chez la truite arc-en-ciel, qui n'est pas utilisé en sélection à ce jour (Verrier et al., 2013).

1.6. Objectifs de la thèse

Cette thèse, inscrite dans le cadre du projet Gènesea, financé par le FEAMP (Fonds Européen pour les Affaires Maritimes et la Pêche) a pour objectif de développer des outils et des méthodes visant à la mise en place de la sélection génomique pour l'amélioration de la résistance aux maladies dans les schémas de sélection chez le bar et la daurade.

Dans ce but, nous avons tout d'abord étudié l'utilisation des données de génotypage, produites par les puces à SNPs de moyennes densités développées dans le cadre du projet, dans l'assignation de parenté. Ensuite, nous avons étudié l'architecture génétique de la résistance à la nodavirose et à la vibriose chez le bar et à la pasteurellose chez la daurade, dans l'objectif de détecter des QTLs impliqués dans cette résistance. Enfin, nous avons exploré les effets de deux paramètres majeurs influençant la précision de la sélection dans le contexte de la sélection génomique : la densité de marqueurs et la taille de la population d'entraînement.

Pour réaliser ce travail, nous nous sommes principalement appuyés sur trois cohortes de bar et une cohorte de daurade, issues des croisements de sélectionneurs français : la Ferme Marine du Douhet (FMD) et l'Eclosérie Marine de Gravelines-Ichtus (EMG). Dans le cadre de l'étude de l'architecture génétique de la résistance à la nodavirose chez le bar, une population expérimentale, composée de quatre familles de plein-frères d'origine différente, a également été créée par l'Ifremer. Ces cohortes ont été challengées par différents pathogènes (deux à la nodavirose, une à la vibriose et une à la pasteurellose) au sein de la plateforme Fortior Genetics, issue de la collaboration entre le SYSAAF et l'ANSES. Les données de génotypage ont été obtenues à partir des puces à SNPs développées dans le cadre du projet Gènesea en collaboration avec le SYSAAF, FMD, EMG, l'Ifremer, INRAE et ThermoFisher. Ces puces, nommées DlabChip 57K pour la puce à SNPs du bar et SaurChip 60K pour celle de la daurade comportent respectivement environ 57 000 et 60 000 marqueurs SNPs. Ces puces ont toutes été traitées par la plateforme Gentyane de l'INRAE de Clermont-Ferrand.

Cette thèse se place dans le cadre du SYSAAF, via le financement d'une bourse CIFRE de l'Association Nationale Recherche Technologie (ANRT) (n° 2017/0731). Les expérimentations et les données utilisées proviennent du programme collaboratif Gènesea (n° R FEA 4700 16 FA 100 0005), financé par le FEAMP (Fonds Européen pour les Affaires Maritimes et la Pêche). Dans ce projet, les sélectionneurs FMD et EMG-Ichtus et le SYSAAF collaborent avec l'Ifremer, INRAE et le CNRS pour avancer vers la mise au point d'outils et méthodes de sélection génomique chez le bar et la daurade





CHAPITRE 2 :
UTILISATION DE MARQUEURS
SNPs POUR L'ASSIGNATION DE
PARENTE

Ce chapitre présente le développement et la validation d'une méthode d'assignation de parenté utilisant les données de puces à SNP.

Malgré la grande diversité de logiciels permettant d'assigner des individus à leurs parents en utilisant différentes méthodes (exclusion, maximum de vraisemblance, bayésien), l'absence de génotypes parentaux reste toujours un problème majeur affectant la fiabilité des assignations. Par ailleurs, les méthodes d'assignation utilisant le maximum de vraisemblance, qui sont en théorie les plus efficaces, en particulier avec les SNP qui n'ont que deux allèles, utilisent des simulations pour fixer les seuils d'assignation. Ces simulations nécessitent la plupart du temps de fournir des estimations des taux d'erreurs de génotypage et de la proportion de parents manquants, qui ne sont pas toujours connues des utilisateurs. Nous avons donc développé une méthode simple qui, par l'utilisation des données observées, est capable par elle-même de restreindre le taux de faux-positifs à un niveau souhaité par l'utilisateur. Cette méthode nous a permis de développer un package R (APIS : Auto-adaptive Parentage inference Software) publiquement disponible sur le CRAN (the Comprehensive R Archive Network), principale source de code et de documentation pour le logiciel R. APIS est disponible ici : <https://cran.r-project.org/web/packages/APIS/index.html>

L'explication théorique de la méthode ainsi que sa validation ont été publiées dans la revue Molecular Ecology Resources.

2.1. Développement d'un logiciel d'assignation de parenté

Cette section a été publiée en 2020 dans *Molecular Ecology Resources* 20 (579 – 590) sous le nom :

« APIS : An auto-adaptive parentage inference software that tolerates missing parents. »

Griot R.^{1,2}, Allal F.³, Brard-Fudulea S.¹, Morvezen R.¹, Haffray P.¹, Phocas F.² & Vandeputte M.^{2,3}

¹ SYSAAF, Station LPGP/INRA, Campus de Beaulieu, Rennes, France

² GABI, INRA, AgroParisTech, Université Paris-Saclay, France

³ MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, Palavas-les-Flots, France

DOI : <https://doi.org/10.1111/1755-0998.13103>

2.1.1. Résumé

En assignation de parenté avec des marqueurs génomiques, les principaux problèmes rencontrés sont les erreurs de génotypage ainsi que l'absence de génotypes parentaux à cause de problèmes d'échantillonnage, de traçabilité ou de génotypage. La plupart des logiciels d'assignation de parenté utilisant des approches de maximum de vraisemblance nécessitent des estimations *a priori* des erreurs de génotypage et de la proportion de parents manquants pour mettre en place des règles d'assignation convenables. Dans cet article, nous présentons le package R APIS, qui peut assigner des individus à leurs parents sans la moindre information *a priori* autre que les génotypes des individus à assigner et de leurs parents potentiels. Le seul paramètre à renseigner est la qualité du résultat souhaité, sous la forme d'un taux d'erreur accepté dans le pedigree, défini par l'utilisateur. Les règles d'assignation utilisent les distributions des probabilités moyennes de transmission mendélienne, qui permettent d'estimer la proportion d'individus dont au moins un génotype parental est manquant. APIS a été comparé à d'autres logiciels (CERVUS et VITASSIGN) sur un jeu de SNP provenant d'une cohorte de bar Européen (*Dicentrarchus labrax*). L'erreur de type I (faux positifs) était plus faible avec APIS qu'avec les autres logiciels, surtout lorsque des génotypes parentaux étaient manquants, mais le taux de vrais positifs était également plus faible, sauf lorsque la puissance d'exclusion théorique atteignait 0,99999. En général, APIS produit des assignations qui satisfont le taux d'erreur accepté par l'utilisateur de 1% ou 5%, même lorsqu'il est testé sur un jeu de données simulé présentant des taux d'erreurs de génotypage élevés (1% ou 3%) et jusqu'à 50% de parents manquants. Du fait de l'utilisation des distributions des probabilités moyennes de transmission mendélienne, APIS est plus fiable lorsque le nombre d'individus à assigner est élevé (> 200). Nous avons démontré que APIS est un logiciel facile d'utilisation et fiable pour l'assignation de parenté, même lorsque 50% des génotypes paternels sont manquants.

2.1.2. Introduction

Pedigree information has many applications, from selective breeding (Durel et al., 1998; Misztal et al., 1992) to ecological and evolutionary studies (Foerster et al., 2007; Wilson et al., 2010). In selective breeding of terrestrial livestock, parental information is usually kept through individual tagging at birth. However, there are many situations where pedigree cannot be maintained with such physical tagging, either because reproduction or birth cannot be observed, as is the case in natural populations (Kruuk, 2004), or because offspring size at birth is too small for them to be tagged. This may be the case in natural populations, or in aquaculture, in batch spawning species or when multiple families are mixed at spawning to avoid the confusion of genetic and common environmental effects (M. Vandeputte and

Haffray, 2014). In all these cases, an alternative method is required to identify the parents of a given offspring.

Parentage assignment with genetic markers is a way to recover pedigree *a posteriori*, using genotypes from both parental and offspring generations. Microsatellite markers have been widely used for this purpose (Pemberton et al., 1995) thanks to their high heterozygosity and number of alleles (De Woody, 2005). Depending on their variability in the population studied, highly reliable assignment could be obtained using from less than 10 to in some cases 50 microsatellite markers (Glaubitz et al., 2003).

More recently, the most widely used genetic markers are single nucleotide polymorphism (SNP) markers (Morin et al., 2004; Vignal et al., 2002). Thousands of SNPs are easily generated in any species with high throughput sequencing techniques (Baird et al., 2008). Although most SNPs are only bi-allelic, and are therefore much less variable than microsatellites, they are widely used in parentage assignment (Anderson and Garza, 2005; Glaubitz et al., 2003) because they are abundant throughout the genome, individually cheaper to genotype than microsatellite loci, and have much lower genotyping error rates than microsatellites (Ranade et al., 2001). Due to their low variability, approximately six times more SNPs than microsatellites are required to reach a given parentage assignment efficiency (Glaubitz et al., 2003).

Several methods have been developed to assign parentage using marker genotypes from the offspring and their potential parents (see the review by Jones, Small, Paczolt, & Ratterman, 2010). These methods can be divided into two main categories: exclusion and likelihood. The exclusion method is the simplest. Knowing the Mendelian inheritance rules, the aim is to exclude all potential parent pairs of an offspring except the true one. In practice, all potential parent pairs are tested for Mendelian incompatibilities with the offspring for all available markers. With a sufficient number of markers, only one parent pair (the true one) should remain compatible with the tested offspring. If the set of markers is not informative enough, multiple pairs remain compatible with the offspring, leading to multiple assignments (Jones et al., 2010).

The likelihood method has been developed to manage most of the situations where the exclusion method fails (Chakraborty et al., 1988). In this method, the likelihood of a parent pair to be the true parents of a given offspring is estimated based on the genotypes over all available markers (Marshall et al., 1998). Then, to infer if the most likely parent pair is the true one, likelihood thresholds are estimated. Most software perform simulations in order to define these threshold values (Jones et al., 2010). A set of parents is generated from the parents present in the data set or from allelic frequencies estimated from the parents genotypes. Then, offspring are simulated based on the genotypes of the parents, following Mendelian inheritance rules. The difference in likelihood (Δ) between the most likely and the second most likely parent pair is calculated, and its distribution between simulated individuals assigned to their true parents and simulated individuals assigned to a wrong parent pair is used to define a threshold value, usually at the 95% confidence interval, established from the values of Δ calculated from the simulation process. If an offspring's Δ is above the threshold, the offspring is assigned to its most likely parent pair. If the offspring's Δ is lower than the threshold, the offspring remains unassigned (Boichard et al., 2014; Kalinowski et al., 2007).

Beyond the practical considerations of cost and tissue collection, parentage assignment has to deal with three main technical issues: lack of power of the marker set, genotyping errors and missing parents, i.e. parents with no genotype available (Jones et al., 2010). Regarding the first issue, the theoretical power of a marker set can be estimated for assignment by exclusion as proposed by Vandeputte (2012), combining allelic frequencies in the parental population and the size of the assignment problem to be solved (number of potential male and female parents). If the power of the

marker set is too low, multiple assignments will occur in an exclusion approach. In such cases likelihood is more efficient than exclusion to assign the offspring to their true parents (Anderson and Garza, 2005), as parent pairs are ordered by decreasing likelihood, while there is no way with exclusion to choose among several parent pairs which are all compatible with the offspring. The likelihood threshold value is set to avoid proposing a wrong parent pair. To keep confidence in the results when the power is low, the threshold value should be conservative enough, which may in turn lead to a lower assignment rate. Concerning genotyping errors, both exclusion and likelihood methods manage them by allowing for some Mendelian incompatibilities. With exclusion, a certain number of mismatches is empirically allowed (Vandeputte et al., 2006). In the likelihood approach, the value of the likelihood for a marker with Mendelian incompatibility is required as an input in the simulation process (Marshall et al., 1998) and generally needs to be a small but greater than zero value, as a zero value would bring the total likelihood to zero, even with a single mismatch (Sancristobal and Chevalet, 1997). Regarding the last main issue, parents with missing genotype (subsequently called “missing parents”), if the marker set is informative enough, offspring with missing parents should not be assigned, but wrong assignment (type I error) may occur if it is not the case. In likelihood approaches, an estimate of the proportion of missing parents is required for the simulation process to accurately establish the threshold values (Marshall et al., 1998). However, providing such an estimate requires a good *a priori* knowledge of the population studied, and any shift from the real values can lead to inconsistent and unreliable results (Jones et al., 2010).

We developed a new parentage assignment method, to efficiently assign offspring to their parents while minimizing the positive assignment error rate (the ratio of false positives to all positives) to a predefined level, set by the user. We designed this method with the aims 1) to be efficient even when some parents are missing, 2) to be simple to use with no *a priori* information required on the genotyping error rate or on the number of missing parents, thus adapting to the characteristics of the population and of the marker set. Using the observed distributions of average Mendelian transmission probability (Boichard et al., 2014) in the offspring population to be assigned, we estimate the number of offspring for which at least one parental genotype is missing, and use this information to set robust parentage assignment rules. This method was implemented in a new software (APIS). We compared its performance to that of existing software, CERVUS (Kalinowski et al., 2007) and VITASSIGN (Vandeputte et al., 2006). We also evaluated its sensitivity to the size of the offspring population used to set parentage assignment rules.

2.1.3. Matériel et Méthodes

2.1.3.1. *Le concept de APIS*

For a given offspring and a potential parent pair, APIS calculates, for each marker, the Mendelian transmission probability, defined as the probability to get the genotype of the offspring knowing the genotype of the parents. These probabilities are averaged over all available markers to give an average Mendelian transmission probability, which is a characteristic of the parent pair, relative to a specific offspring (Boichard et al., 2014). This average Mendelian transmission probability is calculated for all potential parent pairs. The Mendelian transmission probability threshold to declare an offspring as assigned is set up directly from the observed distributions of Mendelian transmission probability in the data analyzed, which requires an offspring data set of sufficient size to have meaningful distributions. The proportion of offspring with at least one missing parent is also estimated from the distributions of Mendelian transmission probability. The only input parameter required from the users is the positive assignment error rate ($e = \text{false positives} / \text{all positives}$) they accept among the individuals assigned to

their parents by APIS. This parameter is used to modulate the value of the Mendelian transmission probability threshold. In the present paper, we only used SNP markers, although APIS can handle other codominant markers, including microsatellites.

2.1.3.2. Méthodes statistiques

For each offspring o and each potential parental pair (sire s , dam d), the average Mendelian probability $P_m(o,s,d)$ is calculated as the average over all genotyped loci (n) of the locus-specific (l) Mendelian transmission probability $p_m(o,s,d,l)$ for a given offspring-sire-dam trio :

$$P_m(o, s, d) = \frac{1}{n} \sum_{l=1}^n p_m(o, s, d, l)$$

The locus-specific Mendelian transmission probability $p_m(o,s,d,l)$ is taken from Tableau 2.1 and Tableau 2.2 as proposed in Boichard et al. (2014).

We consider that the offspring population is representative of the reproduction event(s) that have happened between the parents. Thus, when the genotype of a parent is missing at a specific locus, a reasonable estimate of the probability of allele A at this locus is the frequency of A in the population of offspring, i.e. in the effective subsample of the parental population that has participated to the reproduction event(s). This differs from the approach of Boichard et al. (2014), which use allelic frequencies in the parental population. This is reflected in the Mendelian probabilities estimated for unknown parental genotypes in Tableau 2.1 and Tableau 2.2.

For a given offspring, all potential parent pairs are ordered by decreasing $P_m(o, s, d)$ and noted $PPi(o)$, where i is the rank of the Mendelian transmission probability of the parent pair among all possible pairs. The Mendelian transmission probability of $PP1(o)$ is noted $P1_m(o)$, that of $PP2(o)$ is noted $P2_m(o)$ and that of $PP3(o)$ is noted $P3_m(o)$. The difference between successive ordered $Pi_m(o)$ is $\Delta_i(o) = Pi_m(o) - P(i + 1)_m(o)$. As APIS uses the observed distributions of $Pi_m(o)$, we defined $Pi_m(:)$ the set of all $Pi_m(o)$ among all the offspring o in the data set and $\Delta_i(:)$ the set of all $\Delta_i(o)$. The distributions of $P1_m(:)$, $P2_m(:)$ and $P3_m(:)$, shown in Figure 2., contain information about potential missing parents (i.e. parents with missing genotypes at all loci): when the parents of all offspring are present in the data set, the distribution $P1_m(:)$ has a single peak (Figure 2.a), but when some parents are missing, the distribution of $P1_m(:)$ has two peaks, that represent two distributions (Figure 2.b). The first peak (with the highest Mendelian probability value) represents the distribution of the Mendelian probabilities of the best parent pairs for the offspring which have both parents genotyped in the data set, while second peak represents that of the offspring with at least one missing parent. The distribution of $P1_m(:)$ for the offspring with at least one missing parent is indeed similar to the distribution of $P2_m(:)$ for the offspring which have both parents genotyped (Figure 2.b).

The distributions of $P1_m(:)$ and $P2_m(:)$ are also highly sensitive to the power of the marker set. Reducing the number of markers (and thus the assignment power of the marker set), the distributions of $P1_m(:)$ and $P2_m(:)$ tend to overlap (Figure 2.). Therefore, when all parents are present in the data set, the difference between $P1_m(:)$ and $P2_m(:)$, represented by $\Delta_1(:)$, is directly related to the power of the marker set (Figure 2.). When all parents are present in the data set, the higher the power of the marker set, the lesser the overlap of the distributions of $P1_m(:)$ and $P2_m(:)$.

Tableau 2.1 : Mendelian transmission probabilities for one marker when the offspring is homozygous (AA), conditional on the sire and dam genotypes. Adapted from Boichard et al. (2014)

Sire\dam	AA	AC	CC	Missing
AA	1	0.5	0.01	fa
AC	0.5	0.25	0.01	0.5*fa
CC	0.01	0.01	0.01	0.01
Missing	fa	0.5*fa	0.01	fa ²

C = any allele different from A

fa = frequency of allele A in the offspring population analysed

0.01 = arbitrary value for genotyping error

Tableau 2.2 : Mendelian transmission probabilities for one marker when the offspring is heterozygous (AB), conditional on the sire and dam genotypes. Adapted from Boichard et al. (2014)

Sire\dam	AA	AB	AC	BB	BC	CC	Missing
AA	0.01	0.5	0.01	1	0.5	0.01	fb
AB	0.5	0.5	0.25	0.5	0.25	0.01	0.5(fa + fb)
AC	0.01	0.25	0.01	0.5	0.25	0.01	0.5*fb
BB	1	0.5	0.5	0.01	0.01	0.01	fa
BC	0.5	0.25	0.25	0.01	0.01	0.01	0.5*fa
CC	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Missing	fb	0.5(fa + fb)	0.5*fb	fa	0.5*fb	0.01	2*fa*fb

C = any allele that is not A or B

fa = frequency of allele A in the offspring population analyzed

fb = frequency of allele B in the offspring population analyzed

0.01 = arbitrary value for genotyping error

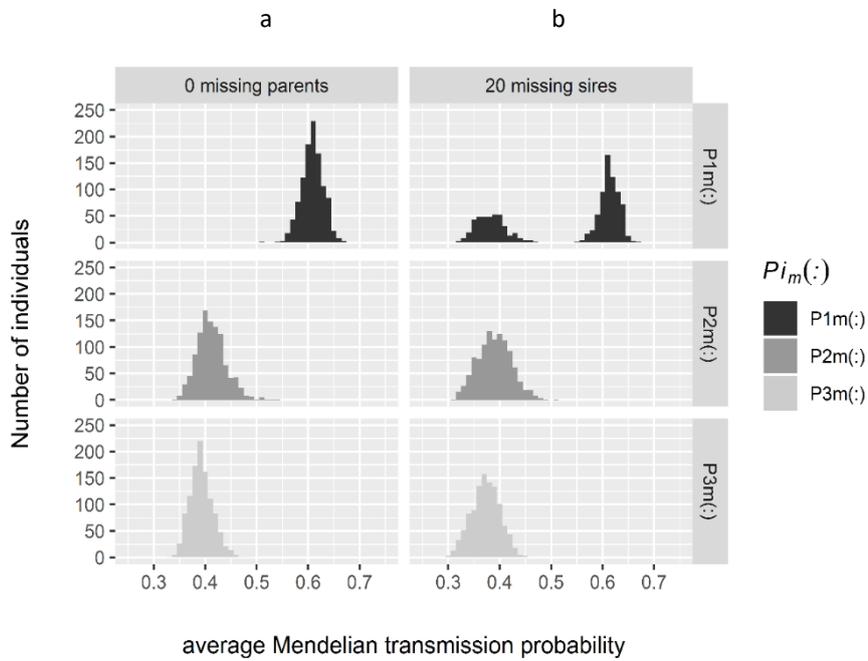


Figure 2.1 : Effect of missing parental genotypes on Mendelian transmission probability distributions in 1,068 offspring of European sea bass from 39 sires and 14 dams genotypes for 200 SNPs. Distributions of Mendelian transmission probabilities of the best parent pairs ($P1_m(\cdot)$, in dark grey), of the second best parent pairs ($P2_m(\cdot)$, in medium grey) and of the third best parent pairs ($P3_m(\cdot)$, in light grey) in the case of (a) no missing parents, and (b) 20 missing sires out of 39

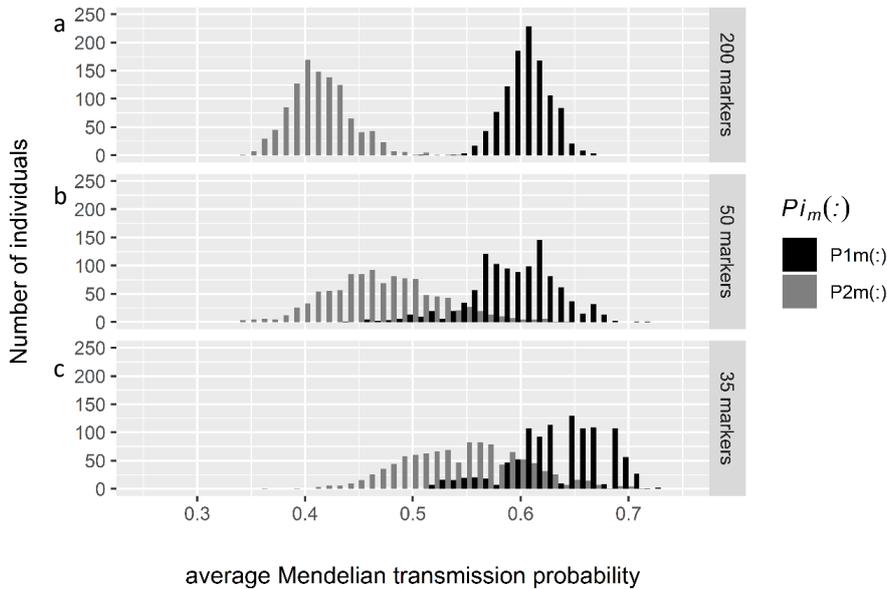


Figure 2.2 : Effect of the number of markers on Mendelian transmission probability distributions in a cohort of 1068 offspring of European sea bass from 39 sires and 14 dams genotyped for 200 (a), 50 (b) or 35 (c) SNPs. The distributions of Mendelian transmission probabilities for the best parent pairs ($P1_m(\cdot)$) are represented in black and those of the second best parent pairs ($P2_m(\cdot)$) are represented in grey.

By construction, $P1_m(o) \geq P2_m(o) \geq P3_m(o) \geq \dots \geq Pn_m(o)$. Parent pair $PP1(o)$, corresponding to $P1_m(o)$, can either be the true parent pair (when both parents are genotyped and the power of the marker set is high enough) or a wrong one (when one or both parents are missing, or the power of the marker set is low). When $PP1(o)$ is the true parent pair, $P1_m(o)$ is much greater than $P2_m(o)$. On the contrary, when $PP1(o)$ is a wrong parent pair, $P1_m(o)$ is only slightly greater than $P2_m(o)$. As a consequence, $\Delta_1(o)$ is high when the true parents are present and low when they are missing (Figure 2.1a and Figure 2.1b, bottom panel). The values of $\Delta_i(o)$ when $i \geq 2$ are always low (Figure 2.1).

The first step to set up the Mendelian transmission probability thresholds is to estimate the number N_{miss} of offspring with at least one missing parent (i.e., a parent missing genotype at all loci). The distribution of $P1_m(\cdot)$ for the progeny with missing parents is expected to be similar to the distribution of $P2_m(\cdot)$ for the progeny with both parents genotyped, and centered on the same median value. Thus, N_{miss} corresponds to the number of individuals in the second peak of the distribution of $P1_m(\cdot)$.

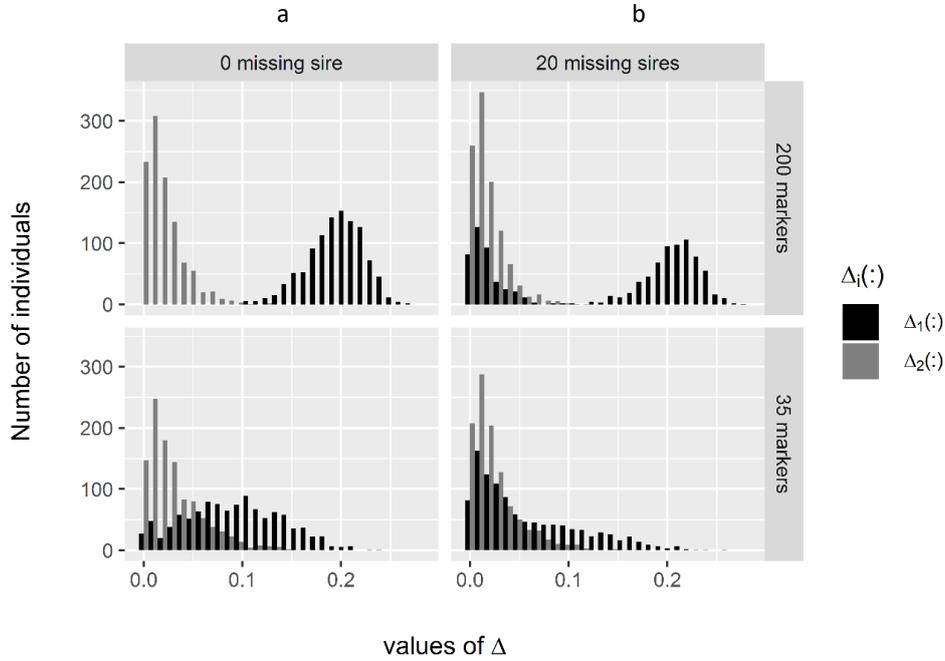


Figure 2.1 : Effect of the number of markers and of missing parental genotypes on Δ distributions in 1068 offspring of European sea bass from 39 sires and 14 dams, with no missing sire (a) or 20 missing sires (b), and with 200 markers (top panels) or 35 markers (bottom panels). $\Delta_1(\cdot)$, in black, is the difference in Mendelian transmission probabilities between the best and second best parent pairs of an offspring, and $\Delta_2(\cdot)$, in grey, is the difference in Mendelian transmission probabilities between the second and third best parent pairs of an offspring.

However, as showed in Figure 2., a lack of power in the marker set also leads to some overlapping of the distributions of $P1_m(\cdot)$ and $P2_m(\cdot)$. Thus, the values of $P1_m(\cdot)$ that are greater than the median of $P2_m(\cdot)$ can either correspond to individuals with at least one missing parent or to individuals with both parents genotyped. Still, we can see on Figure 2. that even with a low power marker set (35 markers), the values of $P1_m(\cdot)$ that are lower than the median of $P2_m(\cdot)$ are very unlikely to be values of $P1_m(\cdot)$ from individuals with both parents genotyped, and thus are mostly from individuals with at least one missing parent. Under the assumption that the distribution of $P1_m(\cdot)$ for the progeny with at least one missing parent is symmetrical, N_{miss} is then estimated as twice the number of offspring for which $P1_m(o)$ is lower than the median of $P2_m(\cdot)$:

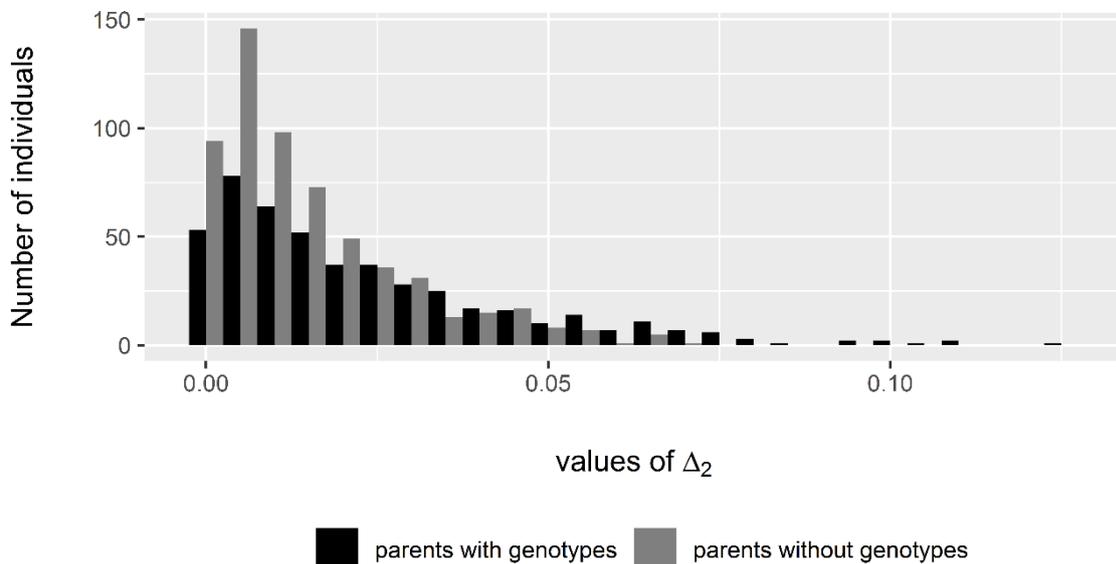
$$N_{miss} = 2 * \sum_{o=1}^N S_o \quad \text{where} \quad \begin{cases} S_o = 0 & \text{if } P1_m(o) > \text{median}(P2_m(\cdot)) \\ S_o = 1 & \text{if } P1_m(o) \leq \text{median}(P2_m(\cdot)) \end{cases}$$

where N is the total number of offspring.

If $\frac{N_{miss}}{N} \leq e$, then all the parent pairs with the highest Mendelian transmission probability are assigned as the parents of the progeny. In such a case, the estimated proportion of offspring with at least one missing parent is lower than the user accepted positive assignment error rate e , and choosing the

parent pair with the highest Mendelian transmission probability is the best solution to obtain a high assignment rate, respecting the accepted positive assignment error rate.

If $\frac{N_{miss}}{N} > e$, the assignment threshold is established as follows: $N - N_{miss}$ is the expected number of offspring with both parents genotyped. We also expect, as explained before, that offspring with both parents genotyped have higher values of $\Delta_2(\cdot)$ than offspring with missing parent(s), and thus the $N - N_{miss}$ highest values of $\Delta_2(\cdot)$ are expected to be those of offspring with both parents genotyped (Supplementary Figure 2.1). Then, the values of $\Delta_2(\cdot)$ are sorted in descending order, and $(1 - e)$ quantile of the $N - N_{miss}$ first values is defined as the threshold value. All the offspring for which $\Delta_1(o)$ is greater than the threshold value is assigned to their $PP1_m(o)$. A consequence of the use of $\Delta_2(\cdot)$ to set the assignment threshold is that the number of potential parents must be such that at least three parent pairs can be tested with each offspring.



Supplementary Figure 2.1 : Effect of missing parents of the distribution of $\Delta_2(\cdot)$ in a cohort of 1068 offspring of European sea bass from 39 sires and 14 dams genotyped for 200 markers. The distribution of $\Delta_2(\cdot)$ for individuals with both parents genotyped is represented in black and that of individuals with one missing parent in grey.

APIS is not able to distinguish between individuals that are not assigned due to missing parents or due to lack of power of the markers, as both missing parents and insufficient power cause an overlap between $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$ distributions. Qualitatively, when the distribution of $\Delta_1(\cdot)$ is composed of two separate sub-distributions, one of which is in the range of the $\Delta_2(\cdot)$ distribution, (like in Figure 2.1 with 20 missing sires and 200 markers), the animals in the lower sub-distribution of $\Delta_1(\cdot)$ are not assigned by APIS, and this is because their parents are missing. However, when the power of the marker set is too low (Figure 2.1 with 35 markers), the distributions of $\Delta_1(\cdot)$ and $\Delta_2(\cdot)$ tend to overlap even without missing parents and then, clearly distinguishing between individuals whose parents are not sampled and individuals with both parents genotyped but low $\Delta_1(\cdot)$ is impossible.

2.1.3.3. Ensemble de validation

A commercial seabass (*Dicentrarchus labrax*) cohort from the selected line of Ferme Marine du Douhet (FMD, la Brée les Bains, France) was genotyped on the 57K seabass SNP array DlabChip at the INRA genotyping platform GENTYANE (Clermont-Ferrand, France). The offspring cohort, composed of 1084 individuals, was obtained from 39 sires and 14 dams mated in a factorial design. After quality controls ($CallRate \geq 0.9$ and $DQC \geq 0.8$), 52813 SNPs were kept. Among those markers, we selected the 500 markers with the greatest minor allele frequency (MAF). This set of markers provides an excess of information for assignment (exclusion power of 1 with the formula proposed by Vandeputte (2012)), so the pedigree obtained from it is considered as the true pedigree. The 1084 individuals of the offspring set were assigned by exclusion using VITASSIGN (Vandeputte et al. 2006) allowing up to 25 mismatches. A total of 1068 offspring were assigned to a unique parent pair and constitute the reference pedigree.

2.1.3.4. Effets du nombre de marqueurs et des parents manquants

Within the genotypes of the individuals from the reference pedigree, four sets of markers (35, 42, 50 and 100 markers) were randomly chosen from the 52813 markers to reach a theoretical exclusion power (Vandeputte, 2012) of 0.90, 0.95, 0.99 and 0.99999, respectively. The corresponding data sets were used to perform parentage assignment with APIS, CERVUS (Kalinowski et al., 2007) and VITASSIGN (Vandeputte et al., 2006). For APIS, the acceptable positive assignment error rate was set to 1% (APIS1) or 5% (APIS5). CERVUS was tested with default simulation parameters (10000 offspring simulated with a genotyping error rate of 1% and a 95% confidence interval – CERVUS95). For VITASSIGN, we allowed for one mismatch (VITASSIGN1). The pedigree proposed by each software was compared to the reference pedigree.

Then, we tested the effect of missing parents on parentage assignment results. In order to test the most challenging situation, we only removed sire genotypes, as the risk of assignment error is higher when the genotype of one of the true parents is missing than when that of both parents is missing (Jamieson and C S Taylor, 1997). We created two series of data sets, one with 4 missing sires (~10% missing sires) and one with 20 missing sires (~50% missing sires), randomly chosen from the sires present in the reference pedigree. Sampling of missing sires was repeated ten times for both cases.

The datasets with missing sires were analyzed with the previously chosen 35, 42, 50 and 100 markers. The results of each software were compared with the reference pedigree.

To compare with the reference pedigree, we counted the number of offspring that were assigned to their true parents (N1), the number of offspring with both parents genotyped that were assigned to a wrong parent pair (N2), the number of offspring with both parents genotyped that were not assigned (N3), the number of offspring from missing sires that were assigned to a wrong parent pair (N4) and the number of offspring from missing sires that were not assigned (N5).

The metrics used to compare the different software were percent of true positives ($N1 / N$), true negatives ($N5 / N$), false positives or type I error ($(N2 + N4) / N$) and false negatives or type II error ($N3 / N$).

2.1.3.5. Effet du nombre de descendants

As APIS uses the observed distributions of Mendelian transmission probabilities to set assignment thresholds, it is expected that the genotypes of a minimum number of offspring need to be used to correctly describe these distributions. The effect of offspring number was tested by randomly choosing

from 50 to 1000 offspring (by steps of 50) in twelve base data sets used for software comparison (0, 4 and 20 missing sires with 35, 42, 50 and 100 markers). For each number of offspring tested, 100 repetitions (resampling of offspring) were done for each base data set, and the positive assignment error rate was calculated as the proportion of type I errors relative to positive results (assigned offspring).

$$\text{positive assignment error rate} = \frac{N2 + N4}{N2 + N4 + N1}$$

2.1.3.6. Validation sur d'autres jeux de données

APIS was tested on other data sets to study the robustness of the method. Using simulated data described in Grashei, Ødegård, & Meuwissen (2018), twelve different data sets of 1000 salmon from 100 sires and 200 dams were randomly chosen among the 50 data sets simulated with 1% (6 datasets) or 3% (6 datasets) genotyping error rate. Each data set was tested by APIS with user-set positive assignment error rates of 1 and 5%. Three levels of missing sires (0%, 10% and 50%) were tested. In each case, the number of markers used (125 to 130 SNPs) was defined to reach a theoretical assignment power of 0.99999 (Vandeputte, 2012).

2.1.4. Résultats

2.1.4.1. Efficacité de l'assignation avec APIS comparée avec d'autres logiciels

Case without missing sires

When no parents were missing (Figure 2.2a), all software were in general equivalent. They all gave about the same true positive and type I error rate with all marker sets. There were two exceptions, APIS1, which gave only 32.7% true positives, and 67.3% type II error when 35 markers were used, and VITASSIGN1 which gave 4.8% type II error with 50 markers. As the number of markers increased, the type I error rate decreased for APIS5, CERVUS95 and VITASSIGN1, from 3.5% for 35 markers to 0% for 100 markers.

Case with 4 missing sires

When 4 sires were missing (Figure 2.2b), APIS gave the lowest type I error rate (except with 100 markers, where VITASSIGN1 performed equally or better), but also the lowest true positive rate. As the number of markers increased, the true positive rate of APIS increased (31.3%, 46.4%, 43.1% and 87.2% for 35, 42, 50 and 100 markers respectively for APIS1 and 58.8%, 73.0%, 69.5% and 89.0% for 35, 42, 50 and 100 markers respectively for APIS5). The type I error rate of CERVUS95 was the highest (12.0%, 10.2%, 5.2% and 3.0% for 35, 42, 50 and 100 markers respectively). VITASSIGN1 had the highest true positive rate for 35 and 42 markers, where it also had an intermediate type I error rate.

Case with 20 missing sires

When 20 sires were missing (Figure 2.2c), the behavior of all software was comparable to the previous case with 4 missing sires. APIS had the lowest type I error rate except when 100 markers were used, but again also the lowest true positive rate. CERVUS95 had the lowest true assignment rate (true positives + true negatives) as well as a very high type I error rate (35.1%, 28.4%, 20.2% and 11.4% for 35, 42, 50 and 100 markers respectively). VITASSIGN1 had the highest true positive rate but an intermediate type I error rate, except when 100 markers were used.

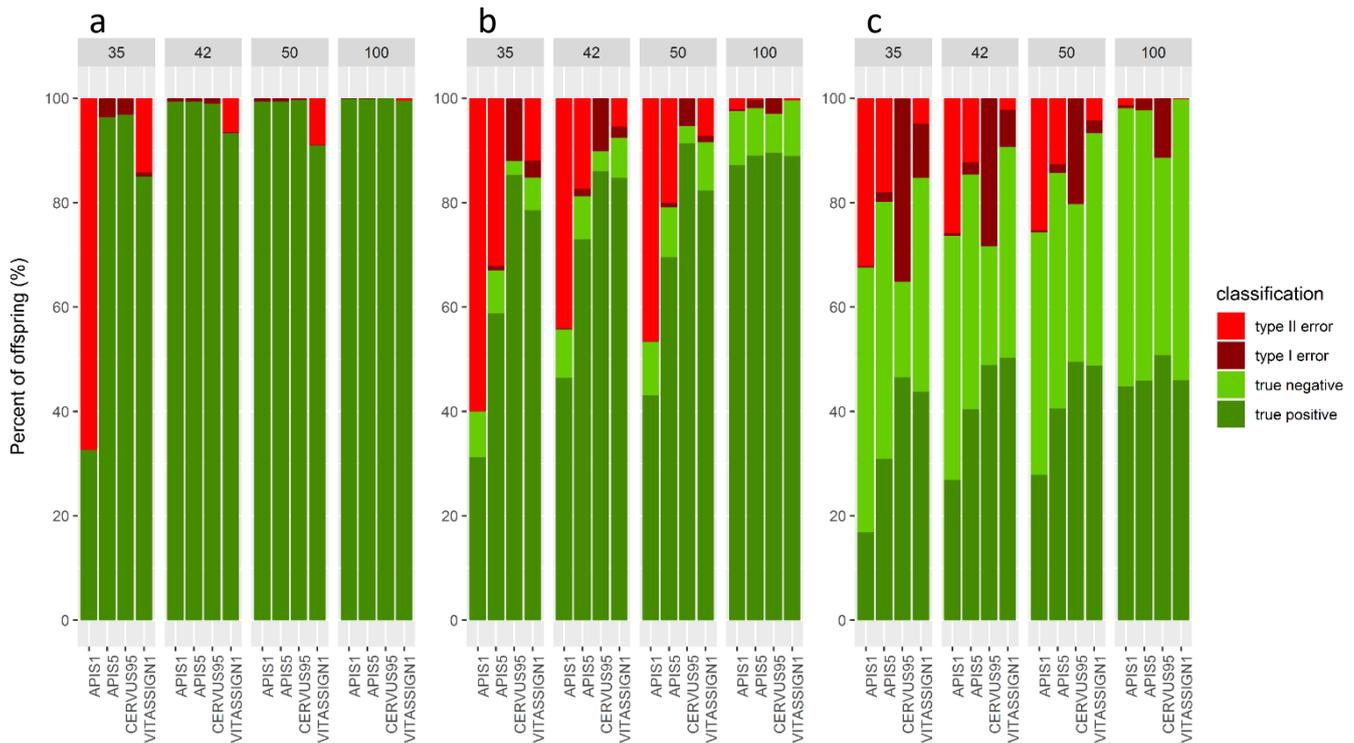
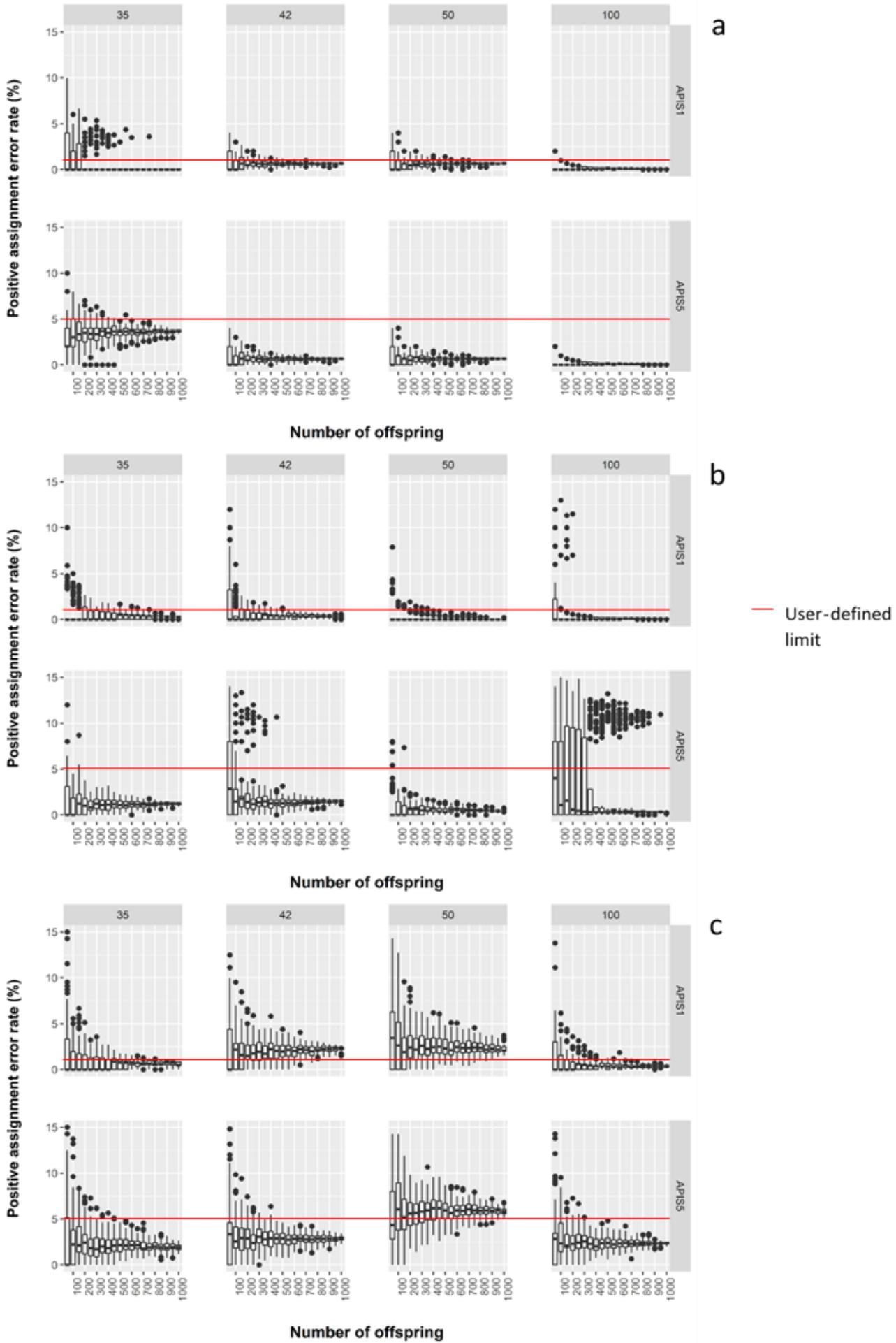


Figure 2.2 : Parentage assignment efficiency with zero (a), four (b) or twenty (c) missing sires in a cohort of 1068 sea bass from 39 sires and 14 dams using APIS1 (1% user-set acceptable positive assignment error rate), APIS5 (5% user-set acceptable positive assignment error rate), CERVUS95 (95% confidence level) and VITASSIGN1 (one mismatch allowed), with four sets of 35, 42, 50 and 100 SNP markers. In panels (b) and (c), the proportions are the means of 10 trials with a random resampling of missing sires.

2.1.4.2. Effet du nombre de descendants

In general, as the number of offspring increased, the variance of the positive assignment error rate decreased (Figure 2.3). In the most difficult cases, where the number of markers was low and/or when some parents were missing, the positive assignment error rate had a high variance in data sets with less than 500 offspring. In data sets with more than 500 offspring, the average positive assignment error rate was lower than the user-defined limit of 1% or 5% except when 20 sires were missing and 42 and 50 markers were used (Figure 2.3c for APIS1 and APIS5). In those cases, the positive assignment error rate converged to an average value slightly higher than the user-defined positive assignment error rate. With 42 markers the positive assignment error rate of APIS1 converged to 2.29% ($\pm 0.19\%$), and with 50 markers, it converged to 2.25% ($\pm 0.45\%$) for APIS1 and 5.83% ($\pm 0.32\%$) for APIS5. Except in these three cases, when the data sets comprised more than 500 offspring, the positive assignment error rate was below the user-defined limit in more than 90% of the simulations. We also can see that

the median of the positive assignment error rate converged to its asymptotic value in data sets with 200 offspring or more.



32 *Figure 2.3 : Effect of the number of offspring on the positive assignment error rate using APIS, with zero (a), four (b) or twenty (c) missing sires in a cohort of 1068 sea bass from 39 sires and 14 dams. Each boxplot represents the results of 100 random samples of offspring for each number of offspring tested. Each panel is divided into two rows (top: 1% user-defined maximum positive assignment error rate %, bottom: 5% user-defined maximum positive assignment error rate). The red line represents the user-defined maximum positive assignment error rate.*

2.1.4.3. Validation sur d'autres jeux de données

The effect of genotyping error on assignments is presented in Figure 2.4. For a 1% genotyping error, the results were consistent with those previously obtained with the real seabass data set, except with VITASSIGN which produced a lower true positive rate and a higher type II error rate. When the genotyping error rate increased from 1% to 3%, the type II error rate increased and true positives decreased with all programs. When the simulated genotyping error rate was 3%, APIS generated the same results whatever the genotyping error rate parameter of the software (1% or 3%), which was not the case for CERVUS where both the true positives and type I error rate increased (Supplementary Figure 2.2).

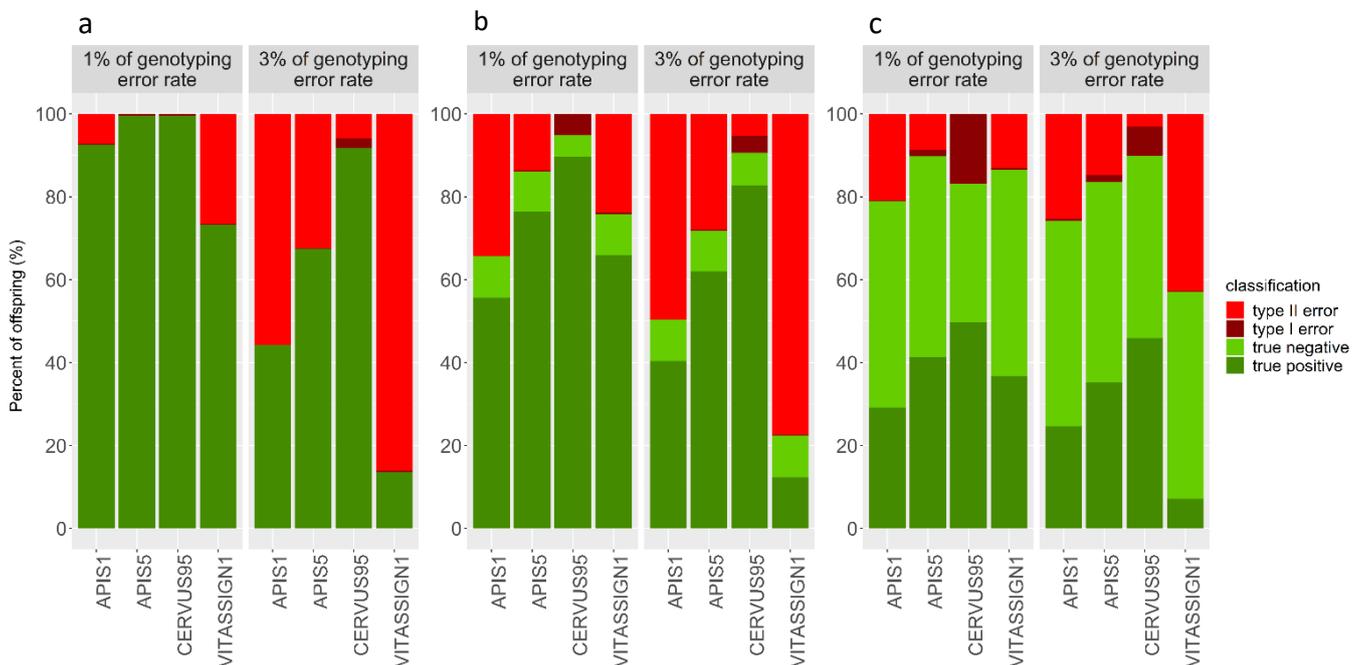
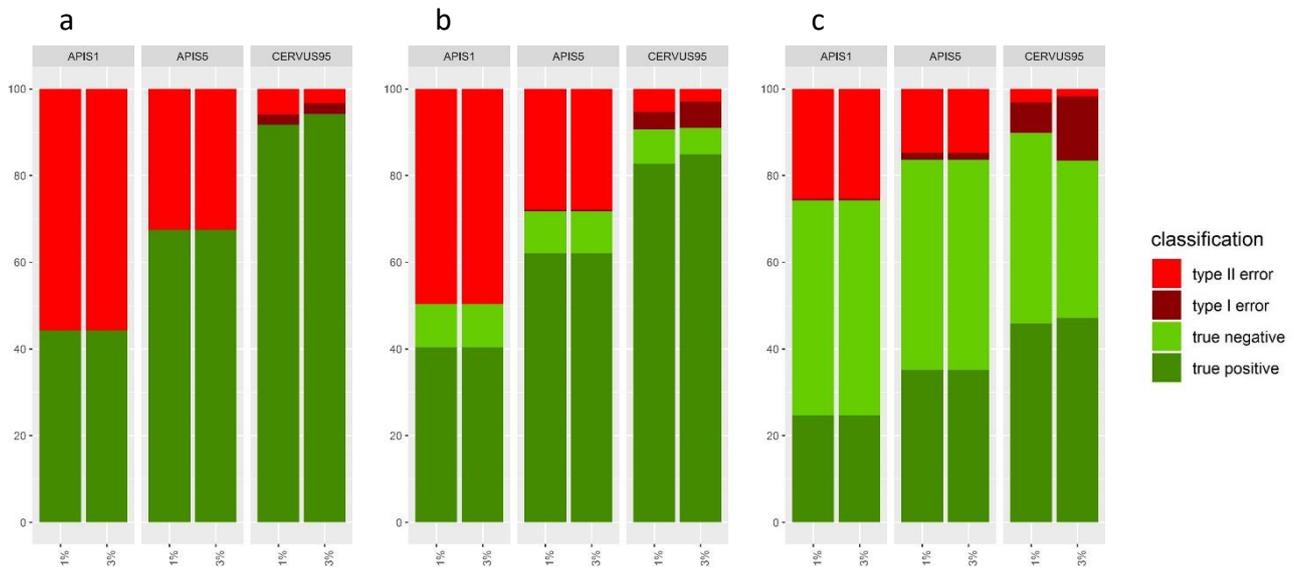


Figure 2.4 : Parentage assignment efficiency with zero (a), ten (b) or fifty (c) missing sires using simulated data from Grashei et al. (2018), with 1000 salmon from 100 sires and 200 dams genotyped on 120 randomly chosen SNP markers, with 1% and 3% genotyping errors tested using APIS1 (1% user-set acceptable positive assignment error rate), APIS5 (5% user-set acceptable positive assignment error rate), CERVUS95 (95% confidence level, 1% genotyping error setting) and VITASSIGN1 (1%, so two mismatches allowed). Assignment efficiency is averaged over six simulated data sets.

2.1.5. Discussion

The new parentage assignment method implemented in APIS, based on the observed distributions of Mendelian transmission probabilities, was designed to provide an alternative to the simulation process proposed by most of the likelihood-based parentage assignment software. When the number of markers is high and all the parents are genotyped, all software are able to assign all the offspring to their true parents without errors. However, when the situation is more complex, with less markers and/or missing parents in the data set, APIS gives more reliable results, by limiting the positive assignment error rate. Such complex situations, especially missing parents, may frequently occur. Missing parents can be caused by traceability issues, technical problems during genotyping (DNA degradation, extraction and genotyping issues), or absence of identified biological samples. Those issues may be known or not by the user. Indeed, offspring with missing parents which are declared as assigned to a parent pair (a wrong one) lead to pedigree errors. Such errors in the pedigree may lead to incorrect estimations of breeding values (Banos et al., 2001; Visscher et al., 2002). In Visscher et al.

(2002), 10% of the pedigree was wrong, leading to a predicted loss of selection response of 2 - 3%. Banos et al. (2001) show a decrease of EBV estimations by 11% - 15% when there are 11% of paternity errors in the pedigree. Pedigree errors also have a high impact in wild populations conservation, where pedigree errors lead to a decrease of genetic diversity (Oliehoek and Bijma, 2009).



Supplementary Figure 2.2 : Assignment efficiency with zero (a), ten (b) or fifty (c) missing sires using simulated data from Grashei et al. (2018), with 1000 salmon from 100 sires and 200 dams genotyped on 120 randomly chosen SNP markers, with a 3% genotyping error tested using APIS1 (1% user-set acceptable positive assignment error rate), APIS5 (5% user-set acceptable positive assignment error rate), CERVUS95 (95% confidence level). Each software program was tested with a 1% or 3% of genotyping error rate setting. Assignment efficiency is averaged over six simulated data sets.

APIS was designed to control the type I error rate through the positive assignment error rate, defined by the user input parameter. While minimizing pedigree errors requires an as low as possible positive assignment error rate, minimizing type II error can also be an objective, especially when the variable of interest is the proportion of unassigned offspring (e.g. in a stock restoration program, identifying animals that are from natural reproduction and not from the restocking operations). However, a general rule is that, all things being equal, minimizing type I error will lead to an increase of type II error (DePoy and Gitlin, 2016). This is the case in our results (Figure 2.2c), and thus APIS is not the most appropriate software when type 2 error has to be minimized.

We tested APIS in many situations, with different data sets. The results obtained are similar across all situations tested. The observed positive assignment error rate was most of the times lower than the user-set maximal value, except in some cases when 50% of the sires were missing, where the error rate exceeded the limit, but generally by a modest amount. We could observe that when the proportion of offspring with at least one missing parent was high, the median of the $P2_m(\cdot)$ was shifted to lower values. This is probably caused by the fact that in these cases, the distribution of $P2_m(\cdot)$ is a mix of $P2_m(\cdot)$ from individuals with both parents genotyped and from individuals with missing parents. The values of $P2_m(\cdot)$ for individuals with missing parents are indeed equivalent to the values of $P3_m(\cdot)$ for individuals with both parents genotyped, which are a little lower than the values of $P2_m(\cdot)$ for individuals with both parents genotyped. As the proportion of individuals with at least one missing parent increases, the distribution of $P2_m(\cdot)$ shifts to lower values, its median

decreases, and the estimate of N_{miss} decreases as well, resulting in more relaxed thresholds and more type I errors in the results than expected. However, although the user-defined threshold of APIS was not respected in a few extreme cases, APIS performed better than CERVUS and VITASSIGN in terms of type I error rate in those cases and in general (Figure 2.2c).

As APIS uses the observed distributions of Mendelian transmission probability, several factors may impact the distributions shape and thus, the assignment efficiency. We showed that the power of the marker set impacts the overlap between $P1_m(\cdot)$ and $P2_m(\cdot)$ and the estimation of the assignment threshold. High proportions of offspring with missing parents may also lead to misestimate the threshold, as discussed before. We showed that APIS requires 500 offspring population to match the aim of controlling the positive assignment error rate (Figure 2.3). However, except in a few cases where the proportion of individuals with missing parents is under-estimated, 200 offspring are generally enough for the median positive assignment error rate to reach its asymptotic value (Figure 2.3). Thus, APIS is not designed to handle small batches of offspring, and 200 individuals seems to be the very minimum. A fourth cause that may impact the distributions and the estimation of the threshold could be inbreeding. It has been shown previously that the type I error rate increases as the degree of relatedness between the members of a trio increases (Anderson & Garza, 2005; Marshall, Slate, Kruuk, & Pemberton, 1998). When the true parents of an offspring are full-sibs or one of the true parents is a full-sib of the offspring, exclusion can be more efficient than a likelihood approach (Anderson & Garza, 2005). In the case of APIS, when the offspring is divided into groups with varying levels of inbreeding, their average $P1_m(\cdot)$ will vary as a function of inbreeding. In an inbred group, AA genotypes will be more frequent in parents and offspring, leading to higher Mendelian transmission probabilities than in an outbred group (Tableau 2.1 : Mendelian transmission probabilities for one marker when the *offspring is homozygous (AA), conditional on the sire and dam genotypes. Adapted from Boichard et al. (2014)*). Depending on the population structure, different peaks may then appear in $P1_m(\cdot)$, leading to unreliable results, mainly caused by an improper estimation of the number of offspring with missing parents. In the European seabass commercial cohort we tested, even if inbreeding was not estimated, it is reasonable to think that it was higher than in a wild population because of the history of domestication and selection. However, as the number of parents at each generation (here 39 sires and 14 dams in a yearly cohort) and the number of families are rather large, we expect inbreeding to be quite evenly distributed in the population. In this situation, we obtained reliable results as shown in the study.

The mating design may have an impact on the reliability of APIS in a similar way as inbreeding does. If one sex repeatedly has very few offspring, this could lead to problem of distributions with multiple peaks. For this reason, the output of APIS shows the distribution of $P1_m(\cdot)$, $P2_m(\cdot)$, $\Delta_1(\cdot)$ and $\Delta_2(o)$ to help the user identify these potential issues.

Another limitation is the knowledge of the sex of the parents. If the sex is unknown for all potential parents and the same genotype file is used for sires and dams, the true parents will be assigned as is (true sire assigned as sire and true dam assigned as dam) but also in the other way round (true sire assigned as dam and true dam assigned as sire). Those two symmetrical parent pairs will be the first and second best pairs. In this situation, $\Delta_1(o)$ will be equal to zero. Thus, APIS will not be able to set up the threshold, leading to no assignment. A solution to this is that the user detects such situation in the log file of APIS and progressively assigns arbitrary sexes to the parents of offspring for which $\Delta_1(o)$ is equal to zero, until all $\Delta_1(o)$ have positive values.

Finally, APIS is not suitable to assign overlapping generations. In this situation, several individuals will be both in offspring and parental genotype data sets. Thus, these individuals may be assigned to

themselves as sire or dam because the Mendelian transmission probability of “parent” pairs containing the individual to assign is very high.

The causes of assignment errors are mainly genotyping errors and missing parental genotypes. In likelihood-based assignment software, the genotyping error rate is required as a parameter of the model (Kalinowski et al., 2007). If this parameter is not correctly estimated, it may lead to assignment errors (Grashei et al., 2018). In APIS, the genotyping error rate is fixed at 1% as proposed in Sancristobal and Chevalet (1997). The value of the genotyping error rate must be greater than zero to allow some mismatches but have a value small enough to penalize incompatibility, but its value is not supposed to be critical for assignment accuracy (Boichard et al., 2014). Still, very high error rates such as the 3% error rate used in some of the salmon simulated datasets had a noticeable effect on assignment efficiency (Figure 2.4 : Parentage assignment efficiency with zero (a), ten (b) or fifty (c) missing sires using simulated data from Grashei et al. (2018), with 1000 salmon from 100 sires and 200 dams genotyped on 120 randomly chosen SNP markers, with 1% and 3% genotyping errors tested using APIS1 (1% user-set acceptable positive assignment error rate), APIS5 (5% user-set acceptable positive assignment error rate), CERVUS95 (95% confidence level, 1% genotyping error setting) and VITASSIGN1 (1%, so two mismatches allowed). Assignment efficiency is averaged over six simulated data sets.). However, such levels of errors are not expected to happen in SNP genotyping (Ranade et al., 2001). The second parameter is the proportion of missing parents. In APIS, this proportion is estimated with the distributions of Mendelian transmission probabilities. Even if APIS sometimes tends to overestimate the proportion of missing parents, leading to higher type II error rates, this parameter is most of the time unknown to the user and incorrect estimation (especially setting it to zero when it is not the case) may lead to incorrect assignment in other software (Oddou-Muratorio et al., 2003). This is what we see in our tests with CERVUS.

APIS was designed to limit the type I error rate. As it does not require any input parameter other than the genotypes of the potential parents and of the offspring, as well as the maximum acceptable positive assignment error rate the user accepts, APIS is especially convenient and efficient when the user does not have any *a priori* information about the population to assign. Although we showed that exclusion, using VITASSIGN, is more efficient in several cases (especially when many markers are available, where type I error rate falls to an acceptable level), it also requires an input parameter which is the maximum allowed number of mismatches. This number is set empirically and depends on many factors such as the variability of the markers used, the genotyping error and the number of markers, without existing decision rules (Vandeputte et al., 2006). As APIS sets thresholds by itself, it is more convenient than exclusion if the user does not have any *a priori* information about how many mismatches to allow.

2.2. Discussion complémentaire : Utilisation de l'assignation de parenté dans un contexte aquacole

Grâce au développement de puces à SNPs, le nombre de polymorphismes est devenu si important qu'il est possible de les utiliser pour identifier les liens de parenté dans les programmes de sélection les utilisant, c'est-à-dire les programmes de sélection génomique. Cependant, même dans ce cas, la fixation des seuils permettant de déclarer qu'un individu est bien descendant d'un couple de parents n'est pas triviale. La principale difficulté réside dans l'absence fréquente de certains génotypes parentaux qui mène facilement à des erreurs de pedigree, en particulier si des apparentés des vrais parents sont présents parmi les parents potentiels, ce qui est par nature le cas dans les programmes de sélection aquacole. APIS a été développé dans l'objectif de répondre à ce problème afin de réduire les erreurs d'assignation. Dans le manuscrit présenté dans la section précédente, nous avons volontairement diminué la quantité d'information afin de tester APIS dans des cas difficiles. En effet, dans nombre de programmes de sélection aquacole, la sélection génomique n'est pas encore en place, et des marqueurs génétiques sont utilisés pour établir le pedigree des animaux. Dans ce cas, afin de limiter les coûts, il est essentiel de pouvoir disposer d'assignations fiables tout en utilisant un nombre minimal de marqueurs. En pratique, lorsque la totalité des marqueurs est disponible, il est plus prudent d'en utiliser entre 500 et 1 000. Même si la méthode mise en œuvre dans le logiciel permet d'obtenir des résultats de bonne qualité de manière complètement automatisée, APIS fournit aussi à l'utilisateur des graphiques diagnostiques, lui permettant de vérifier le bon déroulé du processus ou bien d'en corriger les résultats.

Les graphiques proposés par APIS lorsque le processus d'assignation est terminé permettent de vérifier la validité des résultats obtenus. Tout d'abord, l'utilisateur peut rapidement identifier si des génotypes parentaux sont manquants dans son jeu de données. Dans l'exemple de la Figure 2.5, la superposition des histogrammes rouge (représentant le couple parental le plus probable) et bleu (représentant le deuxième couple parental le plus probable) et dans la partie gauche du panneau central (compris entre des probabilités mendéliennes de transmission allant de 0,3 à 0,55) montre la présence d'individus avec au moins un parent manquant. Les distributions des mismatches (incompatibilités mendéliennes entre le descendant et ses parents potentiels) (Figure 2.5, panneau du bas) permettent de vérifier la qualité globale de l'assignation. En général, le couple de parents le plus probable, ici en rouge, qui est celui avec le moins de mismatches, doit en avoir en très faible proportion par rapport au nombre de marqueurs utilisés. Ici, les couples les plus probables ont moins de trois mismatches avec le descendant auquel ils sont assignés, ce qui correspond à moins de 3% de mismatches. Si le pourcentage de mismatches des couples les plus probables est grand et proche de celui des deuxièmes couples les plus probables, il peut s'agir de génotypes avec de nombreuses erreurs ou d'un jeu de données avec de nombreux parents manquants. Ainsi, les graphiques proposés par APIS permettent de s'assurer du bon déroulement de l'assignation, de vérifier si certains génotypes parentaux sont absents voire même la qualité du génotypage. De plus, il est possible à l'utilisateur de redéfinir les valeurs seuil de probabilités mendéliennes ou de delta (différence entre la probabilité mendelienne du couple de parents le plus probable et la probabilité mendelienne du deuxième couple de parents le plus probable), mais aussi d'utiliser une méthode d'exclusion. Dans ce cas, la distribution observée du nombre de « mismatches » entre les premiers et seconds couples les plus probables permet une décision informée sur ce nombre acceptable de mismatches. Aujourd'hui ce nombre est fixé de manière totalement empirique (cf. Vandeputte et al., 2006), qui plus est sur des modèles développés sur marqueurs microsatellites, dont le nombre d'erreurs de génotypage, cause majeure des mismatches, diffère beaucoup de celui des SNPs (Ranade et al., 2001). Ici, sur la Figure 2.5, nous pouvons voir que les seconds couples ayant le moins de mismatches, en bleu (qui sont par définition des couples faux dans la très grande majorité

des cas, le couple vrai étant en général celui qui présente le moins de mismatches) ont un minimum de 4 mismatches. On peut en conclure que fixer un seuil à 3 mismatches maximum, dans le cas étudié, permet d'assigner les animaux avec un degré de certitude très élevé.

Lorsque le nombre de marqueurs est plus élevé, les graphiques en sortie de APIS sont semblables à ceux dans l'article 2.1 qui présentent des situations d'assignation simple, où des seuils sur le nombre de mismatches ou les distributions des probabilités mendéliennes de transmission peuvent être choisis « à l'œil ». Sur la Figure 2.6, on peut voir que, contrairement à la Figure 2.5, les distributions des probabilités mendéliennes (Figure 2.2, panneau du haut) et celles des mismatches (Figure 2.6, panneau du bas) sont bien séparés. Ici, la distribution des probabilités mendéliennes du couple le plus probable, représentée par les histogrammes en rouge, est divisée en deux pics, l'un à droite du graphique et l'autre à gauche, superposé à la distribution du deuxième couple le plus probable (en bleu). Ainsi, un seuil placé entre les deux, donc vers un delta de 0,12, une probabilité mendélienne de 0,5 ou un nombre de mismatches de 25 permet d'assigner les individus à leurs parents avec un risque quasi-nul de fausse assignation. Dans certains cas, cette fixation visuelle des seuils peut être plus pertinente que l'utilisation de l'algorithme utilisé par APIS, même si nous avons montré sa validité dans un grand nombre de cas.

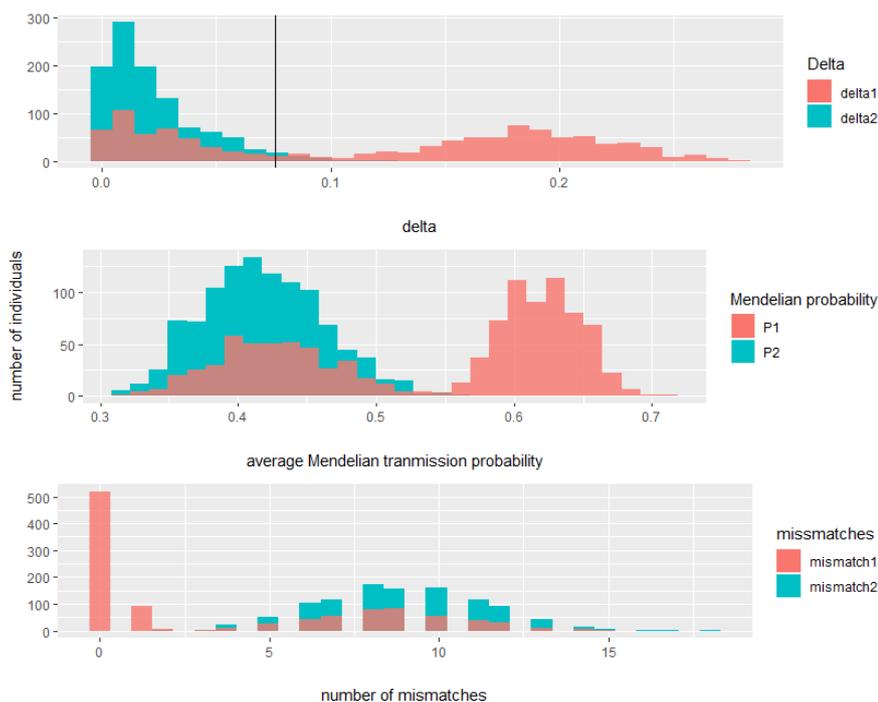


Figure 2.5 : Exemple de graphique produit par APIS lors de l'assignation de 1 068 bars européens génotypés sur 100 marqueurs SNPs. Le panneau du haut représente les distributions des deltas (différence entre deux probabilités mendéliennes de transmission) ainsi que la valeur seuil calculée avec APIS (en trait plein noir). Le panneau du milieu représente les distributions des probabilités mendéliennes de transmission estimées des première et deuxième paires les plus probables. Le panneau du bas représente les distributions des mismatches des premiers et deuxièmes couples les plus probables. Les histogrammes en rouge représentent les distributions du couple le plus probable et ceux en bleus représentent les distributions du deuxième couple le plus probable.

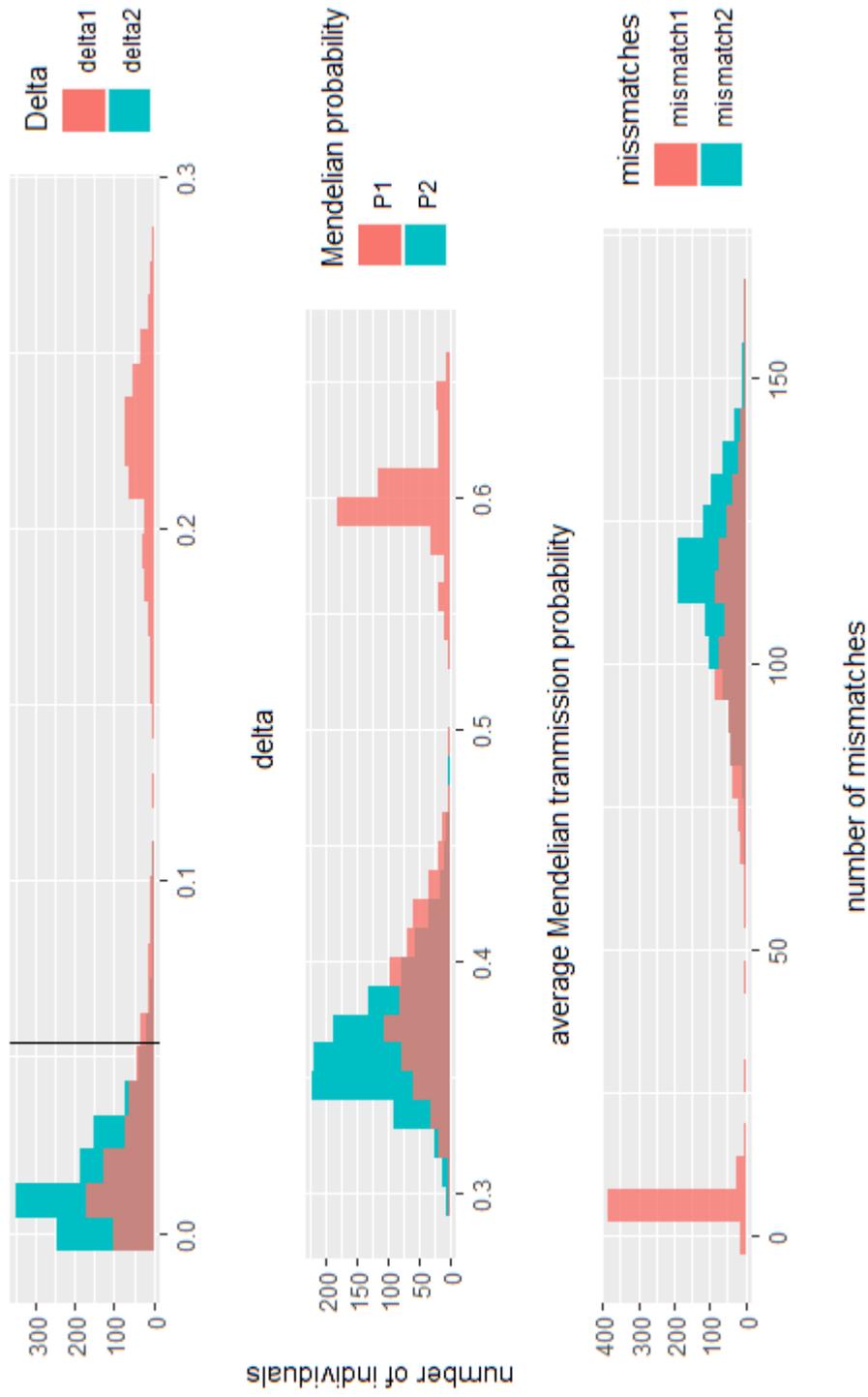


Figure 2.6 : Exemple de graphique produit par APIS lors de l'assignation de 1 068 bars européens génotypés sur 1 000 marqueurs SNPs. Le panneau du haut représente les distributions des deltas (différence entre deux probabilités mendéliennes de transmission) ainsi que la valeur seuil calculée avec APIS (en trait plein noir). Le panneau du milieu représente les distributions des probabilités mendéliennes de transmission estimées des première et deuxième paires les plus probables. Le panneau du bas représente les distributions des mismatches des premiers et deuxième couples les plus probables. Les histogrammes en rouge représentent les distributions du couple le plus probable et ceux en bleu représentent les distributions du deuxième couple le plus probable.

Dans l'article, nous avons annoncé que APIS fonctionnait également avec des marqueurs microsatellites. Afin de valider cela, nous avons utilisé les animaux du croisement décrit en 2.1.3.3, mais génotypés sur douze marqueurs microsatellites. Ici, la puissance du jeu de marqueurs valait 0,9999. Nous avons également testé deux niveaux d'erreurs acceptés par l'utilisateur, un de 1% (APIS1) et un de 5% (APIS5) avec 0, 4 ou 20 génotypes de pères manquants.

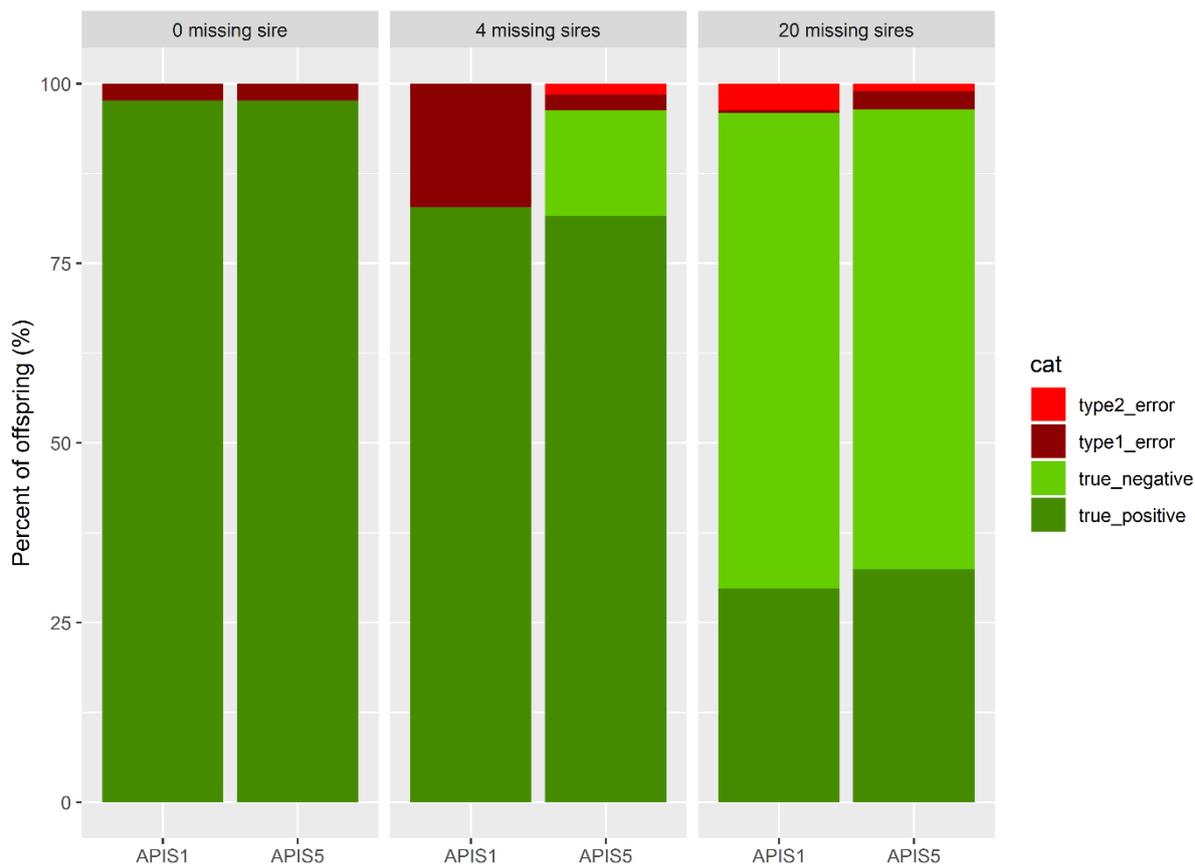


Figure 2.7 : Efficacité de l'assignation avec zero (a), quatre (b) ou vingt (c) pères manquants dans une cohorte de 1068 bars européens issus de 39 pères et 14 mères en utilisant APIS1 (1% d'erreur dans les résultats accepté) et APIS5 (5% d'erreurs dans les résultats accepté).

Les résultats obtenus sont identiques à ceux obtenus avec des marqueurs SNPs. Le taux d'erreurs dans les résultats est de 2,3% lorsque tous les génotypes parentaux étaient présents (Figure 2.7a). Lorsque quatre pères étaient manquants, le taux d'erreurs dans les résultats était de 17,2% avec APIS1 et 2,3% avec APIS5 (Figure 2.7b). Enfin, lorsque vingt pères étaient manquants, APIS1 donnait 0,2% d'erreurs dans les résultats et APIS5 donnait 2,6% d'erreurs dans les résultats. Seul APIS1 lorsque zéro ou quatre génotypes de pères étaient manquants donne un taux d'erreurs dans les résultats supérieur à celui accepté par l'utilisateur. Dans le cas avec tous les génotypes parentaux présents, cela peut être dû à la puissance du jeu de marqueurs ou à des erreurs de génotypage. Ensuite, lorsque quatre pères étaient manquants, un taux d'erreurs dans les résultats élevé peut également être dû à des erreurs de génotypages qui peuvent réduire la puissance réelle du jeu de marqueurs. Nous avons également observé des résultats similaires dans la section 2.1.4.2. Nous avons émis l'hypothèse que ces résultats étaient dus à une mauvaise estimation par l'algorithme d'APIS du nombre d'individus avec au moins

un parent manquant, ce qui pouvait entraîner l'assignation de l'ensemble de la population à un couple et donc, à un fort taux d'erreur. En effet, les résultats présentés ici ne proviennent que d'un exemple et pas de nombreux échantillonnages comme ceux présentés dans l'article. Nous pouvons néanmoins conclure que APIS donne des résultats équivalents, qu'il soit utilisé avec des marqueurs SNPs ou des marqueurs microsatellites. Le paramètre le plus important est la puissance du jeu de données, qui va permettre de discriminer correctement les différents couples de parents possibles et d'assigner chaque individu au bon couple si les deux parents sont génotypés ou ne pas assigner si au moins l'un des deux est manquant.

Dans les programmes de sélection où la fécondation artificielle est majoritairement employée, l'information du plan de croisement est connue. Elle peut en théorie permettre de ne tester que les couples de parents qui ont effectivement été croisés et ainsi éviter d'assigner un individu à un couple de parents qui n'est pas censé avoir de descendants. Or, l'ajout de cette information en amont de l'assignation peut provoquer des erreurs. En effet, il peut y avoir des différences entre le plan de croisement souhaité et celui qui a été réalisé. Elles peuvent être causées par une erreur d'identification des géniteurs lors de la fécondation ou par le mélange de sperme et d'œufs qui ne figurent pas sur le plan de croisement et qui n'a pas été rapporté. Ainsi, le plan de croisement réel doit être considéré comme un résultat de l'assignation de parenté plutôt que comme une variable d'entrée. Lorsque la puissance du jeu de marqueurs est élevée, les résultats obtenus ne contiennent que très peu voire pas d'erreurs. Le pedigree obtenu permet de recréer le plan de croisement réalisé et de le confronter au plan de croisement théorique. Les différences entre les deux peuvent être analysées plus en profondeur grâce aux graphiques diagnostiques de APIS et aussi grâce au fichier *log*, contenant les informations issues de l'assignation.

Si ces fonctions avancées permettent de comprendre et éventuellement résoudre des cas difficiles, le principal avantage d'APIS reste sa facilité d'utilisation, avec une fixation automatique des seuils d'assignation, sans hypothèse nécessaire sur les taux d'erreurs de génotypages ou les parents manquants, tout en garantissant dans ses résultats un taux d'erreurs inférieur à celui souhaité par l'utilisateur.

Ces avantages ont permis d'inclure APIS dans les chaînes d'analyses du SYSAAF lors des traitements de routine des données des programmes de sélection aquacole et avicoles.

CHAPITRE 3 :
ETUDE DE L'ARCHITECTURE
GENETIQUE DE LA RESISTANCE
A DIFFERENTS PATHOGENES
CHEZ LE BAR ET LA DAURADE

Ce chapitre présente l'étude de l'architecture génétique, et plus particulièrement la détection de QTL, de la résistance à la nodavirose et à la vibriose chez le bar ainsi que de la résistance à la pasteurellose chez la daurade. L'objectif était de disséquer plus finement l'architecture génétique de la résistance à ces différents pathogènes, dans le but de découvrir des QTLs à effets modérés ou forts pouvant être utilisés en sélection assistée par marqueurs.

Ce chapitre se divise en trois parties. La première présente la détection de QTL de résistance à la nodavirose chez le bar dans deux cohortes commerciales provenant de sélectionneurs adhérents du SYSAAF (EMG et FMD) ainsi que dans quatre familles de back-cross produites à l'Ifremer de Palavas-les-Flots. Cette partie a été publiée dans *Aquaculture* en 2020 (doi : 10.1016/j.aquaculture.2020.735930). Les deuxième et troisième parties présentent les détections de QTL de résistance à la vibriose chez le bar et à la pasteurellose chez la daurade, respectivement.

3.1. Détection de QTL de résistance à la nodavirose dans trois populations de bar européen par l'utilisation d'une puce de 57k SNP.

Cette section a été publiée en 2020 dans *Aquaculture* 530 (735930) sous le nom :

« Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. »

Griot R.^{a,b,c}, Allal F.^c, Phocas F.^b, Brard-Fudulea S.^d, Morvezen R.^a, Bestin A.^a, Haffray P.^a, François Y.^a, Morin T.^f, Poncet C.^e, Vergnet A.^c, Cariou S.^g, Brunier J.^h, Bruant J.S.^g, Peyrou B.^h, Gagnaire P.A.ⁱ, Vandeputte M.^{b,c}

^a SYSAAF, Station LPGP/INRAE, Campus de Beaulieu, 35042 Rennes, France

^b Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

^c MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, 34250 Palavas-les-Flots, France

^d SYSAAF, UMR BOA, Centre INRAE Val de Loire, 37380 Nouzilly, France

^e INRAE-UCA, UMR GDEC, 63000 Clermont-Ferrand, France

^f ANSES, Ploufragan-Plouzané-Niort Laboratory, Viral Fish Diseases Unit, National Reference Laboratory for Regulated Fish Diseases, Technopôle Brest-Iroise, 29280 Plouzané, France

^g Ferme Marine Du Douhet, 17840 La Brée Les Bains, France

^h Ecloserie Marine de Gravelines-Ichthus, 59273 Gravelines, France

ⁱ ISEM, Univ. Montpellier, CNRS, IRD, 34000 Montpellier, France

DOI : <https://doi.org/10.1016/j.aquaculture.2020.735930>

3.1.1. Résumé

La Nécrose Nerveuse Virale (NNV) ou nodavirose est une menace majeure pour la production aquacole de bar européen (*Dicentrarchus labrax*), pour laquelle l'amélioration de la résistance par sélection génétique pourrait être une solution intéressante. Avec le développement des technologies de séquençage haut-débit, l'identification de régions génomiques impliquées dans la résistance est un moyen potentiel d'améliorer l'efficacité de la sélection. L'objectif de cette étude a été d'identifier les régions du génome (Quantitative Trait Loci - QTL) impliquées dans la résistance à la NNV, et de quantifier leurs effets.

Quatre familles expérimentales issues de rétro-croisements entre pères hybrides résistants-sensibles et mères sensibles, de 378, 454, 291 et 211 individus ont été produites. Par ailleurs, deux populations commerciales (A et B), de 1 027 et 1 042 individus ont été obtenues à partir de plans de croisement factoriels partiels (59♂ x 20♀ pour la population A; 39♂ x 14♀ pour la population B). Ces six groupes (4 familles et 2 populations commerciales) ont été soumis à un challenge infectieux par balnéation avec un nodavirus de type génétique RGNNV (redspotted grouper nervous necrosis virus). Un panel de SNP (Single Nucleotide Polymorphism) de haute densité a été développé pour créer une puce à SNP (ThermoFisher Axiom™ 57k SNP DlabChip), qui a été utilisée pour le génotypage de tous les individus ainsi que pour la création d'une carte génétique de haute résolution. Dans les quatre familles de rétro-croisement, respectivement 30 917, 23 592, 30 656 et 31 490 marqueurs ont été utilisés dans une approche de cartographie d'intervalles. Dans les populations commerciales, 40 263 marqueurs dans la pop A et 41 166 marqueurs dans la pop B ont permis de réaliser des études d'association pangénomiques (GWAS - Genome-Wide Association Studies) avec deux modèles, GBLUP et BayesCπ.

Un QTL a été identifié sur le chromosome LG12 dans trois des quatre familles de rétro-croisement, et un QTL supplémentaire a été détecté sur le LG8, uniquement dans une famille. Chez les populations commerciales, la détection de QTL a révélé un total de sept QTL, parmi lesquels a été retrouvé le QTL positionné sur le LG12, présent dans les deux populations. Ce QTL, qui a été positionné dans un intervalle de confiance de 3,45 cM, explique 9,21% de la variance génétique dans la pop A, alors que les autres QTL identifiés expliquent chacun moins de 1% de la variance génétique.

L'identification de QTL impliqués dans la résistance à la nodavirus chez le bar européen, dont un avec un effet fort, devrait avoir un impact important sur l'industrie aquacole. De futurs travaux pourront se concentrer sur la détection précise de la mutation causale du QTL présent sur le LG12 par une approche de séquençage du génome complet.

3.1.2. Introduction

European sea bass (*Dicentrarchus labrax*) is a major species for Mediterranean aquaculture, with a production of more than 150,000 tons in 2016 (FEAP, 2017). A major threat for the sea bass aquaculture industry is the occurrence of disease outbreaks, and especially of Viral Nervous Necrosis (VNN) disease mostly in warmer growing Mediterranean areas (Breuil et al., 1991; Le Breton et al., 1997; Vendramin et al., 2016). The VNN disease is caused by the Nervous Necrosis Virus (NNV), a RNA virus belonging to the Betanodavirus genus (Mori et al., 1992). NNV is widespread worldwide and can infect more than 70 marine and freshwater species (Doan et al., 2017b). In sea bass production, mortalities up to 90% have been recorded especially at the very susceptible larval and juvenile stages (Le Breton et al., 1997). Several vaccine strategies have been tested, leading to an improvement of survival (Doan et al., 2017b). However, three points limit the use of vaccines in VNN outbreak management: i) vaccines are generally expensive, ii) the number of fish to vaccinate is high which increases the cost issue (Ulmer et al., 2006), and iii) VNN is more infectious in early life stages, while vaccination of small fish is challenging (Gomez-Casado et al., 2011; Sommerset et al., 2005).

One promising solution to reduce the effects of VNN outbreaks in aquaculture farms is selective breeding for improved resistance to VNN. Selective breeding for disease resistance in aquaculture species has demonstrated its interest with a genetic gain in survival of 12.5% per generation on average, across different host-pathogen pairs (Gjedrem and Robinson, 2014). For selective breeding to be successful, enough genetic variation for disease resistance within the species is however required. Moderate to high heritability has already been reported for resistance to VNN in European sea bass (from 0.26 to 0.43), confirming the interest of selective breeding for this trait (Doan et al., 2017a; Palaiokostas et al., 2018a).

Genomic markers can be used to detect Quantitative Trait Loci (QTL), which are regions of the genome involved in the phenotypic variation of the trait (Georges et al., 1993). Their use allows a better understanding of the genetic architecture of complex traits as well as the location of the potential candidate genes involved in the resistance phenotype. With the access to high-throughput genotyping technologies at a reasonable cost for non-model species, some disease resistance QTLs have been found in a limited number on aquaculture species (see review by Ødegård et al. (2011)). In Atlantic salmon, major QTLs were discovered for IPNV resistance, explaining up to 83% of the genetic variance (Houston et al., 2008; Moen et al., 2009). These results can then be used to accurately choose broodstock among the candidates for selection or commercial diffusion by Marker Assisted Selection (MAS) (Houston et al., 2008; Moen et al., 2009). In Asian sea bass (*Lates calcarifer*), QTL mapping revealed nine QTLs involved in VNN resistance, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et al., 2017). In Atlantic cod (*Gadus morhua*), five QTLs were detected, including three

QTL with major effects, explaining between 14.2% and 19.7% of the phenotypic variance (Baranski et al., 2010). For VNN resistance in sea bass, a recent study reported three first QTLs, each of them explaining a small part of the genetic variance (1.5% to 4%) (Palaikostas et al., 2018a).

In this study, we investigated the genetic architecture of VNN resistance in sea bass. We performed Genome Wide Association Studies (GWAS) based on two commercial cohorts and four experimental backcross families, all genotyped on the newly developed ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip. Two approaches were considered to detect associations between SNPs and resistance to VNN, interval mapping and a multi-marker Bayesian variable selection model to estimate the sharing of genetic variance explained by the QTLs, as well as their credibility intervals.

3.1.3. Matériel et Méthodes

3.1.3.1. *Approbation du comité d'éthique*

All infection challenges were carried out in accordance with the European guidelines (Directive 2010–63-EU) and the corresponding French legislation. Animal experiment procedures were approved by the ethics committee on animal experimentation COMETH ANSES/ENVA/UPC No. 16 and were authorized by the French Ministry of Higher Education, Research and Innovation under numbers 2017022816255366, 29/01/13-5 and 10/03/15-1.

3.1.3.2. *Animaux utilisés*

The animals challenged came from two commercial cohorts from the breeding programs of two different companies and four experimental backcross families. The global study workflow was presented in Tableau 3.1. The commercial cohorts were produced by artificial mating and further mentioned as pop A and pop B. Pop A (1680 individuals) was produced from a mating of 59 sires by 20 dams in four partial factorial subsets (15x5, 14x5, 15x5 and 15x5). Pop B (1737 individuals) was generated from 39 sires and 14 dams mated in six factorial subsets (6x3, 6x1, 6x3, 7x2, 7x3, 7x2). Four experimental backcross full-sib families totalizing 2500 individuals were also produced by in vitro fertilization at Ifremer (Palavas-les-Flots, France). In a previous study (Doan et al., 2017a), 60 sires from four different geographic origins (north-eastern, south-eastern and western Mediterranean Sea as well as Atlantic) were mated with nine west Mediterranean dams in a full factorial design (G0). Part of the offspring (G1) was challenged for VNN resistance and the estimated breeding values were estimated for the entire population, including the parents and the sibs of the challenged fish that were the candidates for selection. From these non-challenged candidates, four males were selected from families in which one parent was highly resistant and the other highly susceptible to VNN, and those were from three different geographic origins (north-eastern, south-eastern and western Mediterranean Sea). Those males were mated with four western Mediterranean females that we expected to be more susceptible due to their western-Mediterranean origin, in order to produce four full-sibs backcross families (G2). The G2 families were then named according to their grand-sire origin, NEM10, NEM12, SEM8 and WEM18 (NEM = north-east Mediterranean Sea, SEM = south-east Mediterranean Sea and WEM = western Mediterranean Sea).

Tableau 3.1: Study design and sampling strategy applied for each of the six sea bass populations challenged for VNN and the additional one used to create the genetic map.

	pop A	pop B	NEM10	NEM12	SEM8	WEM18	additional population
number of individuals produced	1680	1737			2500		880
number of parents (sires / dams)	59 / 20	39 / 14	1 / 1	1 / 1	1 / 1	1 / 1	94 / 39
number of individuals challenged	1350	1212			1719		
average survival rate	45.2	59.7	37.8	67.2	53.6	78.7	
number of individuals genotyped	1152	1152			1536		880
number of markers retained for the creation of the genetic map					51179		51179
number of individuals retained after quality control	1089	476	378	454	291	211	
number of markers retained after quality control	40623	41166	30917	23592	30656	31490	
type of analyses	GBLUP BayesCpi		composite interval mapping				building of the genetic map

3.1.3.3. Challenge VNN

All fish were challenged to VNN at the SYSAAF-ANSES Fortior Genetics platform (ANSES, Plouzané, France). For all experiments, fish were maintained in filtered seawater at a temperature of $27^{\circ}\text{C} \pm 2^{\circ}\text{C}$ in an open circuit. Pop A, pop B and backcross families infectious challenges were performed separately but in the same way. All the fish sent to the infection challenge were individually tagged. A total of 1680 from pop A (25g mean weight), 1737 from pop B (20g mean weight) and 2500 fish from the backcross families (8g mean weight) were received and acclimated for a minimum period of three weeks. Then, the whole batch of each population was split into a pre-test batch (120, 150 and 500 individuals for pop A, pop B and backcrosses, respectively), a challenge batch (1350, 1212 and 1719 individuals for pop A, pop B and backcrosses, respectively) and a negative control batch (120, 275 and 150 fish from the corresponding populations). For the pre-test and challenge batches contamination, the fish were immersed for 2 to 3 hours in a static bath of aerated seawater containing 1×10^5 TCID₅₀/ml of the W80 strain of RGNNV (redspotted grouper nervous necrosis virus), that was previously produced on a striped snakehead SSN1 cell line (Thiéry et al., 2004). Then, the batch was split into two tanks, water circulation was restarted and the mortality was recorded during 16 days. The mortality rate was 44% and 51% in the pop A pre-test, 51% and 53% in the pop B pre-test and 22.6% and 23.7% for backcross families. For the infectious challenges, a protocol similar to the pre-test was applied on the challenge batches. The negative controls were immersed for three hours into static seawater containing sterile cell culture medium. The mortality was recorded each day during the challenge period that was 27 days for pop A, 42 days for pop B and 33 days for the backcross families. Bacteriologic and virologic analyses were performed before the pre-test and challenges as well as during the mortality peak to check the sanitary status of the fish. Virologic analyses were done by

injecting a homogenized mixture of eye and brain sampled on random dead fish on SSN1 cells. Then, if cytopathic effects were observed, a virus identification was performed by immunofluorescence using anti-NNV antibodies. For bacteriological analyses, spleen and kidney are randomly sampled and used for bacterial cultures followed by MALDI-TOF identification.

3.1.3.4. Création de la puce à SNP pour le bar européen DlabCHIP 57K

The design of the high-density SNP array was based on the selection of high-quality variants from a database of ~2.6 million SNPs identified through whole-genome resequencing of 8 parents-offspring trios generated by experimental crossing of wild sea bass in Duranton et al. (2018). From this database, we excluded rare variants using a minor allele count threshold of 4 over 16 diploid parental genomes, and only retained SNPs located more than 35bp away from another known variant in order to ensure a high probe specificity. The final set of variants were chosen to cover the whole genome (including ungrouped scaffolds), but with a variable density depending on the estimated local nucleotide diversity (π) reported by Tine et al. (2014). This strategy aimed at increasing the density of SNPs within chromosome regions displaying a higher recombination rate, in order to homogenize recombination distances between two consecutive markers on the map. A list of 57,907 selected markers was submitted to ThermoFisher to develop the Axiom™ Sea Bass 57k SNP DlabChip array.

3.1.3.5. Génotypage et Assignation de parenté

All individuals were genotyped with the ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip at the genotyping platform Gentyane (INRAE, Clermont-Ferrand, France). A total of 1152 individuals were genotyped in each commercial cohort and 1536 fish in total were also genotyped in the experimental backcross families. In pop A and pop B, the genotyped individuals were randomly selected from the challenged ones among the dead and the surviving individuals. In the backcross families, they were selected to have the same average mortality per family as the whole challenged family. SNP calling was done using ThermoFisher software AxiomAnalysisSuite™. Preliminary quality controls were applied with threshold values of 95% for SNP call rate and 90% for sample call rate. Parentage assignment was done using 1000 randomly sampled markers with the R package APIS (Griot et al., 2020) with a positive assignment error rate set to 1%.

3.1.3.6. Création de la carte génétique

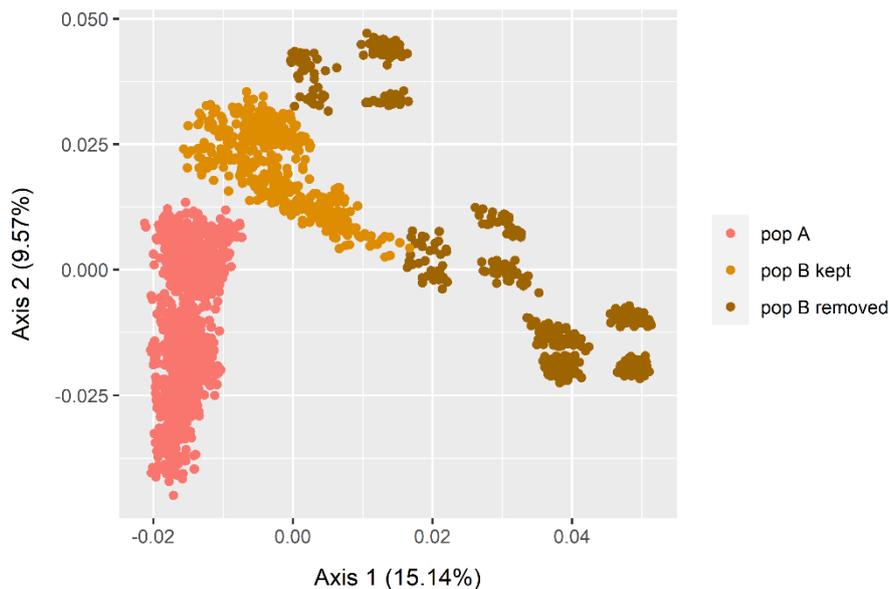
A genetic map was constructed with LepMap3 (Rastas, 2017). Backcross families were merged with another Ifremer sea bass dataset composed of the ThermoFisher Axiom™ Sea Bass 57k SNP DlabChip genotypes of 880 individuals from 94 sires mated with 39 dams in three partial factorial designs. The resulting data set included 2232 individuals genotyped at 51179 markers.

We ran the recommended procedure of LepMap3 with custom settings: 1% segregation distortion for Filtering2, a LOD score of 50 as well as a subset of 25% of the markers for SeparateChromosomes2, and LOD score of 30 for JoinSingles2All.

3.1.3.7. Contrôles qualité des génotypes et des échantillons

For each commercial cohort, the SNPs retained for further analysis had a minor allele frequency (MAF) above 5% and a p-value for the Hardy-Weinberg equilibrium test above a threshold of 10⁻⁸. From the

remaining markers, we subset the ones mapped on the genetic map. After quality controls, 1089 individuals genotyped for 40623 markers were kept for pop A and 1110 individuals genotyped for 41166 markers for pop B. Prior to GWAS analysis, which is known to be highly sensitive to population structures (Hayes, 2013; Pritchard et al., 2000), we performed a principal component analysis (PCA) based on the SNP genotypes, on both commercial cohorts, using PLINK 1.9 (Purcell and Chang, 2015). A strong within population structure was observed for pop B, and consequently we discarded individuals that belonged to minor groups which were distant from the main one (Supplementary Figure 3.1), leading to only 476 individuals retained for the genetic analysis of pop B. All individuals were kept in pop A.



Supplementary Figure 3.1 : PCA with 44772 markers on pop A and pop B cohort. In pop B, individuals in dark orange were removed from the data set and the ones in orange were kept.

In backcross families, we retained only SNPs with a MAF above 5% that were mapped on the genetic map. We obtained 30917, 23592, 30656 and 31490 informative markers for NEM10, NEM12, SEM8 and WEM18 respectively, with a sample size of 378 in NEM10, 454 in NEM12, 291 in SEM8 and 211 in WEM18 family.

Finally, within each cohort, the few missing genotype data were imputed using Fimpute software (Sargolzaei et al., 2014) to obtain complete genotypes for GWAS analysis.

3.1.3.8. Estimation de l'héritabilité

For each data set, we estimated the heritability of VNN resistance under a threshold model using THRGIBBSF90 (Tsurata and Misztal, 2006) and a linear model using AIREMLF90 (Misztal et al., 2002), both from the blupf90 program suite. Only individuals with a phenotype, a genotype and a pedigree were used in heritability estimates, thus the sample size was 1027 in pop A and 476 in pop B. Heritability was also estimated within each of the backcross families, as reliable estimation of heritability within a single family using genomic information was shown to be accurate by Ødegård and Meuwissen (2012).

The following model was computed in each cohort using both threshold and linear models:

$$y = 1b + Zu + e$$

With y the vector of the phenotypes measured as binary dead/survival trait, 1 the incidence (unity) vector of the intercept, b the estimate of the intercept effect, u the vector of breeding values and Z the corresponding incidence matrix. It is assumed that u follows a multivariate normal distribution $N(0, G\sigma_g^2)$ with G the genomic relationship matrix proposed by VanRaden (2008) and σ_g^2 is the additive genetic variance. e is the vector of the random residual errors that follows a normal distribution $N(0, I\sigma_e^2)$ with σ_e^2 the residual variance and I the identity matrix.

With the threshold model, the variance components (σ_g^2 and σ_e^2) were estimated using a Gibbs sampler with 500,000 iterations, 100,000 of burn-in and one sample was kept every 20 iterations for posterior analysis. The posterior distributions were analyzed with the R package *boa* (Smith, 2007). With the linear model, the same components were estimated using a restricted maximum likelihood algorithm, considering the observed binary phenotype as a continuous variable.

The heritability for survival was estimated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Heritability on the observed scale (h_o^2) was estimated using the variance components from the linear model, while the heritability on the underlying liability scale (h_u^2) was computed using the variance components from the threshold model.

3.1.3.9. Détection de QTL dans les populations expérimentales

In the experimental backcross families, we used a composite interval mapping approach from the R package *qtl* (Broman, 2001). The LOD score under the hypothesis of the absence of a QTL on the chromosome was computed for each interval between two consecutive markers of each chromosome using a Haley-Knott regression (Haley and Knott, 1992). Genome-wide significance LOD thresholds were estimated by permutation tests considering 1000 permutations (Churchill and Doerge, 1994).

3.1.3.10. Détection de QTL dans les cohortes commerciales

We performed GWAS under multi-marker linear regression models using GBLUP and Bayesian approaches.

3.1.3.10.1. GWAS avec un modèle GBLUP

We used the *blupf90* suite of programs to perform GWAS by GBLUP analysis for VNN resistance in both commercial data sets. The breeding values were estimated with BLUPF90 using the following linear model:

$$y = 1b + Zu + e$$

With the same notation as in section 2.8. The p-values were computed using POSTGSF90 (Aguilar et al., 2019). The $-\log_{10}$ of the p-values were compared to the chromosome-wide significance threshold and to the genome-wide significance threshold at 5% after Bonferroni correction for the average number of markers per chromosome and the total number of markers, respectively.

3.1.3.10.2. GWAS avec un modèle bayésien

Because GBLUP is known to shrink SNP effects towards 0, we also used a Bayesian variable selection model with a BayesC π approach (Habier et al., 2011) to refine QTL positions as well as to estimate the proportions of genetic variance explained by the QTLs. In this model, a proportion π of the markers is assumed to have a non-zero effect. The marker effects are estimated with a mixture of a proportion π of markers with effects following a normal distribution $N(0, \sigma_a^2)$ and a proportion $1 - \pi$ of markers with a zero effect. σ_a^2 is the part of total genetic variance explained by the SNP markers. In addition to the SNP effect described as above, we added a random polygenic genetic effect to account for the genetic variation that could not be captured by the markers (Legarra et al., 2008; Solberg et al., 2009). This model is defined as follows:

$$y = 1b + Zu + Wg + e$$

With y the vector of the phenotypes measured as binary dead/survival trait, 1 the incidence (unity) vector of the intercept, b the estimate of the intercept effect, u the vector of the polygenic effects, Z the corresponding incidence matrix, g the vector of the SNP random effects, W the corresponding incidence matrix and e the vector of the random residual errors. The total genetic variance was decomposed into one part explained by the polygenic effect and one part explained by the marker effects. As initial priors in pop A, the part of genetic variance explained by the polygenic effect represented 95% of the total genetic variance and the markers genetic variance represented 5% of the total genetic variance.

$$\sigma_g^2 = 0.95\sigma_u^2 + 0.05\sigma_a^2$$

Where σ_g^2 is the total genetic variance, σ_u^2 is the genetic variance explained by the polygenic effect and σ_a^2 is the genetic variance explained by SNP markers. In pop B, the small sample size led to a share of polygenic and SNP variance that remained quasi-identical to the priors, whatever they were, thus the total genetic variance was split between polygenic and SNP variance in the priors with the same proportions (57 and 43 %, respectively) obtained in pop A after 200,000 iterations of Gibbs sampling.

σ_u^2 and σ_e^2 were sampled from a scaled inverse chi-square distribution. The initial values for residual variance and total genetic variance priors were the variance component estimates from the linear model described in section 2.8.

In pop A, the degrees of freedom of both parameter distributions were set to 5. In pop B, as the number of individuals was small and to keep the sampling values of σ_e^2 within a reasonable range, the degrees of freedom of σ_u^2 were set at 5 and the ones of σ_e^2 were set at 10,000 to put a strong degree on belief on its prior and ensure the convergence.

The BESSiE software was used to compute this model (Boerner and Tier, 2016). A total of 200,000 iterations of Gibbs sampling were performed with a burn-in of 10,000 iterations. One Gibbs sample was kept every 20 iterations for further analysis. For every iteration, the proportion $1 - \pi$ was sampled in a beta distribution $B(\alpha, \beta)$. α was set as the total number of makers in each cohort and β was set at 40.

The degree of confidence in the association between the phenotypes and each SNP was computed using the Bayes Factor (Kass and Raftery, 1995) calculated as:

$$BF_i = \frac{P_i/1 - P_i}{\pi/1 - \pi}$$

where P_i is the probability of the SNP i to have a non-zero effect and π is the proportion of markers with a non-zero effect. BF was transformed into $\log BF = 2\log(BF)$ to produce values within the same range as $-\log(p\text{-values})$. Strong evidence for the existence of a QTL was considered when at least one SNP had a $\log BF$ greater than 8 according to Michenet et al. (2016).

As the causative mutation may not be the marker with the highest $\log BF$ (hereafter named the peak SNP), the definition of QTL region was done following the method described in (Michenet et al., 2016). All markers close to the peak SNP and having a $\log BF$ greater than 3 were considered to be in the QTL region. For each chromosome, the algorithm started at the peak SNP. Then, in a sliding window of 0.5 cM starting from the peak SNP, every SNP with a $\log BF$ greater than the noise detection threshold of 3 was included in the QTL region. A sliding window was applied until no SNP had a $\log BF$ greater than 3 in the current window. The border of the QTL region was then defined by the last SNP included in the window. This procedure was applied on both sides of the peak SNP, which leads to a QTL region defined as a credibility interval for the causative mutation.

The genetic variance explained by one SNP was calculated as:

$$\sigma_{SNP}^2 = 2p(1 - p)a^2$$

With p the MAF of the SNP and a the effect of the SNP. To obtain the proportion of the genetic variance explained by one SNP, σ_{SNP}^2 was divided by σ_g^2 . The proportion of the genetic variance explained by the QTL region was the sum of the proportions explained by each SNP located in the QTL region as previously defined.

3.1.4. Résultats

3.1.4.1. Challenge VNN

The three challenges were conducted up to 42 days. VNN presence was confirmed by virologic analyses on several subset of fish dead during the infection kinetic, in absence of significant bacterial coinfection. Survival rates ranging from 37.8% to 78.7% were recorded (Figure 3.1). The peak of mortality was around 10 days after infection.

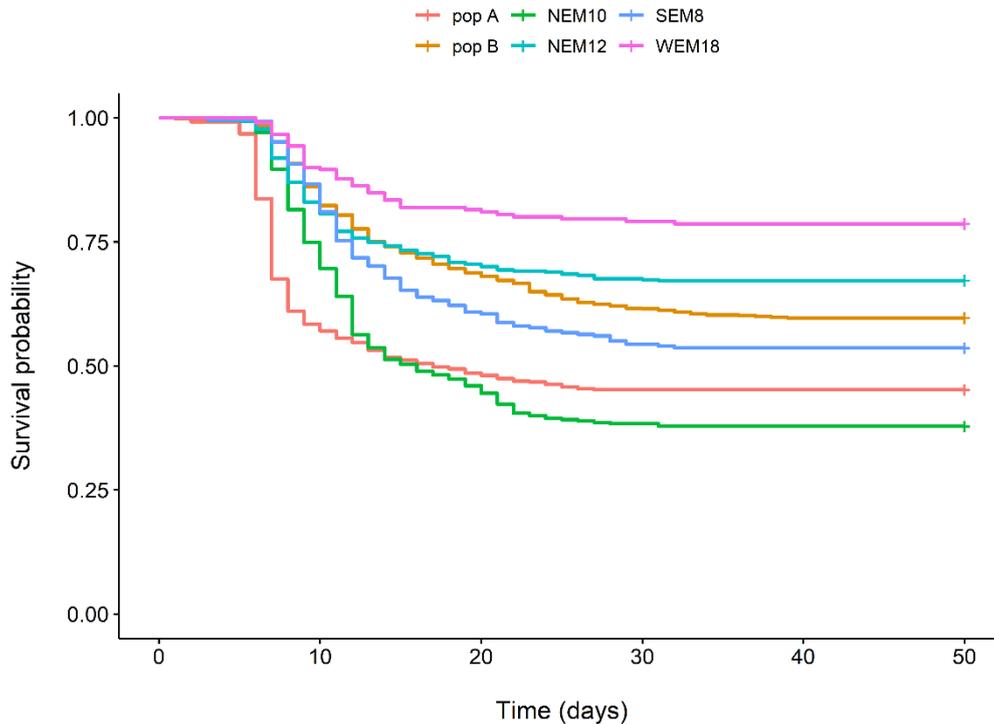


Figure 3.1 : Evolution of the Kaplan-Meier probability of survival of six different European sea bass populations during their respective VNN infection challenge. Pop A and pop B are commercial populations, and NEM10, NEM 12, SEM8 and WEM 18 are backcross families.

3.1.4.2. Performance de la puce à SNP Axiom DlabCHIP dans les populations commerciales

Among submitted SNPs, 56,730 markers were spotted on the SNP array after ThermoFisher internal selection procedure, among which 537 markers were duplicated. Genotyping commercial populations with the new 57K SNP array revealed a large number of polymorphic SNPs with high clustering resolution (PolyHighResolution, MonoHighResolution and NoMinorHom categories) respectively 50,186, 675 and 2,400 (93.1%) in pop A and 51,686, 252 and 2,045 (94.4%) in pop B. For passing samples, the average genotype call rate was >99.6% in pop A and >99.8% in pop B. The minor allele frequency (MAF) distribution for PolyHighResolution SNPs was similar in both populations (Supplementary Table 3.1) with an average frequency of 0.28 for pop A and 0.29 for pop B.

Supplementary Table 3.1 : Number of PolyHighResolution markers per population per categories of Minor Allele Frequency (MAF).

categories of minor allele frequency	pop A	pop B
	PolyHighResolution (%)	PolyHighResolution (%)
MAF < 0.1	3777 (6.7%)	3082 (5.4%)
0.1 ≤ MAF < 0.2	9577 (16.9%)	9756 (17.2%)
0.2 ≤ MAF < 0.3	11680 (20.6%)	12610 (22.2%)
0.3 ≤ MAF < 0.4	12500 (22.0%)	13084 (23.1%)
MAF ≥ 0.4	12584 (22.2%)	13060 (23.0%)

3.1.4.3. Estimation de l'héritabilité

The estimates of genomic heritability using linear and threshold models are summarised in Tableau 3.2. Estimates of the genomic heritability of the death/survival binary trait on the observed scale (h^2_o) were 0.23 (± 0.05) in pop A and 0.08 (± 0.09) in pop B. In backcross families, h^2_o estimates were 0.38 (± 0.09) in the NEM10 family, 0.59 (± 0.07) in the NEM12 family, 0.50 (± 0.09) in the SEM8 family and 0.23 (± 0.15) in the WEM18 family. The heritabilities estimated with a threshold model on the underlying liability scale (h^2_u) were higher than h^2_o estimates, as expected. Estimates of h^2_o obtained by transforming h^2_u with the formula by (Dempster and Lerner, 1950) were similar than the ones obtained with the threshold model in pop A and the WEM18 family, smaller in pop B and higher in the NEM10, NEM12 and SEM8 families (data not shown).

Tableau 3.2 : Variance components and genetics parameters for VNN resistance in European sea bass estimated with three models in six different populations. h^2_o is the heritability on the observed (binary) scale and h^2_u is the heritability estimate on the underlying liability scale.

population	linear model			threshold model			BayesCpi			
	Vg	Ve	h2	Vg	Ve *	h2	Vpoly	Vsnp	Ve	h2
pop A	0.057	0.189	0.231 (± 0.049)	0.627	1.086	0.377 (± 0.065)	0.029	0.039	0.182	0.272
pop B	0.019	0.223	0.078 (± 0.085)	0.404	1.016	0.244 (± 0.140)	0.010	0.012	0.223	0.090
NEM10	0.111	0.180	0.381 (± 0.089)	1.538	1.018	0.584 (± 0.088)				
NEM12	0.187	0.130	0.591 (± 0.067)	5.711	1.013	0.838 (± 0.044)				
SEM8	0.164	0.165	0.499 (± 0.094)	2.328	1.023	0.677 (± 0.078)				
WEM18	0.043	0.147	0.227 (± 0.147)	1.164	1.028	0.478 (± 0.170)				

* : In threshold models, σ^2_e is constrained to a value close to 1

3.1.4.4. Construction de la carte génétique et détection de QTL dans les familles de rétro-croisement

We obtained a new high-density genetic map containing 49638 markers that were homogeneously mapped on 24 linkage groups (LG) corresponding to the 24 known chromosomes in the species' karyotype. The total length of the genetic map was 1873.1 cM, corresponding to a density of 26.5 markers per cM.

In the composite interval mapping analysis done in the backcross families, LG12 had a very high LOD score compared to other chromosomes in the NEM10, NEM12 and SEM8 families (Figure 3.2). At 5% LOD threshold, 725, 322 and 840 markers were detected as potential QTLs on LG12 (out of a total number of 1273, 925 and 1282 on this chromosome for NEM10, NEM12 and SEM8 respectively). In the WEM18 family, no QTL was detected. Additionally, in the NEM12 family, 4 markers of LG8 (out of 925 in this chromosome) were detected as potential QTLs.

3.1.4.5. Détection de QTL avec une approche GBLUP dans les cohortes commerciales

In the GWAS performed by GBLUP analysis, the p-values of 20 markers on LG12 exceeded the chromosome-wide significance threshold in pop A and, among them, 6 exceeded the genome-wide significance threshold (Figure 3.2). One additional marker located on LG8 exceeded the chromosome-wide significance threshold. In pop B, one marker on LG12 and one marker on LG15 exceeded the chromosome-wide significance threshold (Figure 3.2).

3.1.4.6. Détection de QTL avec une approche BayesC π dans les cohortes commerciales

The estimation of the share of genetic variance explained by the polygenic effect represented 57% of the total genetic variance in pop A and 55% in pop B (Tableau 3.2). From the BayesC π model, a total of 5 QTLs were detected with strong evidence ($\log\text{BF}>8$) in pop A and were located on LG3, LG8, LG12, LG14 and LG19 (Figure 3.3). The QTL on LG12 was located between 31.71 and 35.16 cM and explained 9.21% of the total genetic variance. Other QTLs explained less than 1% of the total genetic variance each and their locations are summarized in Tableau 3.3. QTLs on LG3 and LG19 in pop A were single marker QTLs. In pop B, three QTLs located on LG12, LG15 and LG20 were detected. They all explained 1% of the total genetic variance or less. The QTL on LG12 was located in a smaller confidence interval than the one in pop A, between 33.26 and 33.91 cM.

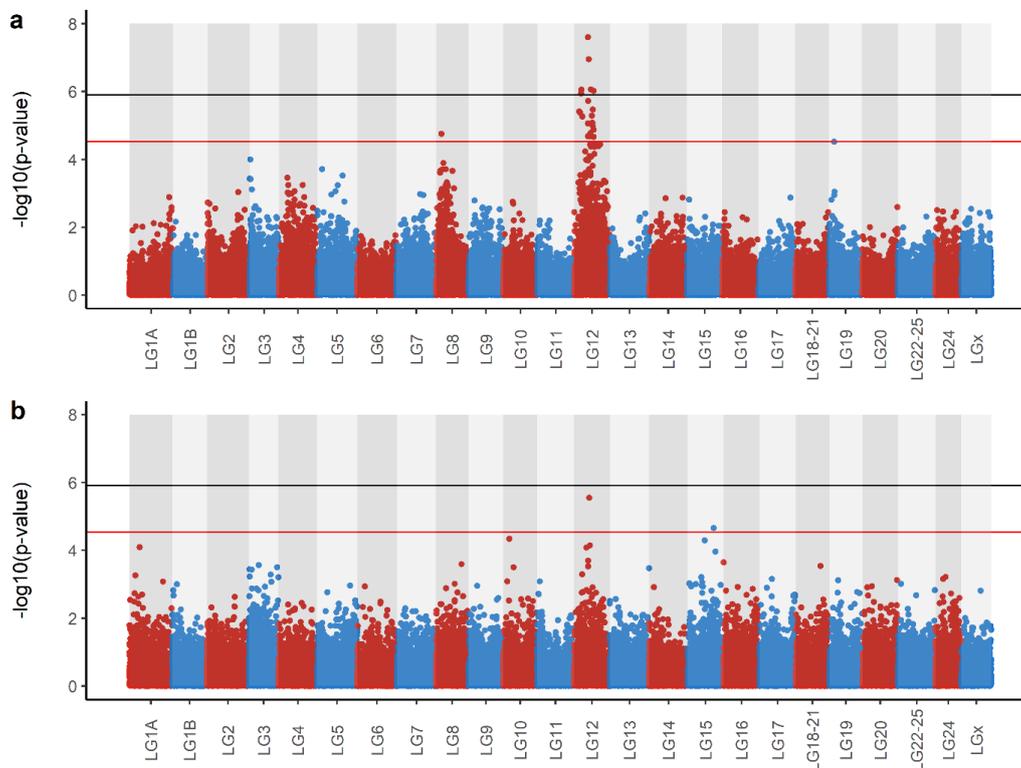


Figure 3.2 : Manhattan plot of $-\log_{10}(\text{p-value})$ obtained from GWAS for VNN resistance QTLs in commercial European sea bass populations pop A (a) and pop B (b). The horizontal black line represents the genome-wide significance threshold and the red line the chromosome-wide significance threshold calculated with the Bonferroni correction.

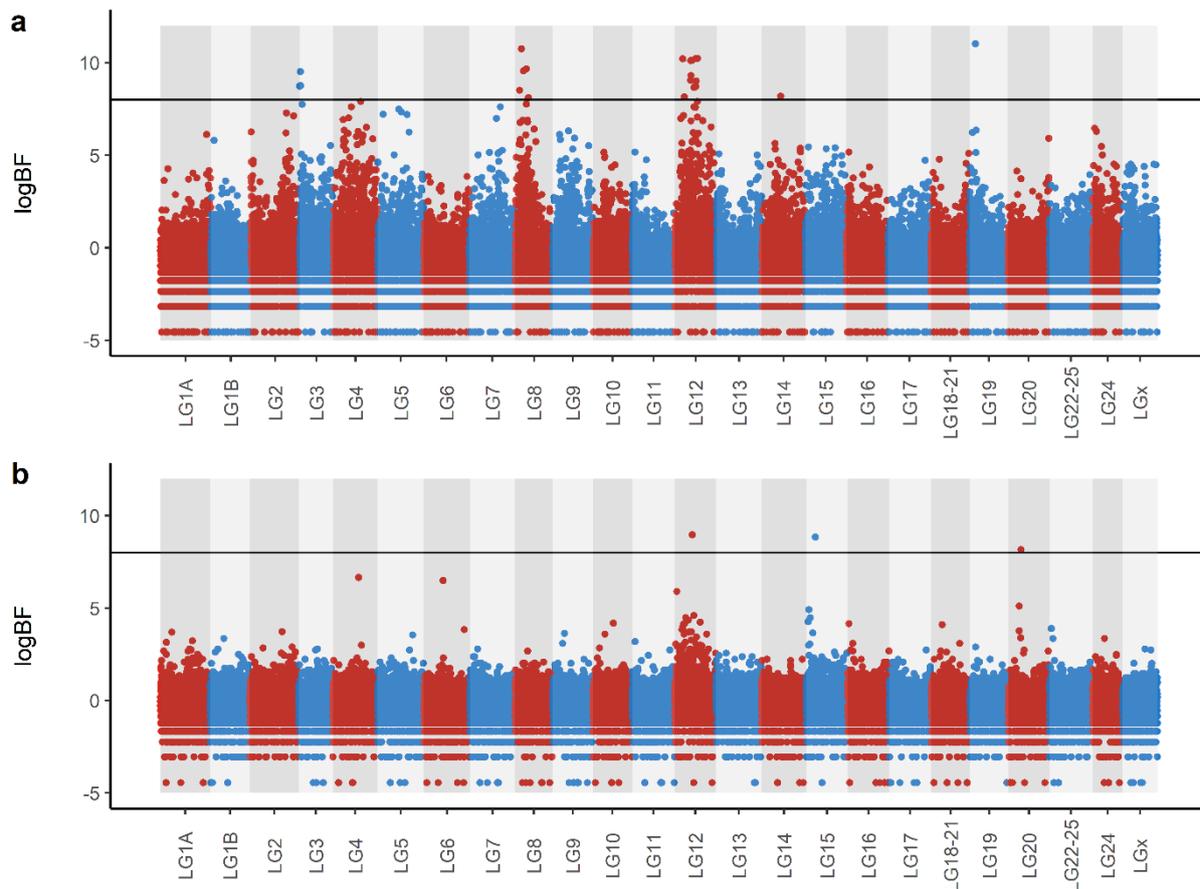


Figure 3.3 : Genome-wide logBF plot for VNN resistance across the genome in the European sea bass populations pop A (a) and pop B (b) using a BayesC π model. Horizontal black lines represent the logBF threshold of 8, corresponding to strong evidence for the presence of a QTL.

Tableau 3.3 : QTL detection by BayesC π for resistance to VNN in two commercial populations of European sea bass : position of the peak SNPs, credibility intervals and proportions of genetic variance explained.

population	chromosome	peak SNP position (cM)	start QTL position (cM)	end QTL position (cM)	share of total genetic variance explained (%)
pop A	LG3	5.89	5.89	5.89	0.45
	LG8	23.65	23.27	23.65	1.10
	LG12	34.47	31.71	35.16	9.21
	LG14	55.93	55.68	55.93	0.29
	LG19	22.24	22.24	22.24	1.04
pop B	LG12	33.64	33.26	33.91	1.09
	LG15	57.94	57.94	57.94	0.50
	LG20	41.58	41.51	41.58	0.39

3.1.5. Discussion

In this study, we designed and used a high-density SNP array specifically developed for mapping studies in the European sea bass. The average physical distance (<12kb) and genetic distance (<0.5cM) between consecutive markers make this array an excellent tool for implementing cost-effective screening of QTLs and genomic selection. The high call rate (>99.6%) and the high number of polymorphic SNPs in the studied populations (>88%) also validate the SNP selection strategy, that permitted to cover the whole genome with high-quality variants, while avoiding technically undesignable SNPs. We used the SNP array to construct a high-density linkage map for composite interval QTL mapping in the backcross families. The accuracy and level of resolution of this new genetic map exceeds that of previous map built in the European sea bass (Palaiokostas et al., 2015).

Using different types of populations (four large backcross families and two commercial admixed populations) and complementary analytical strategies (Composite Interval Mapping, as well as GBLUP and BayesC π based GWAS), we were able to detect QTLs involved in VNN resistance, the major viral threat to European sea bass aquaculture (Vendramin et al., 2016). While the commercial cohorts were used to estimate the genetic parameters and to detect potential QTL for direct improvement in ongoing breeding programs, the backcross families constitute classical QTL mapping populations. We confirmed that VNN resistance in European sea bass has a moderate heritability in the commercial populations (0.24 to 0.38 using a threshold model), similar to the previously reported estimates (0.26 to 0.43; (Doan et al., 2017a; Palaiokostas et al., 2018a)). In the backcross families, the heritability estimates were much higher (0.48 to 0.84). Even though the genomic relationship matrix enables the estimation of genetic parameters in full-sibs families, the estimates can be inaccurate when the family size is too small and when QTLs segregate within the families (Ødegård and Meuwissen, 2012). Except in pop A where the heritability estimated with a threshold model was similar to the one using a linear model and corrected with the Dempster and Lerner formula, all the heritabilities estimated using a threshold model were very different the ones estimated with a linear model and corrected with the Dempster and Lerner formula. In the backcross families, the heritability estimation was challenging due to the single-family structure as well as the small number of individuals. In pop B, it could be because of the small number of individuals, which causes high sampling variance in heritability estimates.

Regardless of the detection method, one QTL was detected on the LG12 in five of the six data sets. Its effect was strong, explaining up to 9.21% of the total genetic variance in the commercial population pop A. The combination of the different results reduced the confidence interval of the QTL to a likely position between 33.26 and 33.91 cM, equivalent to 3.7 Mb and containing 125 of our SNP markers, corresponding to the intersection of all the confidence intervals. From the results we obtained, we cannot yet reach a putative causal mutation and thus, the best markers to use in MAS to predict the phenotype. Another cross-population QTL was detected on LG8, shared by pop A and the NEM12 family. This QTL explained 1.1% of the total genetic variance in pop A. Other QTLs were population-specific as they were only detected in one population.

A previous study led to the detection of minor effect QTLs (Palaiokostas et al., 2018a) located on chromosome 3, 20 and 25 explaining 4%, 1.5% and 2% of the total genetic variance respectively. After chromosome correspondence checking by aligning markers sequences on the European sea bass reference genome (GCA_000689215.1) using BLAST (Altschul et al., 1990), chromosome 3 and 20 from that previous study corresponded to the LG12 and LG6 in our study. The QTL explaining the largest part of genetic variance was located on the same chromosome (LG12) in both studies. In Asian seabass, only minor effect QTLs were found, explaining between 1.6% and 2.7% of the phenotypic variance (Wang et al., 2017). Those QTLs, located on chromosomes 1, 8, 14, 15, 16, 19, 20, 21 and 24 were

linked with potential candidate genes, all involved in stress response. However, using syntenic blocks proposed by Vij et al. (2016), we found that no homologous genomic regions were in common between QTLs detected in European seabass and Asian seabass. In Atlantic cod, five QTLs were detected on LG1, LG6, LG18, LG19 and LG20, explaining up to 19.7% of the phenotypic variance (Baranski et al., 2010). As VNN resistance in Atlantic cod has a high heritability (0.75), the QTLs explaining 14.2%, 18.2% and 19.7% of the phenotypic variance could be considered as strong effect QTLs.

In back-cross full-sib families, we used a Composite Interval mapping approach that is the most appropriate for such data. From those analyses, we obtained very high LOD values for the detection of the QTL on LG12 as well as very wide confidence interval (37.8 cM in NEM10, 9.74 cM in NEM12 and 42.3 cM in SEM8) mainly due to strong linkage-disequilibrium between markers. No QTL was found in the WEM18 even though it was the family with the highest survival rate. This absence of the QTL detection may be due to the fact that the WEM18 family did not segregate for resistance QTLs. The mating that produced those families was designed to produce families that segregate for resistance QTLs (section 2.2), by mating putative Rs (R : resistant ; s : susceptible) sires with putative ss dams. Thus, it is likely that each parent of this particular WEM18 family carried the resistance QTLs in only one version. As the global resistance of the family was high, it is rather likely that the WEM18 family could be from a RR sire x ss dam or from a RR sire x RR dam, thus not segregating for the QTL on LG12.

In commercial data sets, we used GBLUP and Bayesian GWAS. In the GBLUP analysis, only the QTL on LG12 exceeded the genome-wide significance threshold in pop A. In pop A and B, the QTL on LG12 exceeded the chromosome-wide significance threshold, as did one marker on LG19 in pop A. In GWAS by BayesC π , more QTLs were detected due to the variable selection process that increases the power of detection (Tam et al., 2019). In addition, as we had very different number of phenotypes in commercial cohorts, BayesC π allowed us to test different priors to fit the best model for each data set.

Thanks to the high number of individuals in pop A and thus, the high power of the QTL detection, we were able to refine the location of the QTL on LG12. Both GBLUP and BayesC π approaches revealed a strong association between VNN resistance and markers located on LG12 as well as a moderate part of the total genetic variance explained by this QTL. Even though the QTL on LG12 explained 9.21% of the total genetic variance, several markers located on the chromosome and detected in pop A seemed to have a potential interest in MAS. Among them, one marker (LG12_8815613) detected had the highest effect on that chromosome and the third highest logBF on LG12. In addition, it had a low MAF (0.13), which is expected when resistance is not well spread in a population, and which also mechanically decreases the percentage of variance explained. Thus, this marker had interesting properties to be a candidate marker to MAS. Its effect on survival across data sets is shown in Figure 3.4. Interestingly, this marker had a strong effect on survival both in commercial populations (Figure 3.4) and in the backcross families, with an average of 43.3% survival for genotype AA, 72.8% survival for genotype AB and 78.3% survival for genotype BB. In pop B only, the BB genotyped survived less than the AB genotype, but only five individuals were BB in pop B, and thus the condition to quantify the effect of the BB genotype in this population are far from optimal. Although the consistent effect of this marker on six different cohorts highlights its potential interest, further validation in additional populations or cohorts from the same population would be of great value. The low observed frequency for allele B in the commercial populations could be due to a founder effect or maybe from the fact that these populations have been selected for several traits, among which growth rate for at least 6 generations. As there is a moderate but negative genetic correlation of VNN resistance with body weight (-0.35 ± 0.14 ; (Doan et al., 2017a)), this selection for growth might have had a negative effect on the frequency of a VNN resistance allele.

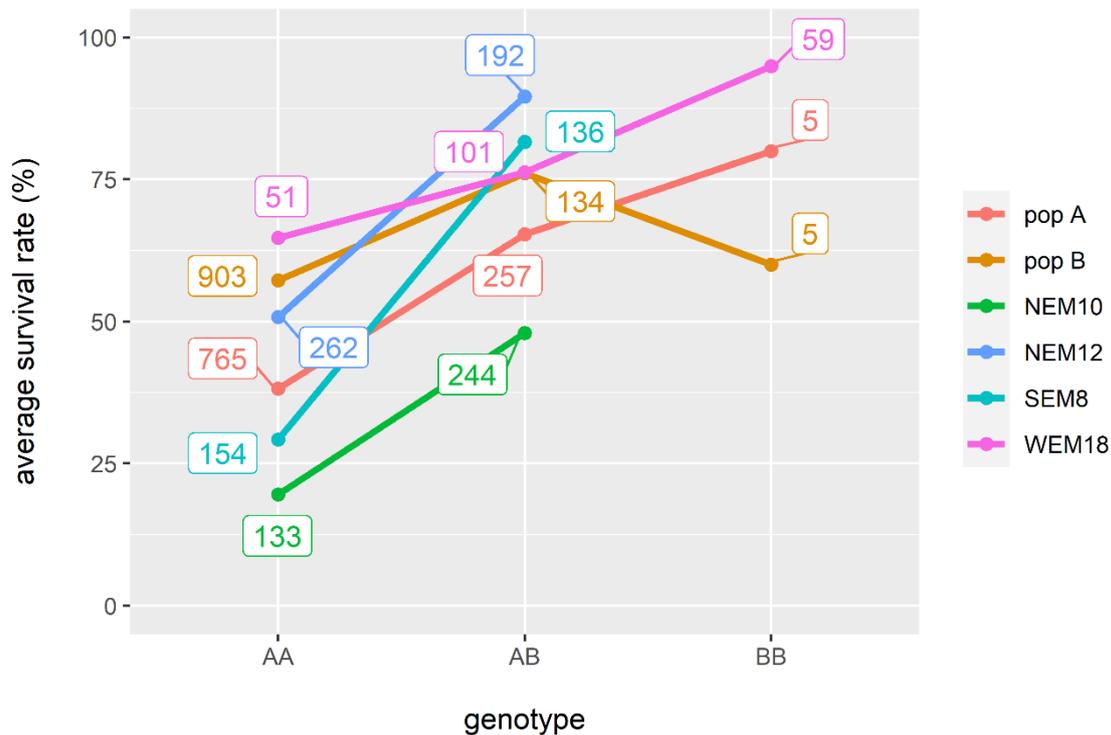


Figure 3.4 : Average survival rate (%) to VNN for each genotype at marker LG12_8815613 in six cohorts of European sea bass. Each colour corresponds to one data set. The numbers in the boxes are the numbers of individuals per genotype.

The QTL detection in pop B was more challenging. The strong population structure that we revealed with the PCA led to a strong decrease in the number of individuals that could be used in this data set and thus, to a decrease in the QTL detection power. We only managed to detect two QTLs at chromosome-wide level using GWAS by GBLUP, and three QTLs with Bayesian GWAS. Bayesian GWAS was very sensitive to the part of genetic variance explained by the markers given as priors. In this population, the proportions estimated after 200,000 iterations of Gibbs sampling was similar to the priors, contrary to pop A that estimated new values for those proportions. Thus, a larger number of individuals would be needed to accurately estimate those priors and improve the overall QTL detection. For example, in Korte and Farlow (2013), the authors showed that 800 individuals are necessary to detect a QTL explaining 5% of the phenotypic variance with a power of 0.8 when applying a false discovery rate of 5%. Here, we only had 476 usable individuals in the pop B cohort.

This study produced encouraging results with the discovery of one important QTL involved in VNN resistance in European sea bass. Even though the part of genetic variance explained by this QTL is limited (9.21%), further work on the location of the causal mutation and the discovery of diagnostic marker informing on the resistance genotype of individuals could greatly improve the global sea bass aquaculture. The best example of the application of MAS in aquaculture is the discovery of one major QTL involved in the Infectious pancreatic necrosis virus (IPNV) resistance in Atlantic salmon and explaining 90% of the genetic variance (Houston et al., 2008; Moen et al., 2009). Since its use by the salmon aquaculture industry to improve selection based on pedigree information, IPNV outbreaks occurrence decreased by 75% (Hjeltnes, 2014). This success was due to the identification of one SNP located on the gene responsible of the resistance and the genotyping of selection candidates to improve the IPNV resistance in aquaculture farms. In the context of VNN resistance in European sea bass, the effect of the QTL on LG12, although significant, is smaller. Its use in MAS will have less benefits due to the smaller part of genetic variance explained. In that situation, genomic selection could be

more appropriate to improve VNN resistance. As shown with the example of marker LG12_8815613 (Figure 3.4), it remains possible that a marker with very large effects on survival would only explain a moderate part of the genetic variance if the resistance allele is rare in the population studied. The presence of a major QTL explaining a large part of the genetic variance in VNN resistance is thus still a valid hypothesis and could be investigated in further work using genome sequence combined with fine mapping approach.

3.2. Validation du QTL de résistance à la nodavirose chez le bar

Dans Griot et al. (2021), nous avons montré qu'un QTL à effet fort est impliqué dans la résistance à la nodavirose chez le bar. Afin de confirmer ces résultats, le consortium du projet Gènesea a proposé de valider la présence de ce QTL dans d'autres cohortes commerciales provenant des mêmes entreprises de sélection.

3.2.1. Matériel et Méthodes

3.2.1.1. Cohortes de validation

Trois nouvelles cohortes en sélection commerciale de bar ont été utilisées lors de la validation du QTL. Elles étaient composées de 709, 536 et 1 260 individus respectivement et seront nommées VAL1, VAL2 et VAL3 dans la suite de ce chapitre.

Ces trois cohortes avaient été challengées au VNN au sein de la plateforme Fortior Genetics (ANSES, Plouzané, France) en suivant le même protocole que celui décrit en 2.1.3.3

3.2.1.2. Génotypage et assignation de parenté

Les trois cohortes ont été génotypées sur la puce ThermoFisher Axiom™ DlabChip 57K par la plateforme de génotypage Gentyane (INRAE, Clermont-Ferrand, France). Les génotypes ont été traités par le même protocole que celui décrit en 2.1.3.7. Après imputation et sélection des marqueurs présents sur la carte génétique, 40 855, 40 602 et 37 398 marqueurs ont été utilisés dans les cohortes VAL1, VAL2 et VAL3 respectivement.

3.2.1.3. Analyses génétiques

3.2.1.3.1. Détection de QTL intra-cohorte

Au sein de chaque cohorte, une détection de QTL avec un modèle GBLUP a été réalisée de la même manière que celle décrite en 3.1.3.10.1.

3.2.1.3.2. Détection de QTL inter-cohorte

Une détection de QTL avec un modèle GBLUP et BayesC π a été réalisée sur l'ensemble des cohortes où le QTL présente sur le LG12 a été localisé. Ainsi, les cohortes popA, popB (ayant servi à la détection de QTL dans Griot et al., 2021), VAL1 et VAL2 ont été agrégées en un jeu de données unique de 2 748

individus et seuls les 37 179 marqueurs en commun ont été conservés. En augmentant le nombre d'individus, la puissance de détection est augmentée, ce qui devrait permettre de préciser la position et la part de variance génétique expliquée par les QTLs qui seront détectés. Afin de prendre en compte l'origine de chaque cohorte ainsi que les potentielles différences entre les challenges, un effet fixe « challenge » a été ajouté dans le modèle.

Le modèle suivant a été établi pour la détection de QTL :

$$y = Xb + Zu + e$$

Avec y le vecteur des phénotypes mesurés comme un caractère binaire mort / survie au challenge, noté 0 pour les individus morts et 1 pour ceux qui ont survécu au challenge, X la matrice d'incidence de l'effet fixe, ici le challenge, b l'estimation de l'effet fixe, u le vecteur des valeurs génétiques et Z la matrice d'incidence correspondante. Ici, u suit une loi normale multivariée $N(0, G\sigma_g^2)$ avec G la matrice d'apparentement génomique proposée par VanRaden (2008) et σ_g^2 est la variance génétique additive. Le vecteur des erreurs résiduelles e suit une loi normale $N(0, I\sigma_e^2)$ avec σ_e^2 la variance résiduelle et I la matrice identité.

Ensuite, une GWAS avec un modèle BayesC π a été réalisée. Dans le modèle BayesC π , les paramètres α et β de la loi Beta permettant de définir la valeur de $1 - \pi$, correspondant à la proportion de marqueurs ayant un effet non nul, ont été fixés à 37 179 et 100 et respectivement. La variance génétique a été divisée en une partie expliquée par un effet polygénique, représentant 95% de la variance génétique totale, et une partie expliquée par l'effet des marqueurs, représentant 5% de la variance génétique totale. Les degrés de liberté des lois permettant d'échantillonner les valeurs de la variance de l'effet polygénique et de la variance résiduelle ont été fixés à 5.

3.2.1.4. Effet du marqueur LG12_8815613

Dans Griot et al. (2021), nous avons montré qu'un marqueur, nommé LG12_8815613, présentait de nombreuses caractéristiques permettant son utilisation dans une sélection assistée par marqueurs. L'effet de ce marqueur sur la survie a été estimé pour les cohortes VAL1, VAL2 et VAL3 en calculant la survie moyenne par génotype au marqueur. Ces effets ont été comparés à ceux observés dans les cohortes précédemment évaluées popA, popB, NEM10, NEM12, SEM8 et WEM18 (Griot et al., 2021)

3.2.2. Résultats

3.2.2.1. Détection de QTL intra-cohorte

Les résultats des GWAS intra-cohorte sont présentés Figure 3.5. Le QTL présent sur le LG12 a été confirmé dans les cohortes VAL1 au seuil génomique et VAL2 au seuil chromosomique (Figure 3.5a, Figure 3.5b). Dans la cohorte VAL3, trois QTLs ont été détectés sur le LG7, le LG9 et le LG14 (Figure 3.5c).

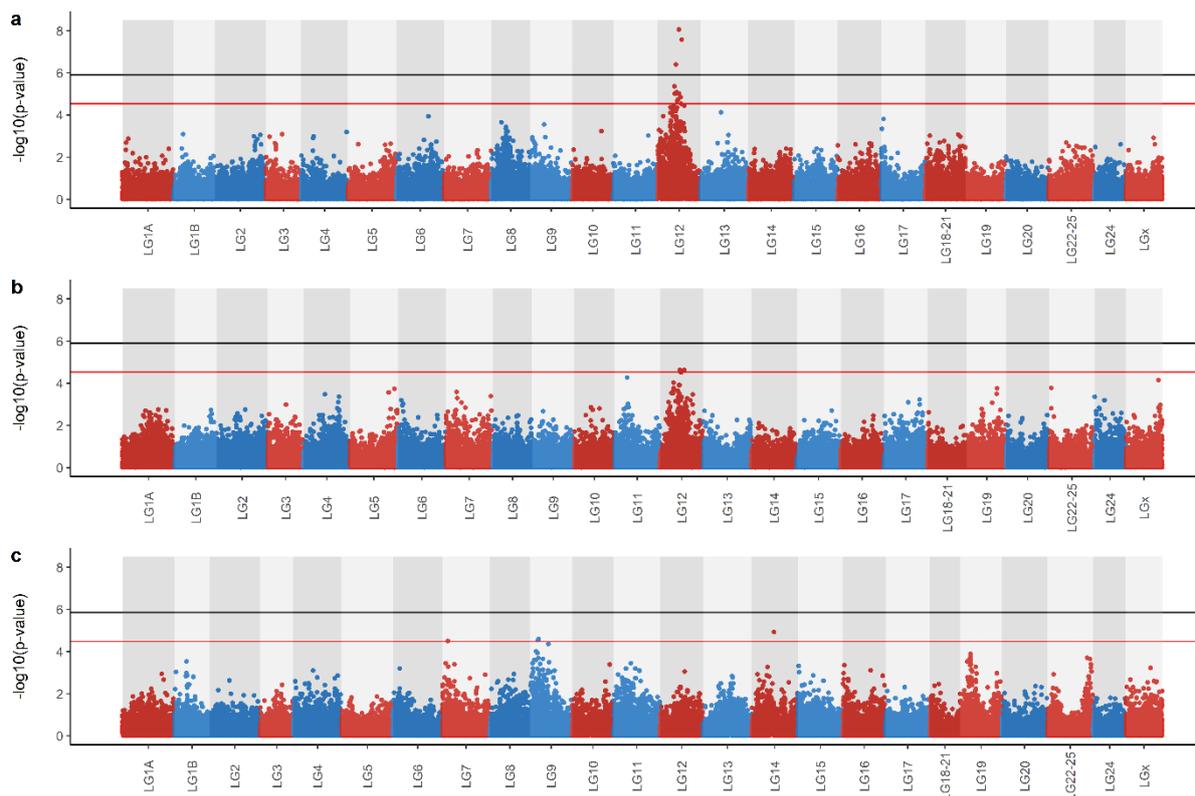


Figure 3.5 : Manhattan plot de $-\log_{10}(p\text{-value})$ obtenus à partir de GWAS pour détecter des QTLs de résistance à la nodaviruse chez le bar européen, réalisées avec un modèle GBLUP dans trois cohortes de 709 (a), 536 (b) et 1 260 (c) individus challengés à la nodaviruse. La ligne rouge représente le seuil de significativité chromosomique et la ligne noire, le seuil de significativité génomique, calculée avec une correction de Bonferroni.

3.2.2.2. Détection de QTL inter-cohorte

Au total, trois QTLs, présents sur le LG1A, LG8 et LG12 ont été détectés (Figure 3.6). Seul le QTL sur le LG12 est révélé par les deux modèles, tandis que ceux sur le LG1A et LG8 le sont uniquement avec le modèle BayesCπ.

Le QTL présent sur le LG12 explique 7,48% de la variance génétique et se situe entre 32,76cM et 34,94cM (Tableau 3.4). Celui sur le LG8 explique 1,89% de la variance génétique et se situe entre 28,50cM et 30,32cM. Le QTL présent sur le LG1A explique quant à lui 0,7% de la variance génétique (Tableau 3.4).

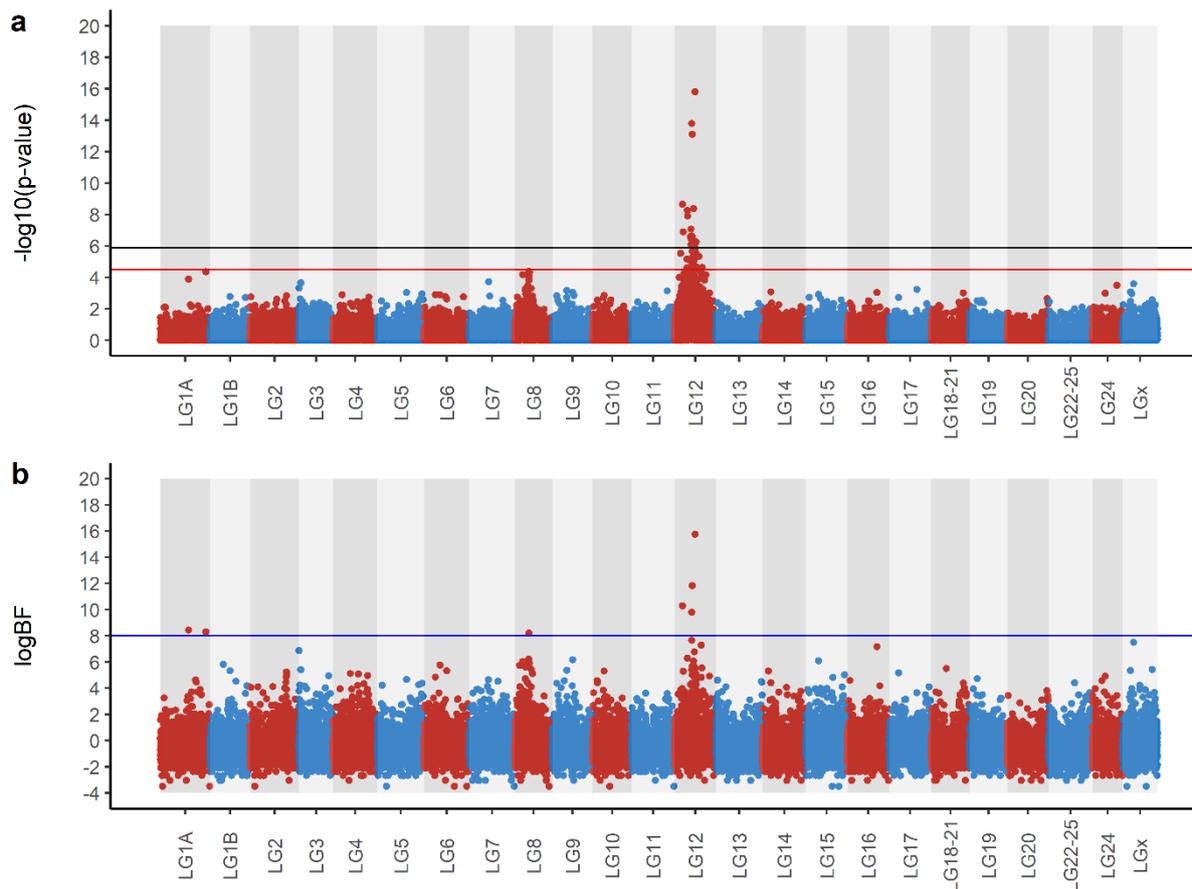


Figure 3.6 : Manhattan plot de $-\log_{10}(p\text{-value})$ obtenus à partir de GWAS pour détecter des QTLs de résistance à la nodavirose chez le bar européen, réalisées avec un modèle GBLUP (a) et un modèle BayesC π (b) dans une population de 2 743 individus composée de quatre cohortes commerciales de challengés à la nodavirose. La ligne rouge représente le seuil de significativité chromosomique et la ligne noire, le seuil de significativité génomique, calculée avec une correction de Bonferroni. La ligne bleue représente un logBF de 8, qui correspond à une preuve forte de la présence d'un QTL.

Tableau 3.4 : Détection de QTL avec un modèle BayesC π pour la résistance à la nodavirose dans une population de bar européen: position de pic SNP, intervalle de crédibilité et proportion de variance génétique expliquée.

chromosome	Position du pic SNP (cM)	Position du début du QTL (cM)	Position de fin du QTL (cM)	Proportion de variance génétique expliquée (%)
LG1A	55,305	55,06	55,89	0,70
LG8	30,203	28,50	30,32	1,89
LG12	34,175	32,76	34,94	7,48

3.2.2.3. Effet du marqueur LG12_8815613

L'effet du marqueur LG12_8815613 sur la survie est présenté dans la Figure 3.7. Ce marqueur est le deuxième plus significatif avec le modèle GBLUP et est le plus significatif avec le modèle BayesCπ (Figure 3.7). L'effet de ce marqueur dans la cohorte VAL2 est proche de celui dans la famille WEM18. Il est additif et les trois génotypes sont présents. L'effet dans la cohorte VAL1 est proche de celui de la cohorte pop B car la survie des hétérozygotes est supérieure à celle des deux homozygotes. De plus, les homozygotes BB sont en très faible nombre ce qui engendre une forte incertitude sur l'estimation de la survie moyenne de ce génotype. Dans la cohorte VAL3, l'effet de ce marqueur n'est pas significatif.

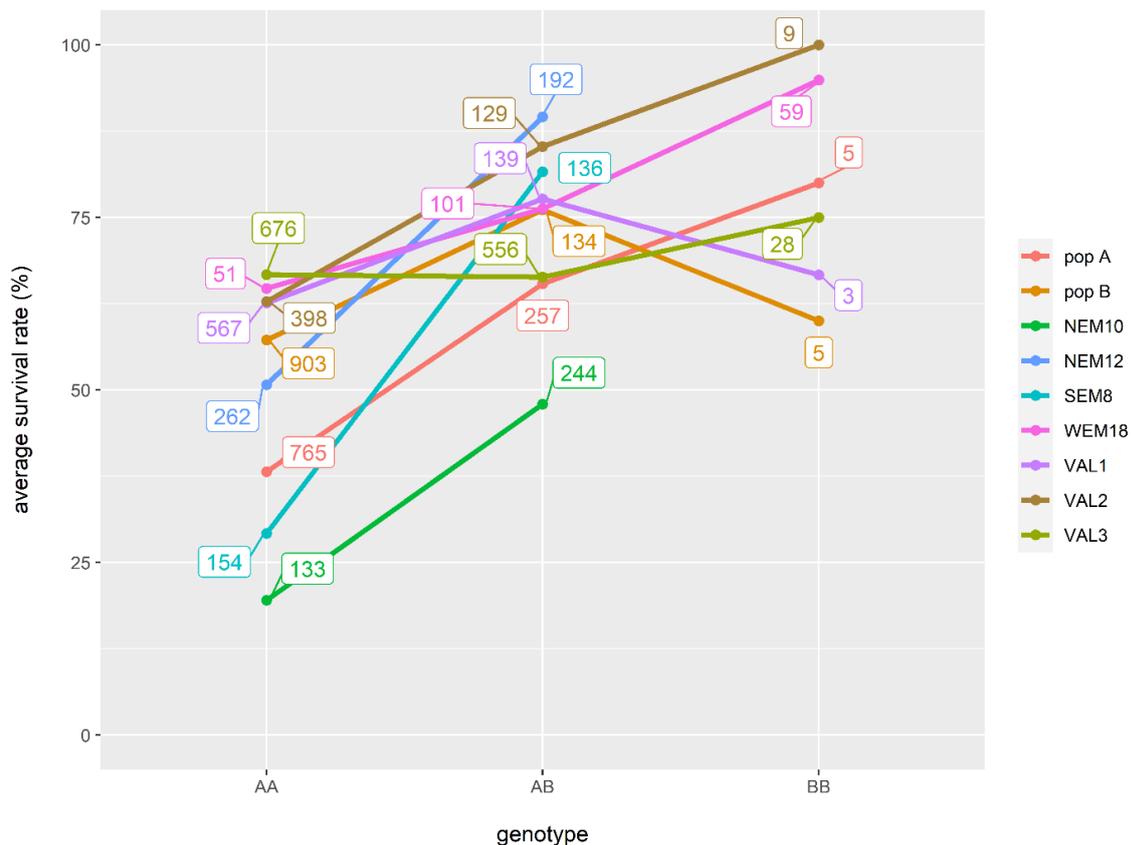


Figure 3.7 : Taux de survie moyen (%) à la nodavirose pour chaque génotype au marqueur LG12_8815613 dans six cohortes de bar européen. Chaque couleur correspond à un jeu de données. Les nombres dans les encadrés représentent le nombre d'individus par génotype.

3.2.3. Discussion

Dans cette partie, nous avons validé la présence d'un QTL de résistance à la nodavirose chez le bar européen dans trois cohortes supplémentaires provenant de EMGI et FMD. Ce QTL, localisé sur le LG12, a été détecté dans deux des trois cohortes additionnelles. Chez VAL1, le QTL a été détecté à un seuil génomique et chez VAL2, à un seuil chromosomique tandis que chez VAL3, il n'a pas été détecté. En plus de la présence, nous avons confirmé le fort effet de QTL en combinant les cohortes dans lesquelles il avait été détecté. Dans cette population diversifiée, il explique 7,48% de la variance génétique totale, ce qui en fait un QTL à effet fort. Pour finir, ce jeu de données a permis de réduire l'intervalle de confiance du QTL, passant de 3,45cM à 2,18cM, ce qui équivaut à une réduction de l'intervalle de confiance de 1Mb.

Dans la cohorte VAL3, le QTL sur le LG12 n'a pas pu être confirmé. Bien que le QTL puisse être absent dans cette cohorte, d'autres hypothèses peuvent expliquer ce résultat. Tout comme la pop B (Griot et al., 2021), la cohorte VAL3 peut présenter une structure génétique qui fausse la détection de QTL et peut masquer la présence du QTL.

Nous avons également validé l'intérêt du marqueur LG12_8815613 dans une sélection assistée par marqueurs. Ce marqueur est le deuxième plus significatif dans la GWAS avec un modèle GBLUP et est le plus significatif dans celle avec un modèle BayesCπ. Dans les cohortes VAL1 et VAL2, l'effet de ce marqueur sur la survie est proche de celui observé dans des cohortes utilisées dans Griot et al. (2021). La fréquence de l'allèle de résistance, ici noté B, est faible dans les trois nouvelles cohortes et son effet sur la résistance à la nodaviriose est fort. Les individus porteurs du génotype AA à ce marqueur résistent à 50% tandis que ceux porteurs du génotype BB résistent à 79,4%. Cependant, l'effet du marqueur LG12_8815614 sur la résistance à la nodaviriose peut être différent d'une cohorte à l'autre. Il semble être additif chez WEM18, pop A et VAL2 alors qu'il semble être dominant chez pop B et VAL1. Cette diversité d'effet peut refléter une association imparfaite entre le marqueur et le QTL. Bien que son utilisation en sélection assistée par marqueurs semble donner une amélioration de la résistance à la nodaviriose, la liaison imparfaite dans certaines cohortes peut entraîner une mauvaise sélection des candidats et donc une stagnation voire régression du gain génétique.

Grâce à cette validation, nous avons confirmé la présence d'un QTL à effet fort impliqué dans la résistance à la nodaviriose chez le bar européen. La combinaison des différentes cohortes en une population de détection a permis de localiser plus précisément ce QTL ainsi que d'en estimer la part de variance génétique expliquée. Bien que l'intervalle de confiance se soit réduit, il reste encore bien trop grand pour y trouver un gène candidat ou la mutation causale. Cette validation a confirmé l'intérêt de poursuivre ces travaux par l'utilisation de données de séquençage qui pourront permettre d'identifier la mutation causale à l'origine de ce QTL.

3.3. Détection de QTL de résistance à la vibriose chez le bar

3.3.1. Matériel et Méthodes

3.3.1.1. Production des animaux et challenge.

Une cohorte commerciale de bar, provenant de l'Écloserie Marine de Graveline-Ichthys (EMG), a été produite par reproduction artificielle de 60 pères et de 18 mères en trois blocs factoriels (20x6, 20x5, 20x7). Parmi les poissons produits, 2 100 individus, de 15g de poids moyen, ont été envoyés à la plateforme Fortior Genetics pour être challengés à la vibriose. À leur réception, les poissons ont été acclimatés une semaine. Parmi les 2100 poissons reçus, 850 n'étaient pas identifiés et ont donc été marqués par une puce RFID. La cohorte a été divisée en trois groupes : 430 individus pour un pré-test, 1 151 individus pour le challenge infectieux et 324 individus comme témoin négatif.

Pour les pré-tests et le challenge infectieux, les poissons ont été injectés avec une solution contenant la souche 94473 1811603 AQN553P2 de *Vibrio harveyi*. Pour les pré-tests, différentes concentrations, allant de $1,6 \times 10^8$ à 8×10^8 UFC/poisson ont été testées. Les mortalités ont été enregistrées chaque jour sur une période de trois semaines. Au terme du pré-test, une concentration de 2×10^8 UFC/poisson a été établie comme optimale pour le challenge. 1 151 poissons ont donc été injectés à 2×10^8

UFC/poisson tandis que les 324 poissons du groupe contrôle ont été injectés avec du milieu de culture stérile. Le challenge a duré 13 jours durant lesquels la mortalité a été enregistrée quotidiennement.

3.3.1.2. *Analyses génétiques*

Les 1151 poissons challengés ont été génotypés avec la puce ThermoFisher Axiom™ DlabChip 57k à la plateforme Gentyane (INRAE Clermont Ferrand). Les mêmes contrôles qualité et analyses génétiques que ceux décrits en 2.1.3.7 et 2.1.3.9 ont été réalisés sur cette cohorte.

1 049 individus génotypés sur 39 421 marqueurs ont été obtenus après contrôles qualité réalisés avec les mêmes paramètres que ceux utilisés précédemment pour la recherche de QTL de résistance à la nodavirus, en plus d'une suppression des marqueurs montrant plus de 10 erreurs mendéliennes dans la population. Tous les marqueurs obtenus ont été gardés pour l'estimation des paramètres génétiques et la réalisation des études d'association (GWAS).

Dans le modèle BayesC π , les paramètres α et β de la loi Beta permettant de définir la valeur de $1 - \pi$, correspondant à la proportion de marqueurs ayant un effet non nul, ont été fixés à 39 421 et 100, respectivement. La variance génétique a été divisée en une partie expliquée par un effet polygénique, représentant 75% de la variance génétique totale, et une partie expliquée par l'effet des marqueurs, représentant 25% de la variance génétique totale. Les degrés de liberté des lois permettant d'échantillonner les valeurs de la variance de l'effet polygénique et de la variance résiduelle ont été fixés à 5.

3.3.2. Résultats

3.3.2.1. *Challenge infectieux*

Le challenge a été conduit pendant 13 jours. La survie moyenne au terme du challenge a été de 59%. Le pic de mortalité a été enregistré deux jours après le début de challenge.

3.3.2.2. *Paramètres génétiques*

L'héritabilité de la résistance à la vibriose était de 0,112 avec un modèle linéaire PBLUP et de 0,105 avec un modèle linéaire GBLUP. En utilisant un modèle à seuil, l'héritabilité est passée à 0,205 en PBLUP et 0,180 avec un modèle GBLUP. L'héritabilité estimée par le modèle BayesC π était de 0,114 (Tableau 3.5).

Après la calibration du modèle BayesC π , la répartition de la variance génétique en une part expliquée par les marqueurs et une part expliquée par un effet polygénique est restée inchangée. L'effet polygénique représentait toujours 75% de la variance génétique totale, contre 25% pour la variance expliquée par l'effet des marqueurs.

Tableau 3.5 : Estimation des paramètres génétiques de la résistance à la vibriose dans une cohorte commerciale de 1 049 bars européens avec trois modèles différents. La variance génétique est divisée en une composante polygénique (V_{poly}) et une composante prenant en compte l'effet des marqueurs (V_{snp}).

		V_{poly}	V_{snp}	V_e	h^2
PBLUP	modèle linéaire	0,027		0,216	0,112 (0,043)
	modèle à seuil	0,268		1,007	0,205 (0,068)
GBLUP	modèle linéaire	0,026		0,218	0,105 (0,039)
	modèle à seuil	0,241		1,006	0,180 (0,061)
BayesC π		0,021	0,007	0,216	0,114

Les nombres entre parenthèses représentent l'écart type de l'estimation

3.3.2.3. Détection de QTL de résistance à la vibriose

Un seul QTL a été détecté avec le modèle GBLUP sur le chromosome LG16, significatif au seuil chromosomique (Figure 3.8a). Avec le modèle BayesC π , aucun QTL n'a pu être détecté à un seuil de $\log BF$ de 8 (Figure 3.8b).

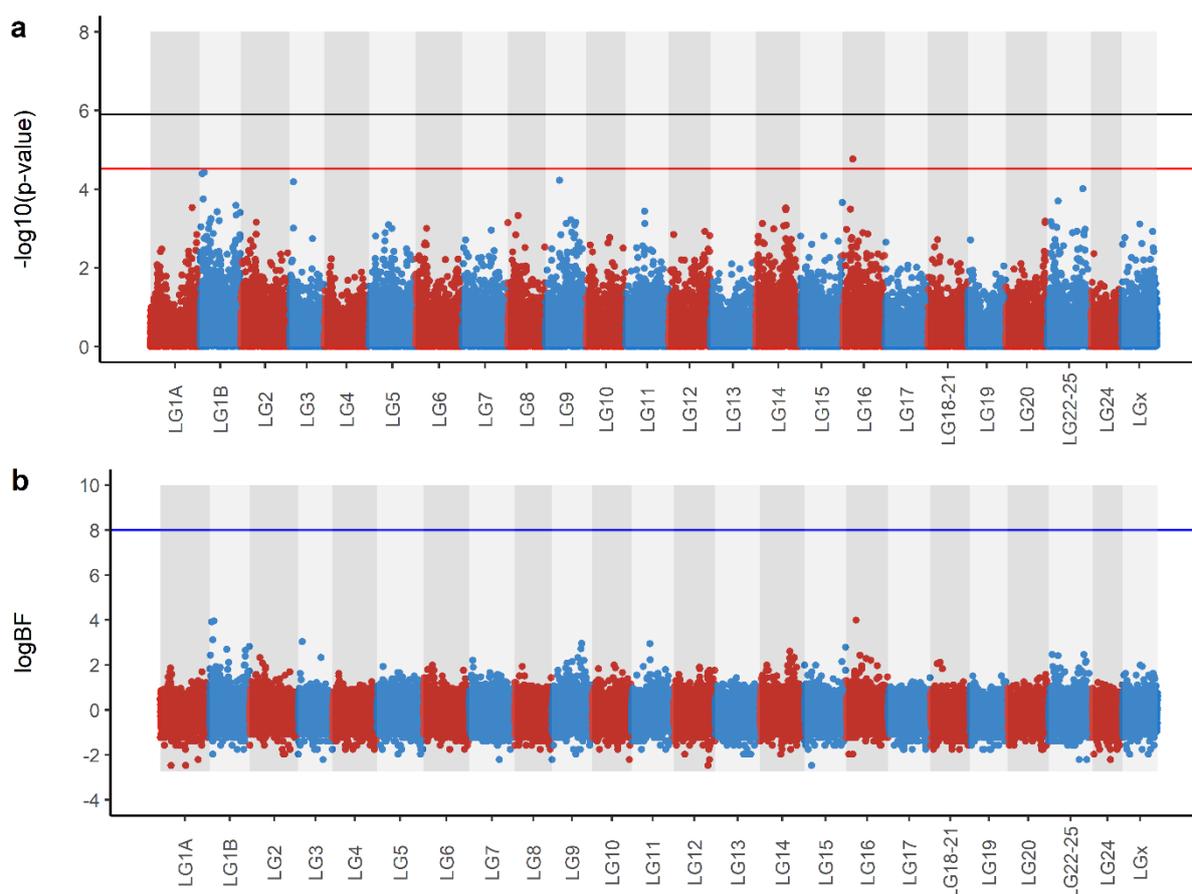


Figure 3.8 : Manhatten plot de $-\log_{10}(p\text{-value})$ obtenus à partir de GWAS pour détecter des QTLs de résistance à la vibriose chez le bar européen, réalisées avec un modèle GBLUP (a) et un modèle BayesC π (b) dans une cohorte de 1 049 bars challengés à la vibriose. La ligne rouge représente le seuil de significativité chromosomique et la ligne noire, le seuil de significativité génomique, calculée avec une correction de Bonferroni. La ligne bleue représente un $\log BF$ de 8, qui correspond à une preuve forte de la présence d'un QTL.

3.3.3. Discussion

Dans cette partie, nous avons présenté les premiers résultats, à notre connaissance, sur l'estimation des paramètres génétiques ainsi que sur la détection de QTL de résistance à la vibriose chez le bar. Son héritabilité, estimée à 0,11 avec un modèle PBLUP, est faible à modérée, ce qui traduit une capacité d'amélioration par sélection. Quel que soit le modèle utilisé, les différentes estimations de l'héritabilité de la résistance à la vibriose sont proches. Chez le cabillaud Atlantique (*Gadus morhua*), l'héritabilité de ce caractère est modérée et a été estimée à 0,16 (Bangera et al., 2011). En revanche, la corrélation génétique estimée entre le poids au challenge et la résistance à la vibriose est assez forte (0,56 avec le modèle PBLUP). Ainsi, sélectionner des individus sur la croissance améliorera la résistance à la vibriose en même temps.

Un seul QTL a pu être détecté avec la GWAS utilisant le modèle GBLUP. Ce QTL, détecté uniquement à un seuil chromosomique, n'a pas été confirmé avec le modèle BayesCπ. Dans ce modèle, la part de variance génétique expliquée par l'effet des marqueurs est restée inchangée après convergence de l'échantillonneur de Gibbs. Par conséquent, les effets des marqueurs ainsi que leur logBF restent faibles et ne permettent pas de mettre en évidence des QTLs. Cette absence de modification dans les parts de la variance génétique expliquée par l'effet polygénique et par l'effet des marqueurs peut traduire un manque de puissance dans le jeu de données, en particulier dû à un faible nombre d'individus. Ceci a été traité plus en détail en 2.1.5.

Cette partie propose la première détection de QTL de résistance à la vibriose chez le bar européen. Un seul QTL a été détecté sur le LG16 avec un niveau de confiance faible et une tendance peut être observée sur les chromosomes LG1B, LG3, LG9 et LG11. Une validation avec un plus grand nombre d'individus permettrait de confirmer la présence de ce QTL voire d'en détecter de nouveaux. A partir de la connaissance de l'architecture génétique de la résistance à la vibriose présentée ici, qui semble être essentiellement polygénique, la sélection génomique devrait être un moyen plus efficace pour sélectionner des individus plus résistants que la sélection assistée par marqueurs.

3.4. Détection de QTL de résistance à la pasteurellose chez la daurade

3.4.1. Matériel et Méthodes

3.4.1.1. Production des animaux et challenge.

Une cohorte commerciale de daurade, provenant de la Ferme Marine du Douhet (FMD), a été produite par reproduction artificielle de 50 pères et 23 mères en six blocs factoriels (8x4, 10x3, 7x4, 10x4, 7x5 and 8x3). Parmi les poissons produits, 1 200 individus, de 3g de poids moyen, ont été envoyés à la plateforme Fortior Genetics pour être challengés à la pasteurellose. A leur réception, les poissons ont été acclimatés une semaine. La cohorte a été divisée en trois groupes : 89 individus pour le pré-test, 960 individus pour le challenge infectieux et 10 individus comme témoin négatif.

Pour les pré-tests et le challenge infectieux, les poissons ont été injectés avec une solution contenant la souche 11822 PP11787 6/94 P2 de *Photobacterium damsela* subsp. *piscicida*. Pour les pré-tests, différentes concentrations, allant de 1×10^{11} à 3×10^{11} UFC/poisson ont été testées. Les mortalités ont été enregistrées chaque jour sur une période de sept jours. Au terme du pré-test, une concentration de 3×10^{11} UFC/poisson a été établie comme optimale pour le challenge. 960 poissons ont donc été injectés à 3×10^{11} UFC/poisson tandis que les 10 du groupe contrôle ont été injectés avec du milieu de

culture stérile. Le challenge a duré 10 jours durant lesquels la mortalité a été enregistrée quotidiennement.

3.4.1.2. *Analyses génétiques*

Les 960 poissons challengés ont été génotypés sur la puce ThermoFisher Axiom™ SaurChip 60k à la plateforme Gentyane. Les mêmes contrôles qualité et analyses génétiques que ceux décrits en 2.1.3.7 ont été réalisés sur cette cohorte.

929 individus génotypés sur 43 400 marqueurs ont été obtenus après contrôles qualité réalisés avec les mêmes paramètres que ceux utilisés chez le bar pour la nodavirose, sauf pour le test d'équilibre de Hardy-Weinberg, où la valeur seuil a été empiriquement fixée à 10^{-4} . A cette valeur, les marqueurs supprimés avaient majoritairement un génotypage de mauvaise qualité (duplication d'amorces, clusters peu espacés). Tous les marqueurs obtenus ont été gardés pour l'estimation des paramètres génétiques et des GWAS.

De la même manière que dans la partie 2.1.3.10, une estimation de l'héritabilité ainsi que des GWAS, l'une basée sur le modèle GBLUP et l'autre sur un modèle BayesC π ont été réalisées. L'estimation de l'héritabilité ainsi que la GWAS basée sur le modèle GBLUP ont été réalisées de la même manière que celle décrite en 2.1.3.9. En revanche, pour celle basée sur le modèle BayesC π , de nombreux paramètres ont été modifiés

Dans le modèle BayesC π , les paramètres α et β de la loi Beta permettant de définir la valeur de $1 - \pi$, correspondant à la proportion de marqueurs ayant un effet non nul, ont été fixés à 43 400 et 60. La variance génétique a été divisée en une partie expliquée par un effet polygénique, représentant 60% de la variance génétique totale, et une partie expliquée par l'effet des marqueurs, représentant 40% de la variance génétique totale. Les degrés de liberté des lois permettant de définir les valeurs de la variance de l'effet polygénique et de la variance résiduelle ont été fixés à 5.

3.4.2. Résultats

3.4.2.1. *Challenge infectieux*

Le challenge a été conduit sur une période de 10 jours. Le taux de survie moyen était de 40% et le pic de mortalité a été observé un jour après le début du challenge.

3.4.2.2. *Paramètres génétiques*

L'héritabilité de la résistance à la pasteurellose était de 0,139 avec un modèle linéaire PBLUP et de 0,159 avec un modèle linéaire GLUP. En utilisant un modèle à seuil, l'héritabilité est passée à 0,291 en PBLUP et 0,295 avec un modèle GBLUP. L'héritabilité estimée par le modèle BayesC π , de 0,160, est proche des estimations des autres modèles (Tableau 3.6).

Après la calibration du modèle bayesC π , la variance génétique expliquée par l'effet polygénique représentait 59% de la variance génétique totale, contre 41% pour la variance expliquée par l'effet des marqueurs.

Tableau 3.6 : Estimation des paramètres génétiques de la résistance à la pasteurellose dans une cohorte commerciale de 929 daurades royales avec trois modèles différents. La variance génétique est divisée en une composante polygénique (V_{poly}) et une composante prenant en compte l'effet des marqueurs (V_{snp}).

		V_{poly}	V_{snp}	V_e	h^2
PBLUP	modèle linéaire	0,032		0,206	0,134 (0,049)
	modèle à seuil	0,411		1,008	0,279 (0,084)
GBLUP	modèle linéaire	0,036		0,203	0,151 (0,044)
	modèle à seuil	0,401		1,007	0,280 (0,064)
BayesC π		0,023	0,016	0,201	0,160

Les nombres entre parenthèses représentent l'écart type de l'estimation

3.4.2.3. Détection de QTL de résistance à la pasteurellose

Avec le modèle GBLUP, aucun QTL impliqué dans la résistance à la pasteurellose chez la daurade n'a été détecté (

Figure 3.9a). Avec le modèle BayesC π , un QTL sur le LG17 a été détecté à un seuil de logBF de 6, correspondant à une probabilité modérée de présence d'un QTL (

Figure 3.9b).

3.4.3. Discussion

Dans cette partie, nous avons présenté les estimations de paramètres génétiques ainsi que la détection de QTL de résistance à la pasteurellose chez la daurade royale. L'héritabilité de ce caractère a été estimée à 0,29 avec un modèle à seuil. Dans Palaiokostas et al. (2016), les auteurs ont estimé l'héritabilité à 0,22 avec un modèle à seuil et utilisant le jour de mort comme phénotype. Dans Aslam et al. (2018), l'héritabilité a été estimée à 0,32 avec un modèle GBLUP. Toutes les valeurs estimées sont proches et indiquent que l'héritabilité de la résistance à la pasteurellose est modérée. De plus, la corrélation génétique avec le poids au challenge est très forte (0,89 avec un modèle PBLUP) ce qui permet une amélioration indirecte de la résistance à la pasteurellose grâce à une amélioration de la croissance.

Aucun QTL n'a pu être détecté quel que soit le modèle utilisé. En revanche, une tendance peut être observée sur le chromosome 17 aussi bien avec le modèle GBLUP que BayesC π , qui indique une probabilité modérée de présence d'un QTL. Une augmentation de la puissance du jeu de données par l'augmentation du nombre d'individus pourrait amener à confirmer la présence d'un QTL sur ce chromosome, sachant qu'il a en été déjà détecté un sur le même chromosome dans une autre étude réalisée sur des daurades provenant du même sélectionneur (Aslam et al., 2018). Ce QTL, détecté grâce à 1 300 individus challengés, était à effet fort et expliquait de 13% à 16% de la variance génétique.

Tout comme la vibriose, l'architecture génétique de la résistance à la pasteurellose chez la daurade semble donc essentiellement polygénique au sein de cette cohorte. La sélection génomique serait donc l'alternative la plus efficace pour améliorer ce caractère. En revanche, si le QTL présent sur le

chromosome 17 était confirmé, une sélection assistée par marqueurs pourrait être envisagée du fait de la large part de variance génétique expliquée.

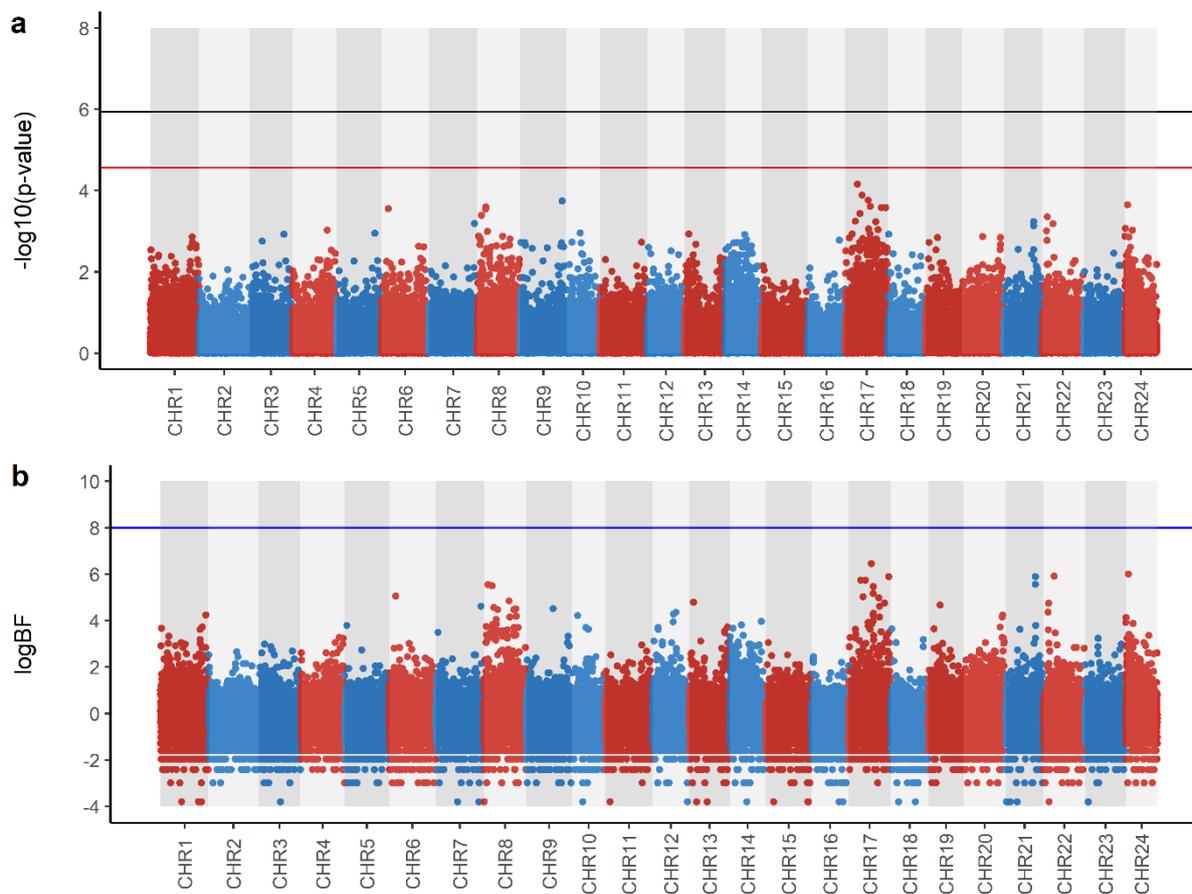


Figure 3.9 : Manhattan plot de $-\log_{10}(\text{p-value})$ obtenus à partir de GWAS pour détecter des QTLs de résistance à la pasteurellose chez la daurade royale, réalisées avec un modèle GBLUP (a) et un modèle BayesCπ (b) dans une cohorte de 929 daurades challengées à la pasteurellose. La ligne rouge représente le seuil de significativité chromosomique et la ligne noire, le seuil de significativité génomique, calculée avec une correction de Bonferroni. La ligne bleue représente un $\log\text{BF}$ de 8, qui correspond à une preuve forte de la présence d'un QTL.

CHAPITRE 4 :

EVALUATION DE L'EFFICACITE DE LA SELECTION GENOMIQUE

Ce chapitre présente l'évaluation de l'efficacité de la sélection génomique pour améliorer les résistances à la nodavirose et à la vibriose chez le bar et la résistance à la pasteurellose chez la daurade

L'objectif était d'évaluer l'efficacité de la sélection génomique, par l'estimation de la précision de sélection, en comparaison avec une sélection sur pedigree. Pour cela, nous avons estimée la précision de sélection par validation croisée, en faisant varier à la fois la taille de la population de référence et le nombre de marqueurs pour évaluer l'impact de ces deux paramètres sur la précision. Enfin, pour la résistance à la nodavirose du bar, nous avons estimé la précision en incluant ou non l'information du QTL de résistance qui a été détecté dans le chapitre précédent et publié dans *Aquaculture* (Griot et al., 2021).

Ce chapitre a été soumis pour publication dans la revue *Frontiers in Genetics* en 2021.

4.1. Optimisation de la sélection génomique pour améliorer la résistance aux maladies dans deux espèces de poissons marins, le bar européen (*Dicentrarchus labrax*) et la daurade royale (*Sparus aurata*)

Cette section a été soumise pour publication dans la revue *Frontiers in Genetics* sous le nom :

« Optimization of genomic selection to improve disease resistance in two marine fishes, the European sea bass (*Dicentrarchus labrax*) and the gilthead sea bream (*Sparus aurata*) »

Griot R.^{a,b,c}, Allal F.^c, Phocas F.^b, Brard-Fudulea S.^a, Morvezen R.^a, Haffray P.^a, François Y.^a, Morin T.^d, Bestin A.^a, Bruant J.S.^e, Cariou S.^e, Peyrou B.^f, Brunier J.^f, Vandeputte M.^{b,c}

^a SYSAAF, Station LPGP/INRAE, Campus de Beaulieu, 35042 Rennes, France

^b Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

^c MARBEC, Univ. Montpellier, Ifremer, CNRS, IRD, 34250 Palavas-les-Flots, France

^d ANSES, Ploufragan-Plouzané-Niort Laboratory, Viral Fish Diseases Unit, National Reference Laboratory for Regulated Fish Diseases, Technopôle Brest-Iroise, 29280 Plouzané, France

^e Ferme Marine du Douhet, 17840 La Brée Les Bains, France

^f Ecloserie Marie Gravelines-Ichthus, 59273 Gravelines, France

4.1.1. Résumé

Les épidémies sont une menace majeure pour l'aquaculture, et la sélection génétique est une stratégie prometteuse pour améliorer la résistance aux maladies des poissons. Avec le développement des technologies de séquençage et de génotypage haut-débit, la sélection génomique est devenue accessible même chez des espèces mineures. La taille de la population d'entraînement et la densité de marqueurs sont deux des paramètres principaux influençant la précision de prédiction, et ont également tous deux un impact élevé sur le coût de la sélection. Dans cette étude, nous avons évalué l'impact de la taille de la population d'entraînement et de la densité de marqueurs sur la précision de prédiction de la résistance aux maladies chez le bar européen (*Dicentrarchus labrax*) et la daurade royale (*Sparus aurata*). Nous avons challengé deux cohortes de bars au virus de la nécrose nerveuse, une cohorte de bars à *Vibrio harveyi* et une cohorte de daurades à *Photobacterium damsela* subsp. *piscicida*. Tous les individus challengés ont été génotypés sur une puce à SNP de moyenne densité de 57K-60K marqueurs. Des marqueurs de ces puces de moyenne densité ont été échantillonnés pour créer des puces à SNP virtuelles de basse densité de 1K, 3K, 6K et 10K marqueurs. De même, les individus challengés ont été échantillonnés pour faire varier la taille de la population d'entraînement de 50 à 800 individus. La précision des valeurs génomiques estimées (modèle GBLUP, utilisant l'information génomique) et des valeurs génétiques estimées (modèle PBLUP, utilisant uniquement l'information sur le pedigree) ont été calculées pour chaque taille de population d'entraînement. Les valeurs génomiques ont également été estimées à partir des puces virtuelles de basse densité pour étudier l'effet de la densité de marqueurs. Pour la résistance à la nodavirose, pour laquelle un QTL a été détecté dans une précédente étude, l'opportunité d'une sélection assistée par marqueurs a été évaluée par l'ajout de l'effet du QTL dans les modèles de prédiction génomique et génétique. Nous avons montré que lorsque la taille de la population d'entraînement augmentait, la précision augmentait également pour atteindre des valeurs comprises entre 0,51 et 0,65 pour les puces

complètes. La précision aurait pu encore augmenter encore avec plus d'individus dans la population d'entraînement, car elle n'avait pas atteint de plateau avec la taille maximale testée de 800 individus. Par ailleurs, avec une puce de 6K marqueurs, la précision atteignait une valeur égale à au moins 90% de celle obtenue à partir de la puce complète. Enfin, l'ajout de l'effet du QTL pour la résistance à la nodavirus augmentait la précision du modèle PBLUP jusqu'à des valeurs supérieures à celles obtenues avec le modèle GBLUP sans l'effet du QTL. Cette étude établit un cadre pour mettre en place la sélection génomique pour améliorer les résistances aux maladies chez le bar et la daurade.

4.1.2. Introduction

Viral and bacterial infectious diseases are a major threat to the development of aquaculture production (Gjedrem, 2015). In the Mediterranean Sea, fish culture is mainly focused on European sea bass (*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*) with a production of 157,000 and 160,000 tons respectively in 2016 (FEAP, 2017). Viral Nervous Necrosis (VNN) caused by the nervous necrosis virus (NNV), vibriosis caused by *Vibrio harveyi* and pasteurellosis due to *Photobacterium damsela* subsp. *piscicida* are considered as the most impacting diseases for Mediterranean aquaculture (Vendramin et al., 2016). They cause high mortality in aquaculture farms and both prophylaxis and therapeutics remain insufficient to control outbreaks.

Selective breeding to improve the genetic resistance to those pathogens is a promising approach to prevent outbreaks. It has been successfully applied in many aquaculture species and led to a genetic gain of 12.5% in disease resistance per generation on average over a number of host-pathogen pairs (Gjedrem and Robinson, 2014). To select for genetically resistant individuals, enough genetic variability must exist within the species. Moderate to high heritability (0.24 to 0.43) has been reported for resistance to VNN in European sea bass (Doan et al., 2017a; Griot et al., 2021; Palaiokostas et al., 2018b) and for resistance to pasteurellosis in gilthead sea bream (from 0.22 to 0.32 ; Aslam et al., 2018; Palaiokostas et al., 2016), thus leaving room for genetic improvement.

Genomic selection is a tool that increases the efficiency of selective breeding. Classical pedigree-based selective breeding derives genetic relationships between the individuals from pedigree records and combines them with phenotypes to estimate breeding values. Genomic selection uses genomic markers spread along the genome to estimate those genetic relationships more accurately, leading to better estimates of breeding values (Meuwissen et al., 2001). For most of the traits studied in aquaculture, genomic selection indeed outperforms pedigree-based selection and increases the genetic gain per generation (Nielsen et al., 2011; Zenger et al., 2019).

Genomic selection is particularly interesting to improve disease resistance because phenotypes are collected on related individuals instead of the selection candidates themselves, in order to avoid contacts between the pathogen and the selection candidates. The genomic breeding values of the selection candidates are predicted based on the phenotypes recorded on challenged relatives and the markers genotyped on both challenged individuals and selection candidates (see review of Ødegård et al. (2011)). When the phenotype is collected on relatives and not on the candidates, pedigree-based selection only accounts for between-family genetic variance, but genomic selection can account for both between and within-family genetic variances and thus, allows to rank the individuals within their family, which makes it more efficient.

From an economic perspective, the efficiency of genomic selection is limited by two main factors. The first one is the cost of the phenotypes. In disease resistance breeding programs, the phenotypes must be recorded on individuals that are closely related to the selection candidates (full and half-sibs) and

only produced for this phenotyping purpose. Although natural challenge on the field could be used (Fraslin et al., 2019), the disease challenge is preferred to take place in a dedicated facility to control the infection process (pathogen strain, concentration, route of infection), to record daily the mortality and to avoid the risk of spreading the pathogen in the wild and to commercial lines (Fjalestad et al., 1993). All those factors make disease resistance phenotyping expensive. The second limiting factor is the necessary genotyping of both selection candidates and challenged relatives. Even though genotyping becomes more and more affordable, it remains expensive for the majority of breeding companies (Ødegård et al., 2011).

To limit the cost of a disease resistance breeding program, the options available are 1) to reduce the number of phenotypes recorded, which results in reduced costs for both genotyping and phenotyping, and/or 2) to reduce the number of markers genotyped, which only decreases the cost of genotyping. Limiting the number of phenotyped individuals, that constitute the training population, affects the accuracy of genomic prediction (Dufflocq et al., 2019). In general, as the number of individuals in the training population increases, the accuracy increases until it reaches a plateau. The density of markers has similar effects, as prediction accuracy increases with the number of markers, until it also reaches a plateau (Kriaridou et al., 2020; Tsai et al., 2016).

The aim of this study was to assess the impact of training population size as well as marker density on the accuracy of genomic selection, compared to a pedigree-based selection, to improve disease resistance in fish. To do so, we used one data set from gilthead sea bream challenged to *Photobacterium damsela* subsp. *piscicida*, one data set from European sea bass challenged to *V. harveyi* and two data sets from European sea bass challenged to NNV.

4.1.3. Matériel et Méthodes

4.1.3.1. Approbation du comité d'éthique

All infection challenges were carried out in accordance with the European guidelines (Directive 2010–63-EU) and the corresponding French legislation. Animal experiment procedures were approved by the ethics committee COMETH n°16 (ANSES, ENVA and UPEC) and authorized under numbers 14/03/17-10 (n° APAFiS: 2017022816255366), 29/01/13-05 and 10/03/15-01 by the French Ministry of Higher Education, Research and Innovation.

4.1.3.2. Animaux utilisés

The animals challenged came from four commercial cohorts from the breeding programs of two different companies. The commercial cohorts were produced by artificial mating and are identified as VNN_A and VNN_B for the two European sea bass cohorts challenged to NNV, VIB for the European sea bass cohort challenged to *V. harveyi* and PAS for the gilthead sea bream cohort challenged to *Photobacterium damsela* subsp. *piscicida*. Cohort VNN_A (1,680 individuals) was produced by mating 59 sires with 20 dams in four partial factorial designs (15x5, 15x5, 15x5 and 14x5). Cohort VNN_B (1,737 individuals) was generated from 39 sires and 14 dams mated in six factorial subsets (6x3, 6x1, 6x3, 7x2, 7x3 and 7x2). Cohort VIB (2,100 individuals) was produced by mating 60 sires with 18 dams in three factorial subsets (20x6, 20x7 and 20x5) and cohort PAS (1,200 individuals) was produced by mating 50 sires with 23 dams in 6 factorial subsets (8x4, 10x3, 7x4, 10x4, 7x5 and 8x3). The cohorts VNN_A and VNN_B were referred as popA and popB in Griot et al. (2021).

4.1.3.3. Challenges infectieux

All infection challenges were performed at the SYSAAF-ANSES Fortior Genetics platform (ANSES, Plouzané, France). All fish were individually tagged with RFID glass tags. Some fish were sampled from the quarantine time to check their sanitary status through virological and bacteriological analysis before the challenge. In each infection challenges, pre-tests were made using 180, 150, 430 and 89 randomly sampled individuals from the VNN_A, VNN_B, VIB and PAS cohorts.

For the challenge tests themselves, 1,350 individuals from VNN_A (mean body weight = 25g) and 1,212 individuals from VNN_B (mean body weight = 20g) were challenged to NNV in filtered and UV sterilized seawater at 27°C ±. The infection was done by immersing them in static seawater containing 1x10⁵ Tissue Culture Infectious Dose (TCID₅₀)/mL of the W80 strain of RGVVN (redspotted grouper nervous necrosis virus; Thiéry et al., 2004) produced on SSN-1 (snakehead fish) cell line for two hours and fifteen minutes. Then, mortality was recorded daily for 27 days for VNN_A and 42 days for VNN_B.

In the VIB cohort, 1,475 individuals (mean body weight = 15g) were challenged to *V. harveyi* at 22°C ±2. The fish were infected by intraperitoneal injection of the strain 94473 1811603 with a concentration of 2x10⁸ Colony Forming Unit (CFU)/fish. The mortality was recorded daily for 13 days.

In the PAS cohort, 960 individuals (mean body weight = 3g) were challenged to *Photobacterium damselae subsp. piscicida* at 24°C ±2. They were infected by intraperitoneal injection of the strain PP11787 6/94 with a concentration of 3x10¹¹ CFU/fish. The mortality was recorded daily for 10 days.

In each challenge, random samples were taken at different times of the mortality kinetics to confirm the presence of the pathogen using cell culture followed by an immunofluorescence antibody test (IFAT) for NNV and bacterial isolation followed by a Maldi Tof identification culture for *V. harveyi* and *Photobacterium damselae subsp. piscicida*.

The details of the infection challenge applied to each cohort is summarized in Tableau 4.1.

Tableau 4.1 : Summary of infection challenges procedure followed by the four commercial cohorts (VNN_A, VNN_B, VIB and PAS).

	VNN_A	VNN_B	VIB	PAS
number of individuals	1680	1737	2100	1200
number of parents (sires / dams)	59 / 20	39 / 20	60 / 18	50 / 23
number of individuals in pre-test	180	150	430	89
number of individuals challenged	1350	1212	1475	960
pathogen	RGNNV	RGNNV	<i>Vibrio harveyi</i>	<i>Photobacterium damsela</i> subsp. <i>piscicida</i>
strain	W80	W80	94473 1811603 AQN553P2	PP11787 6/94
infection method	immersion	immersion	IP injection	IP injection
concentration	1x10 ⁵ TCID/mL	1x10 ⁵ TCID/mL	2x10 ⁸ CFU/fish	3x10 ¹¹ CFU/fish
water temperature (°C)	27 ±2	27 ±2	22 ±2	24 ±2
duration of the challenge (in days)	27	42	13	10
average survival rate	45.2%	59.7%	59.0%	40.0%

4.1.3.4. Génotypage, Contrôles qualité et Assignment de parenté

Genotyping was performed at the Gentyane genotyping platform (INRAE, Clermont-Ferrand, France). From the challenged individuals and their parents, 1,152 individuals from VNN_A, VNN_B and 1,151 individuals from VIB were genotyped on the ThermoFisher Axiom™ 57k SNP DlabChip. In the PAS cohort, 1,026 individuals were genotyped on the ThermoFisher Axiom™ 60k SNP SaurChip. Genotyped individuals were sub-sampled from the challenged ones ensuring each sub-sample had the same average survival rate as the whole challenge batch. SNP calling was performed using the ThermoFisher AxiomAnalysisSuite software and quality controls with PLINK 1.9 (Purcell and Chang, 2015). First, individuals with a genotyping rate lower than 90% were discarded. Then, for the sea bass data sets (VNN_A, VNN_B and VIB), genotyping quality controls were performed on a global population composed of the three data sets. A common set of markers was subset by keeping only markers with a call rate higher than 95%, a minor allele frequency higher than 0.05 and a p-value for the departure from Hardy-Weinberg equilibrium test higher than 10⁻⁸, resulting in 44,772 common markers to be used for the three cohorts. For the PAS data set, markers with a call rate higher than 95%, a minor allele frequency higher than 0.05 and a p-value for the departure from Hardy-Weinberg equilibrium test higher than 10⁻⁴ were retained, leaving 43,618 usable markers. Finally, all missing genotypes were imputed with Flmpuete (Sargolzaei et al., 2014).

Parentage assignment was done using 1,000 randomly sampled markers, analysed with the R package APIS (Griot et al., 2020) with a positive assignment error rate set to 1%.

4.1.3.5. Estimation de l'héritabilité

For each data set, we estimated the heritability of disease resistance with either a threshold model using THRGIBBSF90 (Tsuruta and Misztal, 2006) or a linear model using AIREMLF90 (Misztal et al., 2002). Only individuals with a phenotype, a genotype and a pedigree were used in heritability estimates, thus the sample size was 1,027 in VNN_A and, 1042 in VNN_B, 1,049 in VIB and 916 in PAS.

The following model was computed in each data set using both threshold and linear models:

$$\mathbf{y} = \mathbf{1}b + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

With \mathbf{y} the vector of the phenotypes measured as binary dead/survival trait, $\mathbf{1}$ the incidence (unity) vector of the intercept, b the estimate of the intercept effect, \mathbf{u} the vector of breeding values and \mathbf{Z} the corresponding incidence matrix. To compare pedigree-based and genomic-based heritability estimation, \mathbf{u} followed either a multivariate normal distribution $N(0, \mathbf{A}\sigma_g^2)$ with \mathbf{A} the pedigree-based relationship matrix or a multivariate normal distribution $N(0, \mathbf{G}\sigma_g^2)$ with \mathbf{G} the genomic relationship matrix proposed by VanRaden (2008). σ_g^2 is the additive genetic variance and \mathbf{e} is the vector of the random residual errors that follows a normal distribution $N(0, \mathbf{I}\sigma_e^2)$ with σ_e^2 the residual variance and \mathbf{I} the identity matrix.

With the threshold model, the variance components (σ_g^2 and σ_e^2) were estimated using a Gibbs sampler with 500,000 iterations, 100,000 of burn-in and one sample was kept every 20 iterations for posterior analysis. The residual variance σ_e^2 was set to a value of 1. The posterior distributions were analyzed with the R package `boa` (Smith, 2007). With the linear model, the same components were estimated using a restricted maximum likelihood algorithm, considering the observed binary phenotype as a continuous variable.

The heritability for survival was estimated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Heritability on the observed scale (h^2_o) was estimated using the variance components from the linear model, while the heritability on the underlying liability scale (h^2_u) was computed using the variance components from the threshold model.

4.1.3.6. Création de puces virtuelles de basse densité

From the SNP markers obtained after quality controls, four low-density (1K, 3K, 6K and 10K) virtual SNP chips were created. To do so, we used a marker pruning method based on the linkage-disequilibrium (LD) (Porto-Neto et al., 2013). In a user-defined sliding window, every pairwise LD between markers was estimated by the r^2 metrics. Then, SNPs were pruned until no pair had a r^2 greater than a given value, until we reached the desired number of SNPs. For sea bass data sets, we used an iterative method to create the low-density chips.

For the creation of the sea bass 10K chip, we used an iterative method with a sliding window of 1 Mb and a r^2 value of 0.434. First, the SNP were pruned in the cohort VNN_A data set with the LD method explained above. Then, the remaining markers in cohort VNN_A were subset from cohort VNN_B and the same pruning method was applied on the remaining ones. The same process was done one last time on the cohort VIB data set to obtain the desired number of markers. Finally, the remaining markers in the VIB data set were subset in cohort VNN_A and cohort VNN_B data sets to obtain 10,020 markers in each cohort.

The same process was repeated to create the 1K, 3K and 6K chips by modifying the sliding window size of 100, 200 and 500 kb and the r^2 threshold value of 0.175, 0.268 and 0.344 for 1K, 3K and 6K respectively. The virtual chips for sea bass contained 1,007, 3,022, 6,010 and 10,020 markers for the 1K, 3K, 6K and 10K respectively.

For the PAS data set, the pruning was done by applying the same protocol, but on one dataset only. The chips contained 1,007, 2,999, 6,011 and 10,010 markers for the 1K, 3K, 6K and 10K respectively and were done by using a sliding window of 100, 200, 500, 1,000 kb and a r^2 threshold of 0.044, 0.191, 0.318 and 0.434 for the 1K, 3K, 6K and 10K respectively. In each species, the size of the sliding window as well as the r^2 value for LD were empirically chosen to uniformly sample the desired number of markers.

4.1.3.7. *Création de la population d'entraînement*

We fixed the validation population size to 200 individuals randomly chosen from the whole population in each data set. The training population was composed of a minimum of 50 individuals up to 800 in sea bass data sets and 700 in sea bream data set. From 50 to 200, we added 50 individuals at each step and then, from 200, we added 100 individuals per step. The initial 50 individuals were randomly sampled from remaining individuals after the validation population had been chosen. Then, to increase the training population size, the added individuals were randomly chosen from the remaining individuals and added to the previous training population. This process was repeated until the training validation population size reached the maximum limit. To account for stochastic sampling effects, the entire process was repeated 100 times for each data set.

4.1.3.8. *Effet de la densité de marqueurs et de la taille de la population d'entraînement sur la précision de prédiction*

For each replicate of training population size, we tested six densities of markers. We tested the low-density SNP chips (1K, 3K, 6K and 10K) as well as the full SNP chip (44K for sea bass data sets and 43K for sea bream data set) and a control case with only pedigree information and no genomic information.

The phenotypes of the individuals in the validation population were masked and the breeding values were estimated with the same linear model as the one described in “Heritability estimation” using the genomic relationship matrix when genomic information was available or the pedigree-based relationship matrix otherwise. The estimation using the genomic relationship matrix was reported as GBLUP and the estimation with the pedigree relationship matrix as PBLUP. The accuracy was computed as:

$$r = \frac{\text{cor}(EBV, y)}{h}$$

With $\text{cor}(EBV, y)$ the correlation between the estimated breeding values (EBV) and the phenotypes y of the 200 individuals belonging to the validation population and h the square-root of the heritability estimated with a linear model, using the pedigree-based relationship matrix and the whole data set as in the section “Heritability estimation”.

4.1.3.9. Ajout de l'effet du QTL sur la précision de prédiction pour la résistance à la nodavirose

In Griot et al. (2021), the authors showed that one strong effect QTL is involved in VNN resistance in European sea bass. Among the markers within the confidence interval, they highlighted one marker (LG12_8815613) with high potential for marker assisted selection. Here, we assessed the impact of adding the QTL effect in the estimation of breeding values on the selection accuracy. To do so, in the cohorts VNN_A and VNN_B, the model described in section “Effect of the density of markers and training population on selection accuracy” was replaced by the following model as proposed by Kennedy et al. (1992):

$$y = Xb + Zu + e$$

With all parameters remaining the same except for Xb where X is the genotype matrix for the LG12_8815613 marker coded as 0 for the homozygous, 1 for the heterozygous and 2 for the alternative homozygous and b is the marker fixed effect.

Then, the GEBV were computed as:

$$GEBV = \hat{u} + X\hat{b}$$

With \hat{u} the estimated additive effect of the polygenic breeding value and \hat{b} the estimated QTL effect.

Finally, accuracy was estimated as in section “Effect of the density of markers and training population on selection accuracy” only for a training population of 800 individuals.

4.1.4. Résultats

4.1.4.1. Challenges infectieux

The four challenges were conducted up to 42 days. The presence of the different pathogen agents in their respective challenges was confirmed by virologic or bacteriological analyses on several subsets of fish that died during the challenges. Survival rates ranging from 40.0% to 59.7% were recorded (Figure 4.1). The mortality peaks were early for challenges to *V. harveyi* and *Photobacterium damsela subsp. piscicida*, at three and two days respectively. For NNV, the peak was at 9 days post infection, and was much sharper in cohort VNN_A than in cohort VNN_B.

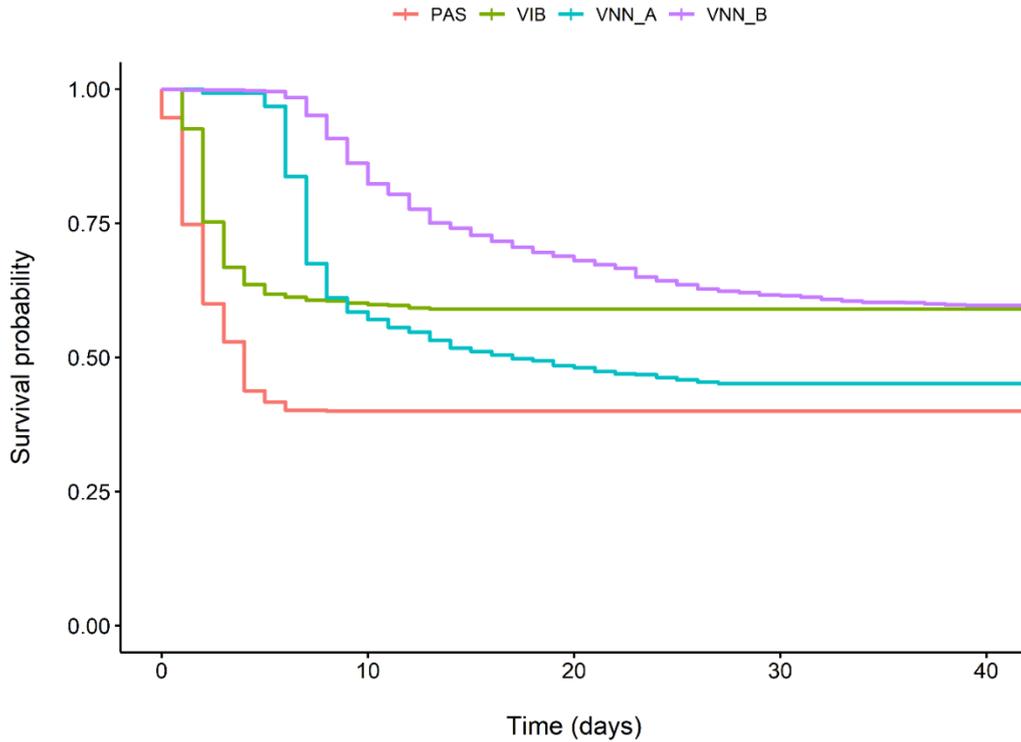


Figure 4.1 : Kaplan-Meier probability of survival over time following infection for two European sea bass commercial cohorts challenged to NNV (VNN_A, VNN_B), one European sea bass commercial cohort challenged to *V. harveyi* (VIB) and one gilthead sea bream commercial cohort challenged to *Photobacterium damsela* subsp. *piscicida* (PAS).

4.1.4.2. Estimation de l'héritabilité

Heritability was moderate for all diseases, ranging from 0.103 to 0.238 using a pedigree-based linear model and from 0.118 to 0.232 using a genomic linear model (Tableau 4.2). For all challenges, genomic and pedigree based heritability estimates were very similar. When using the threshold model, the estimates were much higher than those estimated with the linear models, ranging from 0.198 to 0.421 for pedigree-based heritability and from 0.198 to 0.379 for genomic heritability.

Tableau 4.2 : Heritability estimated for VNN resistance in two European sea bass commercial cohorts (VNN_A and VNN_B), vibriosis resistance in one European sea bass commercial cohort (VIB) and pasteurellosis resistance in one gilthead sea bream commercial cohort (PAS) with pedigree-BLUP (PBLUP) or genomic-BLUP (GBLUP) using linear or threshold models.

population	PBLUP		GBLUP	
	linear model	threshold model	linear model	threshold model
VNN_A	0.238 (± 0.063)	0.421 (± 0.106)	0.232 (± 0.049)	0.379 (± 0.065)
VNN_B	0.103 (± 0.048)	0.214 (± 0.087)	0.118 (± 0.043)	0.217 (± 0.068)
VIB	0.109 (± 0.043)	0.198 (± 0.068)	0.111 (± 0.040)	0.198 (± 0.064)
PAS	0.139 (± 0.051)	0.291 (± 0.086)	0.159 (± 0.045)	0.295 (± 0.066)

Values in parenthesis are the standard errors

4.1.4.3. *Effet de la densité de marqueurs et de la taille de la population d'entraînement sur la précision de prédiction*

In all data sets, accuracy increased with the size of the training population (Figure 4.2). From 50 to 150 individuals in the training population, the increase in accuracy was the greatest. From 150 to 400-500 individuals in the training population, the increase in prediction accuracy was intermediate and with more than 400-500 individuals in the training population, the prediction accuracy increased slowly. With the full density chip, GBLUP with a training population size of 200, 500, 500 and 300 individuals reached the same accuracy as PBLUP for VNN_A, VNN_B, VIB and PAS respectively.

In general, accuracy increased with the density of markers (Figure 4.2). The addition of genomic information improved the accuracy compared to that of pedigree-based estimation, except in VNN_B where the accuracy estimated with PBLUP was greater than that estimated with 1,000 markers (Figure 4.2b). With the maximum training population and the full density chip, genomic evaluation led to an increase in accuracy, compared to PBLUP, of 24.5%, 8.9%, 11.6% and 12.9% for VNN_A, VNN_B, VIB and PAS respectively. Except in VNN_B, the use of one thousand markers for genomic evaluation increased the accuracy by 12.5%, 11.0% and 6.6% for VNN_A, VIB and PAS compared to PBLUP. In general, a density of 6K was enough to reach at least 90% of the accuracy obtained with the full density chip (Figure 4.3).

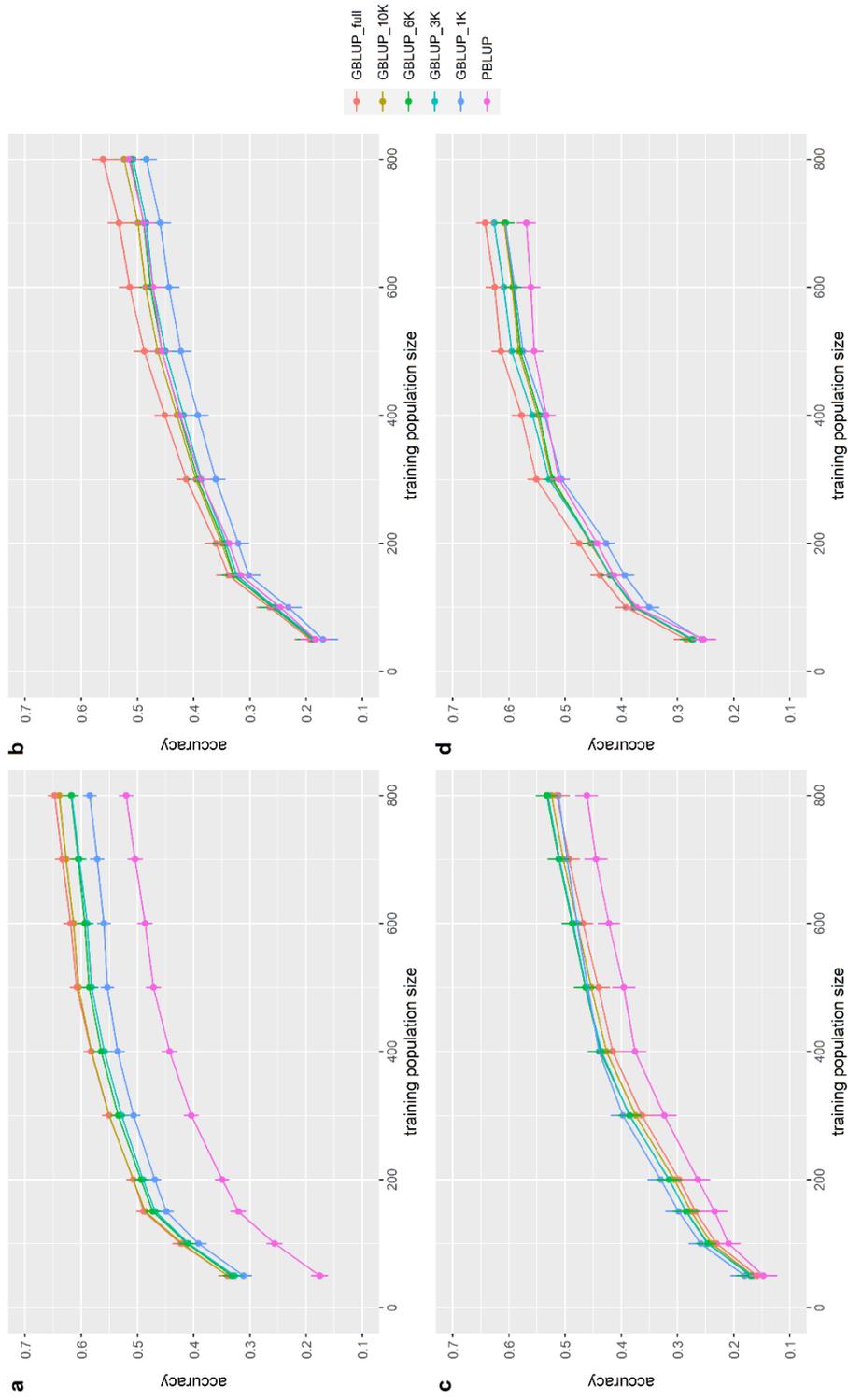


Figure 4.2 : Accuracy of genomic (GBLUP) and pedigree-based (PBLUP) estimated breeding values for disease resistance as a function of the number of individuals in the training population, and for different marker densities, in (a), European sea bass commercial cohort VNN_A challenged to NNV (b) European sea bass commercial cohort VNN_B challenged to NNV, (c) European sea bass commercial cohort VIB challenged to *V. harveyi* and (d) gitthead sea bream commercial cohort PAS challenged to *Photobacterium damselae subsp. piscicida*. Each point is the average of 100 replicates. Error bars represent the standard error of the mean of 100 replicates.

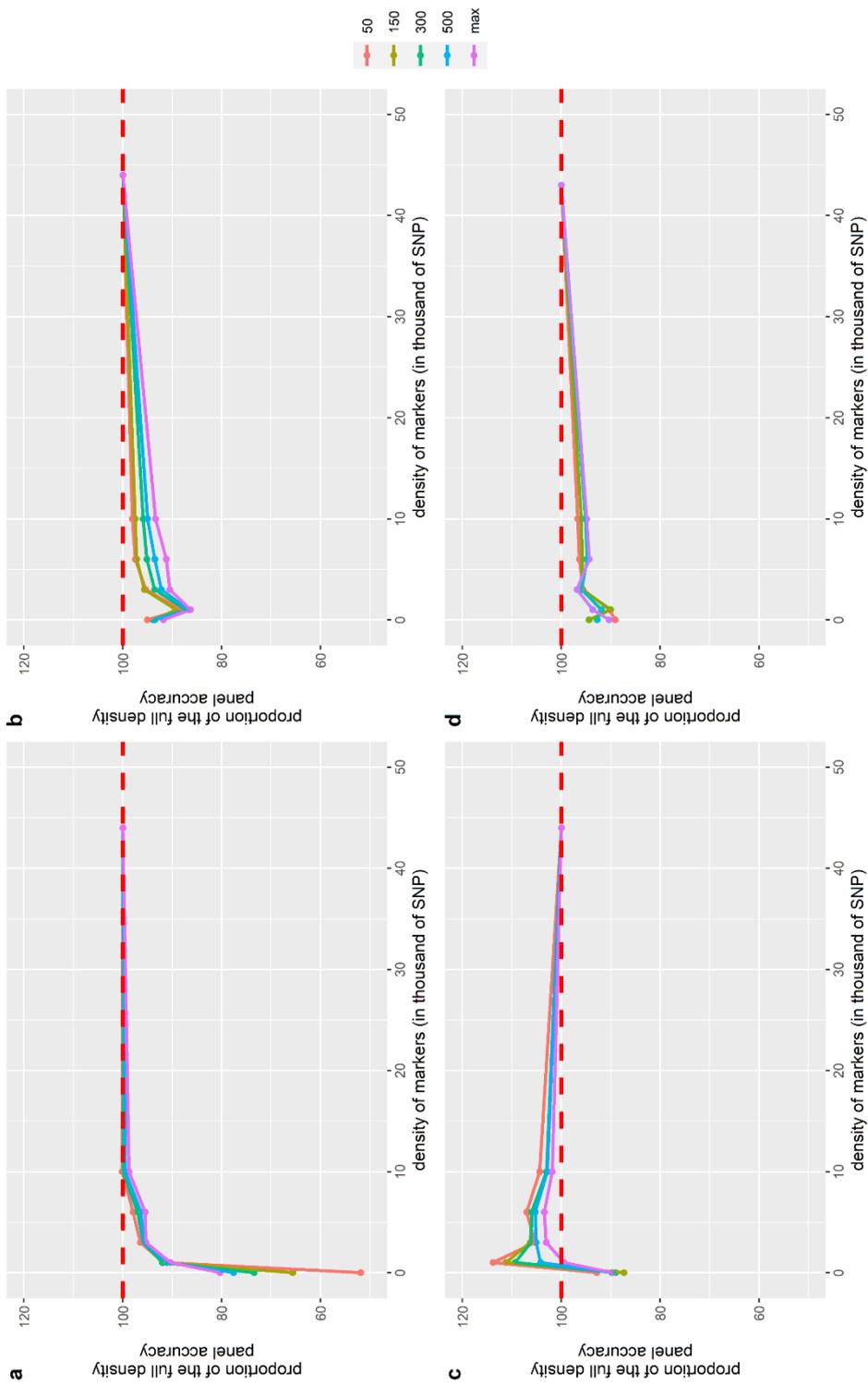


Figure 4.3 : Proportion of the full-density SNP panel accuracy for genomic breeding value estimates as a function of the density of markers in (a), European sea bass commercial cohort VNN_A challenged to NNV (b) European sea bass commercial cohort VNN_B challenged to NNV (c) European sea bass commercial cohort VIB challenged to *V. harveyi* and (d) gilthead sea bream commercial cohort PAS challenged to *Photobacterium damsela* subsp. *piscicida*. The density of markers is expressed in thousands of SNP. Only training population size of 50, 150, 300 500 and 700 for PAS and 800 for others are displayed and represented by the colour palette. Each point is the average of 100 replicates.

4.1.4.4. Ajout de l'effet du QTL sur la précision de prédiction pour la résistance à la nodavirose

The addition of the QTL effect increased the prediction accuracy in a range of 10.5% to 26.3% compared to the prediction accuracy estimated without it (Figure 4.4). In every cohort and for every virtual chip density, it led to an increase in prediction accuracy. The prediction accuracy using just the pedigree and the QTL information was slightly higher than that of the full density chip GBLUP for either VNN_A and VNN_B (Figure 4.4c, Figure 4.4d).

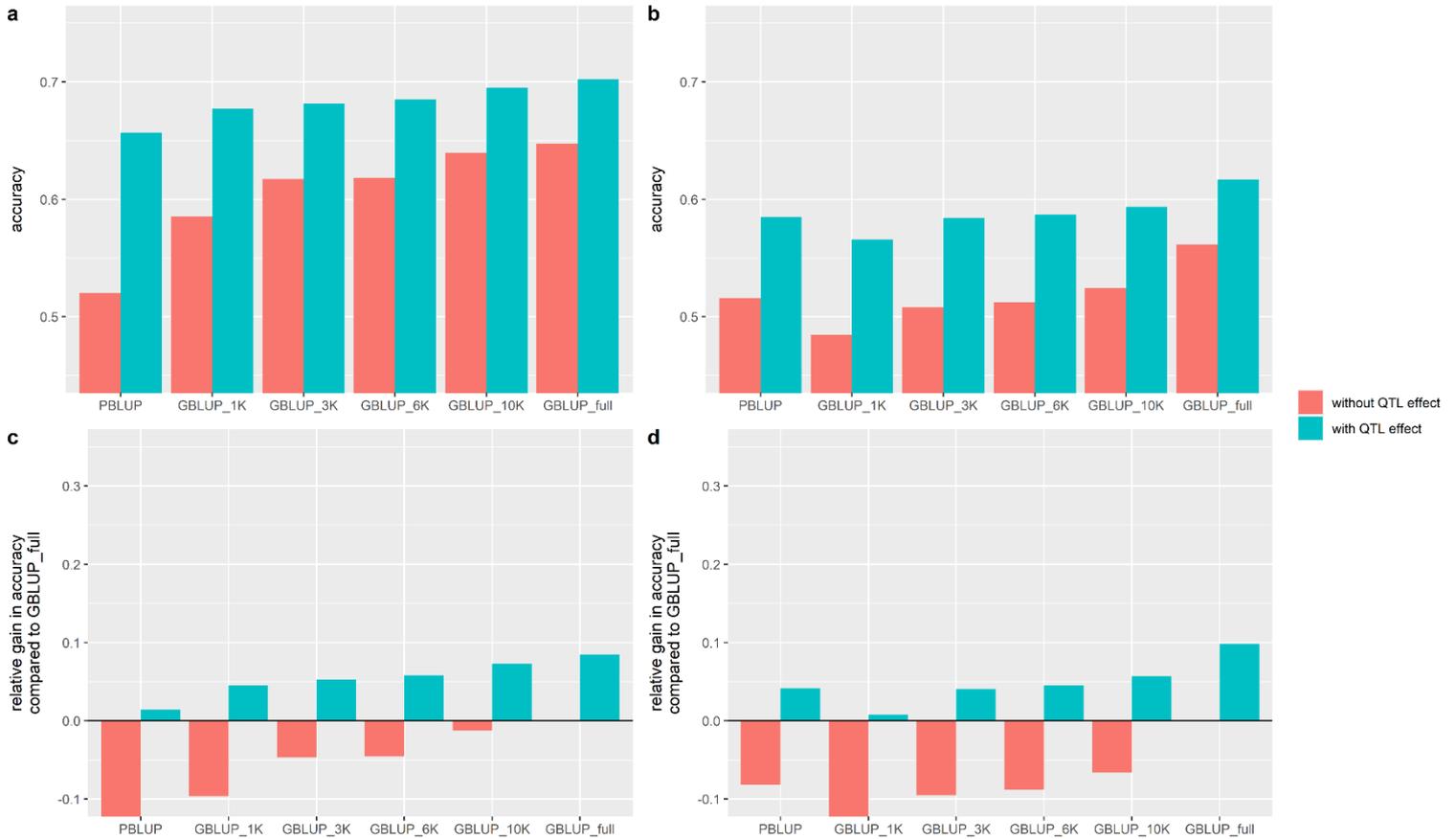


Figure 4.4 : Accuracy of genomic (GBLUP) and pedigree-based (PBLUP) estimated breeding values for VNN resistance in two European seabass commercial cohorts (VNN_A, a and VNN_B, b) with different SNP chip density and with (in blue) or without (in red) the QTL effect and a training population of 800 individuals. Relative gain in accuracy compared to the GBLUP_full model ignoring the QTL effect in cohort VNN_A (c) and VNN_B (d).

4.1.5. Discussion

Genomic selection is widely recognized as having a great potential to improve selective breeding (Meuwissen et al., 2016). In the context of improving pathogen resistance in aquaculture species, its relevance has already been shown (Houston, 2017). In the present work, we provided essential data to implement genomic selection to improve disease resistance to the most common pathogens in European sea bass and gilthead sea bream aquaculture. First, we obtained moderate heritability estimates for resistance to the three pathogens. With the threshold model, the pedigree-based and genomic heritability estimates were rather similar, except for VNN_A for which the pedigree-based heritability estimate was higher than the genomic heritability estimate. The use of genomic information mainly reduced the standard error of the estimates and thus, improved their precision. The heritability estimates were very different between linear model and threshold model, even after applying the correction of the linear estimates proposed by Dempster and Lerner (1950) (data not shown). Our heritability estimates were similar to those reported in former studies, when available: for VNN resistance in sea bass, threshold model heritability estimates were in the range 0.21-0.42, to be compared to previous values of 0.24-0.43 estimated using threshold model or linear model corrected using Dempster and Lerner formula (Doan et al., 2017a; Griot et al., 2021; Palaiokostas et al., 2018b). For pasteurellosis resistance in sea bream, our estimates from linear model (0.14-0.16) were in the lower range of those from previous studies, estimated using linear model (0.22 to 0.45 ; Antonello et al., 2009; Aslam et al., 2018; Palaiokostas et al., 2016). For vibriosis resistance in sea bass, we presented the first genetic parameters estimates to our knowledge. In other aquaculture species, vibriosis resistance has shown low to moderate heritability, ranging from 0.13 in Atlantic salmon on *V. salmonicida* to 0.19 in Chinese tongue sole on *V. harveyi* using a linear model and 0.16 in Atlantic cod on *V. anguillarum* using a threshold sire-dam model (Bangera et al., 2011; Gjedrem and Gjøen, 1995; Li et al., 2019). Similar results were found on shellfish, with moderate heritability of 0.11 in white shrimp on *V. parahaemolyticus* and 0.09 to 0.33 in Pacific oysters on *V. aestuarianus* (Azéma et al., 2017; Lyu et al., 2020).

We showed that genomic selection using a 44K SNP chip can improve the accuracy of breeding values in the range 9% - 25% compared to pedigree-based selection. Palaiokostas et al. (2018) showed an increase of 8% of the prediction accuracy using a GBLUP model with 9,195 markers compared to a PBLUP model for VNN resistance in sea bass. In this paper, they used the AUC (Area under Curve) as a metric for measuring accuracy. As our measure is very different, we cannot compare the results. However, we both showed that genomic selection outperformed pedigree-based selection to improve VNN resistance in sea bass. In Palaiokostas et al. (2016), the authors reached a prediction accuracy of 0.44 using a GBLUP model compared to 0.30 when using a PBLUP model for pasteurellosis resistance in gilthead seabream using 578 individuals in the training population and 200 individuals in the validation population. Here, even though the improvement of prediction accuracy using GBLUP was lower than in that previous study (+43% in Palaiokostas et al. (2016) versus +13% in our work), we obtained much higher absolute values (0.56 using PBLUP and 0.63 using GBLUP with 600 individuals in the training population). Such a difference could be explained by different mating design or genetic diversity within each population.

We showed that an increase in training population size led to an increase in prediction accuracy. Accuracy changed with training population size in a specific way in each data set. Population structure and heritability are major two drivers for the change in accuracy as function of the training population size (Goddard, 2009; Meuwissen et al., 2013). Even though cohorts VNN_A and VIB came from similar mating designs and were genotyped with the same markers, they had very different accuracy profiles (Figure 4.2a, Figure 4.2c). The absolute values of accuracy were different and the increase in accuracy

between 50 and 150 individuals in the training population was greater in VNN_A than in VIB. The heritability was very different between the two traits, as well as the average survival rate, and both have impact on the accuracy. Across all the data sets, even though accuracy increased less and less as the training population size increased, we did not reach the plateau and thus, adding more individuals from the same generation or of successive generations would still increase prediction accuracy (Figure 4.2). Another way to increase prediction accuracy at a constant training population size could be an optimized choice of the training population. By doing this using an optimization algorithm, the prediction accuracy could be significantly improved compared to a random selection of the training population (Akdemir et al., 2015; Rincint et al., 2012). However, the absolute size of the training population is the main driver compared to its composition (Bradford et al., 2017)

Across all diseases and species we studied, between 3K and 6K SNP markers were enough to obtain a high prediction accuracy. Similar results were found in other aquaculture species (Kriaridou et al., 2020). In Kriaridou et al. (2020), the authors found that a marker density between 1K and 2K was sufficient to keep the prediction accuracy close to that of the full density chip. In our study, such a density would lead to a significant decrease in prediction accuracy. An explanation could be that we compared the prediction accuracy of low-density SNP panels to a 43K or 44K SNP chip, while Kriaridou et al. (2020) compared the low-density SNP panels to a 10K SNP chip. If we compared the prediction accuracy of the 1K, 3K and 6K SNP chips to the 10K, the 3K chip maintained the accuracy to 95% of the value of the 10K and the 1K chip to 90% of the value of the 10K, which is close to the values obtained in the study of Kriaridou et al. (2020).

Training population size and marker density are two major cost drivers in a genomic selection breeding program (Rajacic et al., 2016; Riedelsheimer and Melchinger, 2013). Both impact the cost of genotyping, while training population size also involves the cost of phenotyping. We showed that a 6K to 10K SNP chip was enough to reach at least 90% of accuracy obtained with the full density chip (43-44K). Combining medium to high density genotyping on parents and low density genotyping followed by an imputation on the offspring can be a viable genotyping strategies to reduce the overall cost of the breeding program (Cleveland and Hickey, 2013; Tsai et al., 2017; Tsairidou et al., 2020). Both Tsai et al. (2017) and Tsairidou et al. (2020) showed a significant improvement of the prediction accuracy using imputed SNP data rather than low-density panels. Tsai et al. (2017) showed that the imputation from 256 SNP to 25K increased the prediction accuracy by 45% (from 0.4 to 0.58) for sea lice resistance in Atlantic salmon. Tsairidou et al. (2020) showed that genotyping the parents with a 5K chip and the offspring with only 200 SNP markers then imputing them to the 5K chip led to a prediction accuracy close to the value obtained by the medium density chip, while decreasing the genotyping cost of the breeding program by 62%. With a constant budget for genotyping and infection challenge, a breeder could increase the number of individuals in the training population and, thus, the prediction accuracy or, with the money saved, start other breeding programs on other traits.

In Griot et al. (2021), we showed that one strong effect QTL was involved in VNN resistance in European sea bass. By adding the information of the marker proposed as marker-assisted selection in the prediction model, we showed a significant increase in the prediction accuracy. The information of the marker genotype in a PBLUP model led to an accuracy slightly higher than that obtained with the 44K chip. In these populations and for this generation, marker-assisted selection would thus seem to be a very relevant choice. However, as mentioned in Griot et al. (2021), this result may not be consistent in other populations and/or generations as this marker was not in complete LD with the QTL. In addition, even though the marker LG12_8815613 had a strong effect in all the populations from Griot et al. (2021), it was selected partly from the same populations in which its effect was tested, this could lead

to an overestimation of its effect and thus, a greater accuracy than the one that could be expected in an unrelated population (Phocas et al., pers. Communication, 2020).

In this study, we presented a framework to implement genomic selection for disease resistance in European sea bass and gilthead sea bream. The results showed that, across all diseases and cohorts, 6,000 SNP markers were sufficient to get high prediction accuracy, equivalent to at least 90% of accuracy reached with the full density chip. For the training population size, as the plateau of accuracy was not reached with 800 individuals, an increase in its size would lead to an increase in accuracy and thus, in genetic gain. For VNN resistance, as one major effect QTL was detected, we showed that marker-assisted selection was an efficient method to improve the prediction accuracy.

CHAPITRE 5 :

DISCUSSION GENERALE

5.1. Résumé des principaux résultats

Jusqu'à présent, la sélection pour améliorer la résistance aux maladies en aquaculture est faite par sélection familiale sur collatéraux : des apparentés des candidats à la sélection sont challengés contre un pathogène puis, par génotypage sur un ensemble de quelques marqueurs microsatellites et assignation de parenté, les relations familiales sont établies et les valeurs génétiques des candidats sont estimées, avec des valeurs identiques pour tous les candidats d'une même famille de pleins-frères. Avec le développement des techniques de séquençage haut-débit et la baisse des coûts de production et d'utilisation d'outils génomiques, le génotypage sur des puces à SNPs de plusieurs dizaines de milliers de marqueurs a été rendu possible chez les espèces mineures d'élevage. Cette facilité d'accès au génotypage ainsi que les performances de la sélection génomique dans des espèces majeures d'élevage, en premier lieu les bovins, mais aussi le saumon en aquaculture, ont éveillé l'intérêt des sélectionneurs français de bar et de daurade, par le biais du SYSAAF, pour la sélection génomique.

L'objectif principal de cette thèse a été d'évaluer les gains permis par la sélection génomique et de développer un cadre méthodologique et des outils pour en faciliter sa mise en pratique dans les entreprises de sélection françaises.

Dans le **Chapitre 2**, nous avons développé un logiciel d'assignation de parenté, APIS, qui utilise des données de génotypage issus des puces à SNPs. Ce logiciel est basé sur une méthode d'assignation utilisant les probabilités mendéliennes de transmission observées dans une population, calculées à partir des données de génotypages, aussi bien par des marqueurs SNPs que microsatellites. La méthode a également été développée pour garantir un taux d'erreurs d'assignation inférieur à une valeur choisie par l'utilisateur. Nous l'avons testé dans des situations où l'assignation de parenté est difficile, c'est-à-dire lorsque le nombre de marqueurs est faible et qu'une partie des génotypes des parents des individus à assigner sont manquants. En comparaison avec deux autres logiciels d'assignation sur un même jeu de données, CERVUS qui utilise une approche de maximum de vraisemblance et VITASSIGN qui utilise une approche d'exclusion, APIS permet d'atteindre des taux d'erreurs d'assignations (faux positifs) plus faibles. Lorsque le nombre de marqueurs est suffisant pour atteindre une puissance d'assignation théorique de plus de 99,99%, tous les logiciels donnent des résultats fiables, avec des taux d'assignation élevés. Cependant, lorsque cette puissance diminue et que le taux de génotypes parentaux manquants augmente, APIS produit des résultats bien plus fiables que les autres logiciels, tout en garantissant un taux d'erreurs réel inférieur ou égale à celui qui est fixé par l'utilisateur. Nous avons pu montrer que APIS, avec un taux d'erreur maximum de 5% donne, quelle que soit la situation, un taux d'erreur observé inférieur à 5% tandis que CERVUS, lorsque le nombre de marqueurs est inférieur à 42 et que 10% des génotypes parentaux sont manquants, donne un taux d'erreur de plus de 10%, pouvant atteindre plus de 35% lorsque 50% des génotypes parentaux sont manquants. Nous avons également montré que le nombre d'individus à assigner a un effet sur la qualité de l'assignation avec APIS et qu'un minimum de 200 individus doivent être assignés pour garantir la fiabilité des résultats. Enfin, nous avons montré que le taux d'erreur de génotypage ne modifie que le taux d'assignation et que le taux d'erreurs d'assignation est toujours maintenu en deçà de la valeur choisie par l'utilisateur. APIS permet donc d'obtenir des résultats d'assignation fiables quelle que soit la qualité du génotypage (nombre de marqueurs génotypés et potentiels génotypes parentaux manquants).

Dans le **Chapitre 3**, nous avons présenté deux types de résultats. Le premier a été axé sur des outils, avec la validation de la qualité de la puce à SNPs pour le bar ThermoFisher Axiom™ DlabCHIP 57K, et la création d'une carte génétique haute densité. La puce, comportant 56 730 marqueurs, a été validée

dans deux cohortes commerciales de 1 089 et 1 110 individus respectivement. Dans ces deux cohortes, la proportion de marqueurs de haute qualité a été de 93,1% et de 94,4% avec un taux de génotypage moyen de 99,6% et 99,8% respectivement. Grâce à cette puce à SNPs, nous avons créé une carte génétique à partir d'une population composée de 1 334 individus issus de quatre familles de plein-frères et 880 individus provenant d'un plan de croisement factoriel de 94 pères par 39 mères. La carte génétique obtenue était composée de 49 638 marqueurs positionnés sur 23 241 positions uniques.

Dans ce même chapitre, nous avons étudié l'architecture génétique de la résistance à la nodavirose chez le bar, à la vibriose chez le bar et à la pasteurellose chez la daurade. A partir de quatre familles de back-cross composées de 378, 454, 291 et 211 individus issus de trois origines géographiques différentes et de deux cohortes commerciales de 1 027 et 476 individus, nous avons estimé l'héritabilité de la résistance à la nodavirose et détecté des QTLs associés à cette résistance avec différentes méthodes. L'héritabilité de la résistance à la nodavirose s'est montrée modérée, entre 0,24 et 0,38 avec un modèle à seuil. Nous avons détecté un total de sept QTLs avec trois méthodes différentes, localisés sur les LG3, LG8, LG12, LG14, LG15, LG19 et LG20. Le QTL localisé sur le LG12 a été détecté dans cinq des six populations étudiées et explique 9,2% de la variance génétique totale. Les autres QTLs n'expliquent que 1% ou moins de la variance génétique totale. Nous avons également attiré l'attention sur un marqueur particulier, LG12_8815613, qui présente des caractéristiques faisant de lui un candidat potentiel pour une sélection assistée par marqueurs. Ces résultats ont été confirmés chez trois autres cohortes commerciales de 703, 536 et 1 260 individus. Le QTL localisé sur le LG12 a été détecté dans deux des trois cohortes. En associant tous les jeux de données où ce QTL a été repéré, nous avons pu préciser sa position et la part de variance expliquée. L'intervalle de confiance du QTL a été réduit, passant de 3,45cM à 2,18cM, tout comme la part de variance génétique expliquée, qui a été estimée à 7,5%. De plus, nous avons confirmé l'intérêt du marqueur LG12_8815613 dans une sélection assistée par marqueurs.

L'architecture génétique de la résistance à la vibriose chez le bar et celle de la résistance à la pasteurellose chez la daurade sont bien différentes. Dans une cohorte commerciale de bar de 1 049 individus, nous avons estimé l'héritabilité à 0,18 avec un modèle à seuil. Quelle que soit la méthode employée, aucun QTL n'a pu être détecté. Chez la daurade, nous avons estimé l'héritabilité de la résistance à la pasteurellose à 0,28 dans une cohorte commerciale de 929 individus. Un seul QTL, localisé sur le chromosome 17 a été détecté et uniquement avec un modèle BayesCπ. Les résistances à la vibriose et à la pasteurellose sont deux caractères moins héréditaires que la résistance à la nodavirose et ont une architecture génétique polygénique.

Dans le **Chapitre 4**, nous avons évalué l'efficacité de la sélection génomique pour des caractères de résistance à différents pathogènes chez le bar et la daurade, en comparant la précision de sélection de la sélection génomique et celle de la sélection basée sur le pedigree pour différentes tailles de population de référence et densités de marqueurs. Pour deux cohortes commerciales de 1 027 et 1 042 bars challengés à la nodavirose, une cohorte commerciale de 1 049 bars challengés à la vibriose et une cohorte commerciale de 916 daurades challengées à la pasteurellose génotypés sur des puces 57K et 60K SNPs, nous avons généré des puces virtuelles basse densité par sélection de marqueurs, ainsi que différentes populations d'entraînement allant de 50 à 800 individus choisis au hasard. Nous avons montré que, quelle que soit la densité de marqueurs, la sélection génomique est supérieure à la sélection basée sur le pedigree. Lorsque la taille de la population d'entraînement et la densité de marqueurs sont maximales, la précision atteint 0,65 et 0,56 pour les deux cohortes challengées à la nodavirose, 0,51 pour celle challengée à la vibriose et 0,64 pour celle challengée à la pasteurellose. En général, la précision a augmenté avec la taille de la population d'entraînement et avec la densité de marqueurs. Nous avons également montré qu'une densité de 6K marqueurs permet d'atteindre au

moins 90% de la précision de la puce complète. Dans le cas de la nodavirose, nous avons montré que la prise en compte du marqueur LG12_8815613, repéré comme étant un candidat pour une sélection assistée par marqueurs, dans le modèle de prédiction permettait d'atteindre une précision supérieure à celle obtenue sans cette information. La précision d'une sélection basée sur le pedigree et avec uniquement l'information du marqueur et supérieure à celle d'une sélection génomique avec la puce complète. Dans le cas d'une sélection génomique, la prise en compte du génotype du marqueur LG12_8815613 permet d'augmenter la précision de 10,5%.

5.2. Applications des résultats dans les programmes de sélection

5.2.1. Optimisation de la sélection génomique

5.2.1.1. *De la précision au gain phénotypique*

L'objectif principal de cette thèse était d'établir un cadre pour mettre en place la sélection génomique pour améliorer la résistance aux maladies dans les programmes de sélection de bar et de daurade. Nous avons montré dans le chapitre 3 que la sélection génomique augmente la précision de sélection quelles que soient la maladie, la densité de marqueurs ou la taille de la population d'entraînement. Cependant, la précision de sélection n'est pas une valeur facilement appréhendable, en particulier lorsque l'on traite d'un caractère binaire comme c'est le cas avec la survie. Contrairement aux caractères continus, une augmentation de 20% de la précision n'entraîne pas une augmentation de la survie de 20%, en supposant que l'intensité de sélection, la variance génétique et l'intervalle de génération restent constants. Ainsi, pour mieux appréhender l'impact de la sélection génomique sur l'amélioration de la survie, nous allons convertir les valeurs de précision obtenues dans le chapitre 3 en gains de survie.

Pour cela, nous nous sommes placés dans un cas simple. Nous avons supposé une pression de sélection de 20%, une incidence équivalente au taux de survie observée dans les différents challenges et une égalité des précisions obtenues entre un modèle linéaire et un modèle à seuil. Pour justifier cette dernière hypothèse, nous avons estimé les valeurs génétiques de la cohorte VNN_1 avec un modèle à seuil et avec un modèle linéaire. La corrélation entre les deux estimations était de 0,9991 (Figure 5.1).

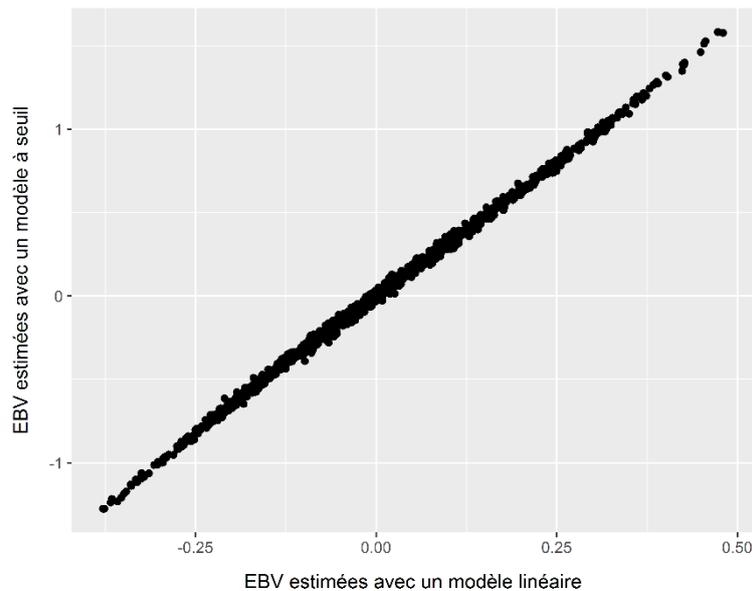


Figure 5.1 : Relation entre les valeurs génétiques estimées pour la résistance à la nodavirus dans une cohorte de 1027 bars européens avec un modèle linéaire et avec un modèle à seuil.

Pour convertir une valeur de précision en gain de survie il faut dans un premier temps estimer le gain génétique réalisé sur l'échelle sous-jacente (Dempster and Lerner, 1950) puis exprimer cette progression de la moyenne des valeurs génétiques sur l'échelle sous-jacente en une valeur sur l'échelle observée grâce à la loi normale.

Ainsi :

$$\Delta G = i * r * \sigma_{g_u}$$

Avec ΔG le gain génétique sur l'échelle sous-jacente, i l'intensité de sélection pour une pression de 20% fixée à 1.4, r la précision de sélection estimée par un modèle linéaire et σ_{g_u} l'écart-type génétique sur l'échelle sous-jacente estimée par un modèle à seuil. Ici, σ_{e_u} a été fixée à 1 donc $\sigma_{g_u} = h$.

Le gain de survie est la traduction du gain génétique réalisée sur l'échelle sous-jacente en un gain phénotypique réalisé sur l'échelle [0, 1] et est calculé par :

$$\Delta P = F(F^{-1}(Sn) + \Delta G) - Sn$$

Avec ΔP le gain de survie, exprimée comme la différence entre la survie estimée entre deux générations consécutives, F la fonction de répartition de la loi normale $N(0, \sigma_{p_u})$ et F^{-1} son inverse et Sn le taux de survie à la génération.

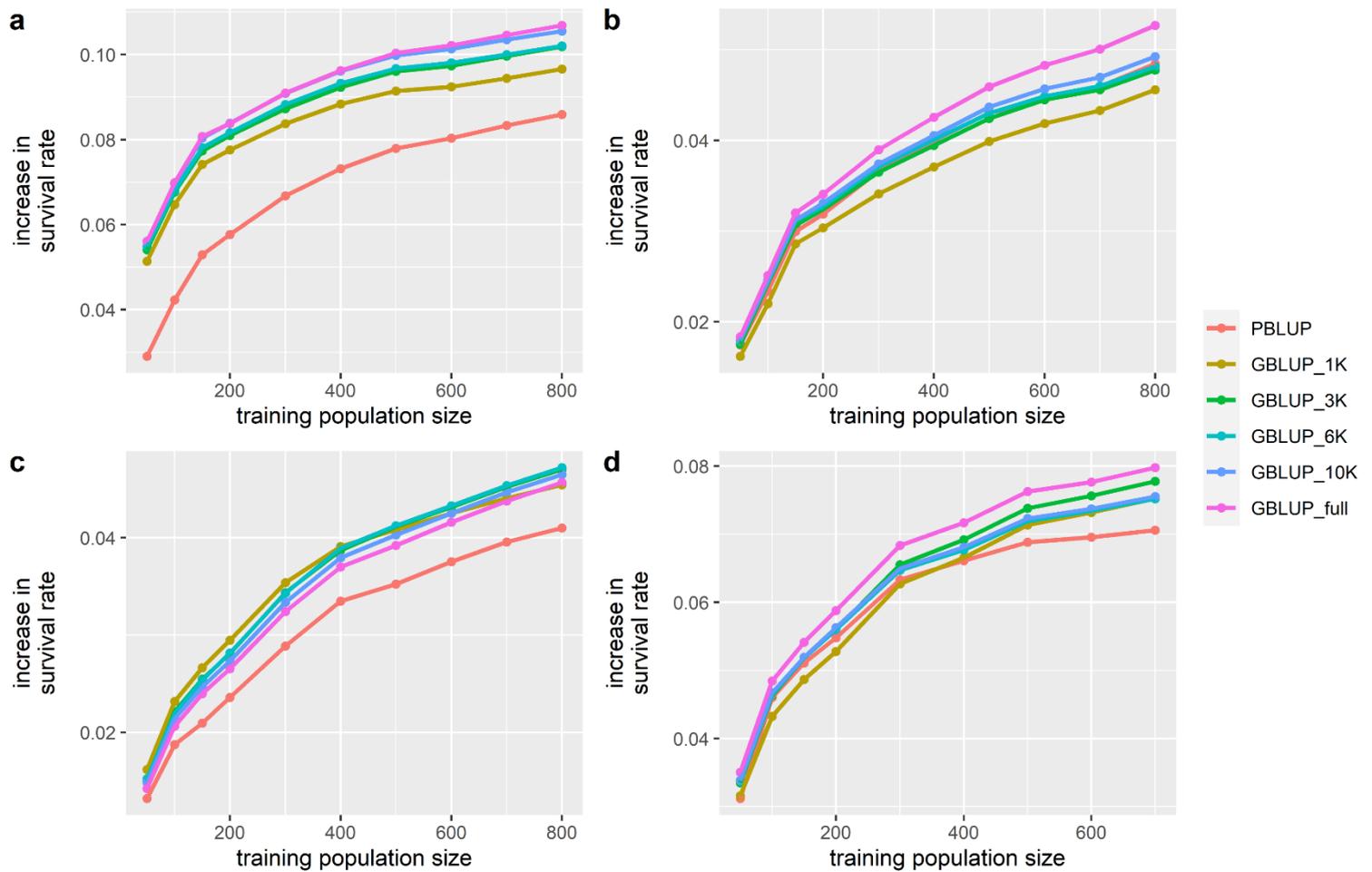


Figure 5.2 : Gain phénotypique relatif estimé pour une population d'entraînement de 50 à 800 individus et pour différentes densités de marqueurs dans deux cohortes de bars challengés au nodavirus (a, b), une cohorte de bars challengés à *V. harveyi* (c) et une cohorte de daurade challengé à *Photobacterium damsela subsp. piscicida* (d). La palette de couleur représente les différentes densités de marqueurs.

Il n'y a pas de différences de classement entre les différentes densités de marqueurs par rapport à la Figure 4.3 du Chapitre 3. Cependant, les gains de survie sont bien plus faibles que les gains de précision estimés auparavant. En général, il y a moins de 3% de gain de survie entre la puce 1K et la puce complète. Les seuls écarts clairement visibles sont pour les deux cohortes challengées au nodavirus. Pour la cohorte VNN_A (Figure 5.2a), l'écart entre la puce 1K et les autres est d'environ 1% tandis que dans la cohorte VNN_B (Figure 5.2b), la puce 1K donne un gain de survie inférieur à une sélection sur pedigree.

Pour observer l'effet de la densité de marqueurs sur le gain phénotypique sur plusieurs générations, nous avons simulé 10 générations. Le taux de survie initial était celui du challenge et, à chaque génération, il était estimé comme précédemment. Nous avons également supposé que la précision restait constante au cours des générations, ce qui présuppose un maintien du taux de survie au challenge au alentours de 50% par l'utilisation de pré-tests. Pour prendre en compte un effet de la sélection sur la variance génétique, nous avons appliqué un effet Bulmer (Bulmer, 1971). L'effet Bulmer permet de calculer la modification de la variance génétique due à la sélection. Tout d'abord, on calcule la variance génétique estimée au sein des individus sélectionnés (d'après Falconer and Mackay (1996)) :

$$Vg^* = Vg_n * (1 - (i(i - x) * r^2))$$

Avec Vg^* la variance génétique au sein des individus sélectionnés, Vg_n la variance génétique à la génération n , i l'intensité de sélection ici fixée à 1.4 pour une pression de sélection de 20%, x le point de troncature de la loi normale $N(0, 1)$ en $1 - p$, avec p la pression de sélection, ici de 20% et r la précision de sélection.

Puis, on calcule la variance génétique estimée à la génération suivante par (d'après Falconer and Mackay (1996)) :

$$Vg_{n+1} = \frac{1}{2}Vg^* + \frac{1}{2}Vg_n$$

Seule une taille maximale de population d'entraînement a été testée (N=800 pour le bar, N=700 pour la daurade). Tous les autres paramètres sont restés identiques.

Ces résultats montrent un avantage de la sélection génomique sur la sélection sur pedigree d'environ une génération (Figure 5.3). Par exemple, dans la population VNN_1 (Figure 5.3a), la survie atteint 95% en 4 générations avec une sélection sur pedigree (PBLUP) et en 3 générations seulement avec la sélection génomique (GBLUP). Chez VIB et PAS (Figure 5.3c, Figure 5.3d), on atteint 95% de survie en 5 générations en sélection sur pedigree et 4 en sélection génomique. En revanche, dans la cohorte VNN_2, tous les scénarios sont identiques et une survie de 95% est atteinte après 4 générations de sélection. Il convient de noter que le nombre de générations pour atteindre 95% de survie est dépendant du niveau de survie initial. Cependant, la différence entre la puce 6K et la puce complète est minime dans tous les scénarios.

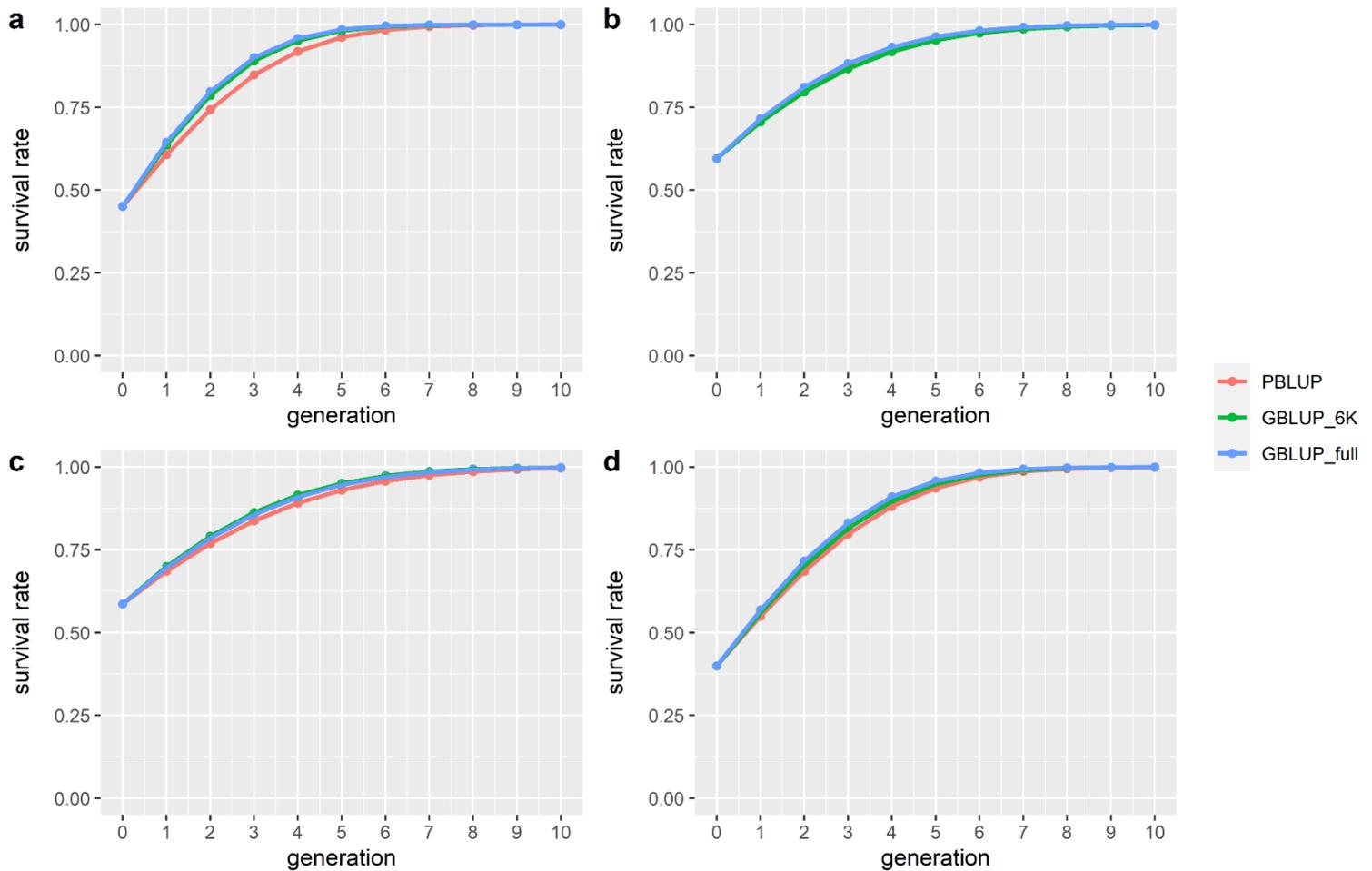


Figure 5.3 : Simulation du taux de survie pour 10 générations de sélection pour différentes densités de marqueurs dans deux cohortes de bars challengés au nodavirus (a, b), une cohorte de bars challengés à *V. harveyi* (c) et une cohorte de daurade challengé à *Photobacterium damselae subsp. piscicida* (d). La taille de la population d'entraînement était de 800 individus pour les cohortes de bars et 700 pour celle de daurades. La palette de couleur représente les différentes densités de marqueurs.

Enfin, dans le cas de la résistance à la nodavirose, nous avons estimé le gain phénotypique sur 12 générations en prenant en compte l'ajout de l'effet du QTL dans les modèles de prédiction PBLUP et GBLUP. Le gain phénotypique a été estimé de la même manière que précédemment.

L'ajout de l'effet du QTL dans le modèle de prédiction permet d'accélérer le progrès génétique d'environ une génération quelle que soit la densité de marqueurs (Figure 5.4). L'avantage de la prise en compte du QTL est plus visible dans un modèle PBLUP et peut atteindre un gain de deux générations dans la cohorte VNN_A (Figure 5.4a, panneau gauche). Dans un modèle GBLUP, l'effet du QTL sur le gain phénotypique se réduit avec la densité de marqueurs. Quelle que soit la méthode employée pour prendre en compte le QTL dans la prédiction des valeurs génétiques des candidats à la sélection, le génotypage du marqueur LG12_8815613 n'entraîne pas de surcoût car il sera présent sur les puces basses densités ou les panels d'assignation qui pourront être développés. En revanche, sur le long terme, la prise en compte du marqueur LG12_8815613 pourrait diminuer l'efficacité de la sélection et donc ralentir le progrès génétique. En effet, avec une sélection assistée par marqueur, la fréquence de ce marqueur va rapidement augmenter dans les populations en sélection jusqu'à se fixer (fréquence allélique = 1). A partir de ce stade, voire même lorsque la fréquence allélique sera élevée, sélectionner des candidats à l'aide de ce marqueur ne permettra plus d'augmenter la résistance (Villanueva et al., 1999). Comme il explique moins de 10% de la variance génétique, le gain génétique passera par une

sélection sur la partie polygénique du caractère et donc un passage à la sélection génomique tout en s'assurant du maintien du QTL dans les populations en sélection. De plus, au cours des générations de sélection, le déséquilibre de liaison entre le marqueur et le QTL peut se réduire, ce qui va diminuer la précision de sélection et donc, l'efficacité d'une sélection assistée par marqueurs (Lande and Thompson, 1990). A ce stade, nous n'avons pas d'estimation de la valeur du déséquilibre de liaison entre le marqueur LG12_8815613 et le QTL de résistance à la nodavirus. Pour éviter cette situation, il convient de trouver un ou des marqueurs très proches du QTL et qui seraient en très fort déséquilibre de liaison. Une recombinaison entre le marqueur et le QTL serait alors très peu probable et maintiendrait l'efficacité de la sélection sur ce QTL jusqu'à la fixation de celui-ci dans les populations en sélection.

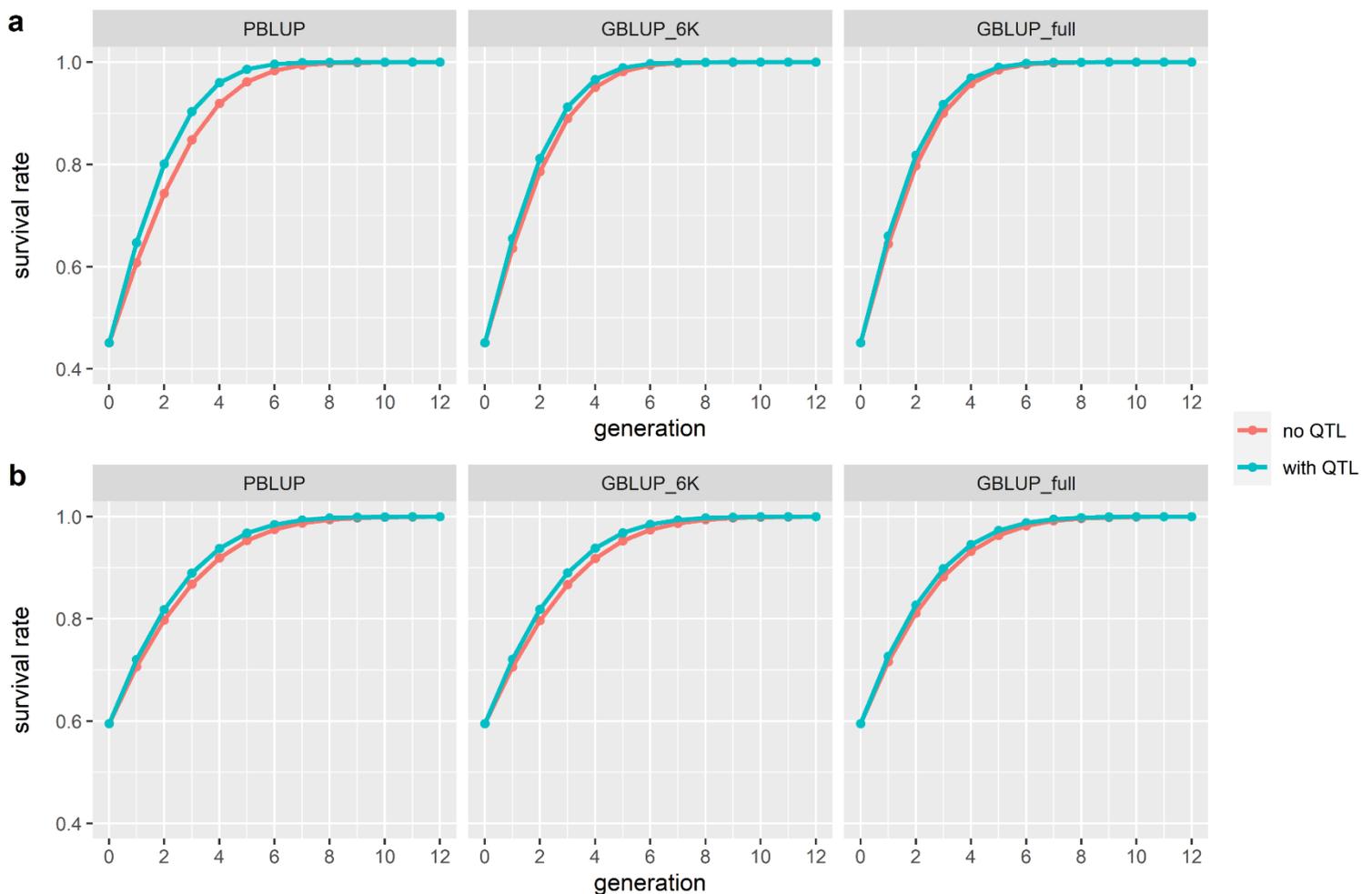


Figure 5.4 : Simulation du taux de survie pour 12 générations de sélection pour différentes densités de marqueurs et en prenant en compte (bleu) ou non (rouge) l'effet du QTL de résistance dans deux cohortes de bars challengés au nodavirus (a, b). La taille de la population d'entraînement était de 800 individus.

Quel que soit le modèle utilisé, la densité de marqueurs ou la taille de la population d'entraînement, l'utilisation de pré-tests, permettant d'obtenir une mortalité aux alentours de 50%, est nécessaire pour maximiser la précision estimée à partir des données issues du challenge. Les conditions dans lesquelles se déroulent les challenges infectieux évolueront avec le progrès réalisé sur les populations et permettront un maintien du gain génétique au cours des générations de sélection.

5.2.1.2. Quels paramètres pour optimiser la sélection génomique ?

Avec ces résultats supplémentaires, nous confirmons qu'une puce basse densité entre 3K et 6K marqueurs est suffisante pour obtenir un gain de survie proche de celui obtenu avec une puce d'environ 50K marqueurs. Cette puce basse densité permettait d'obtenir une précision d'au moins 90% de celle de la puce complète et permet d'obtenir un gain phénotypique équivalent. De plus sur plusieurs générations, le gain phénotypique entre les deux puces est quasi identique. En aquaculture, le nombre de marqueurs permettant d'atteindre une précision équivalente à au moins 90% de celle d'une puce haute densité (environ 50K marqueurs) est variable selon les espèces et les caractères, et est estimé se trouver entre 1 000 à 5 000 marqueurs selon les études (Kriaridou et al., 2020; Tsai et al., 2016). Quel que soit ce chiffre, il montre que des puces basses densités permettent des précisions proches d'une puce moyenne densité, ce qui doit permettre de réduire les coûts de la sélection génomique chez ces espèces. En revanche, les précisions obtenues avec des puces très haute densité, avec un nombre de marqueurs allant de 300K à 600K ne sont pas connues et pourraient démontrer qu'une puce 50K serait le meilleur outil pour obtenir une précision élevée au coût le plus bas. C'est le cas chez les bovins laitiers, chez qui la puce 50K offre une précision proche de celle d'une puce 500K (moins de 2% de différence) d'après des simulations faites par VanRaden et al. (2011). A l'inverse, la différence de précision entre une puce 50K et une puce 3K est bien plus importante, de l'ordre de 20%. Des résultats similaires ont été rapportés chez le mouton (Moghaddar et al., 2017).

Un autre paramètre important est le nombre d'individus qui composent la population d'entraînement. Dans le chapitre 3, nous avons étudié son influence sur la précision de sélection. Bien que nous ayons constaté que plus la population d'entraînement était grande, plus la précision était élevée, nous n'avons pas observé de stagnation de la précision à partir d'une certaine taille de population d'entraînement (Brard and Ricard, 2015). Cela laisse supposer qu'une augmentation de la taille de la population d'entraînement au-delà de 800 individus devrait encore augmenter la précision. Cependant, nous avons vu que malgré une différence d'environ 0,03 entre la précision obtenue par la puce complète et la puce 6K, il n'y a pas de claires différences entre les gains de survie obtenus par les deux puces. Ainsi, pour les cohortes challengées au nodavirus et à *Photobacterium damsela subsp. piscicida*, où l'augmentation de la précision avec la taille de la population d'entraînement commence à ralentir, une population d'entraînement d'au moins 600 individus devrait être suffisante pour obtenir un gain phénotypique quasi identique celui obtenu avec une population d'entraînement plus importante. Dans le cas de la résistance à la vibriose, la précision semble toujours augmenter de manière linéaire au-delà de 600 individus dans la population d'entraînement. Il serait préférable d'évaluer la précision de sélection avec une population d'entraînement plus importante pour identifier à partir de quelle taille la stagnation est atteinte et ensuite donner des recommandations quant à un optimum. Cette recommandation de 600 individus n'est valable que dans le cas qui nous a intéressé au cours de cette thèse et ne sera probablement valable que pour quelques générations. On peut voir que ce chiffre est très variable selon les caractères et les populations étudiées puisque dépendant de la taille de population efficace et de l'héritabilité du caractère (Goddard, 2009) mais aussi de la quantité d'information disponible. Chez les bovins, l'optimum se situe aux alentours de 5 000 individus dans la population de référence (Hayes et al., 2009; Liu et al., 2011). En aquaculture, la majorité des résultats proviennent d'études réalisées chez les salmonidés, et montrent des optimums allant de 1 000 à 5 000 individus (Ødegård et al., 2014; Liu et al., 2011). Dans la majorité des cas, les auteurs insistent sur le fait qu'une augmentation du nombre d'individus pourrait entraîner un gain de précision du fait du faible nombre d'individus utilisé lors de l'étude.

Un autre facteur ayant démontré son importance dans cette thèse est la structuration éventuelle de la population, en l'occurrence celle qui est composée des candidats et de leurs collatéraux. Dans cette thèse, seule la cohorte nommée popB dans le chapitre 2 ou VNN_B dans le chapitre 3 a présenté une

structuration, c'est à dire la présence de sous-populations au sein de la population étudiée dans laquelle les fréquences alléliques ne sont pas semblables à celles estimées dans la population globale. Grâce à une analyse en composantes principales (ACP) à partir des génotypes, on peut observer des individus formant des clusters qui sont éloignés du reste de la population, identifié par des cercles rouges sur la Figure 5. Dans le chapitre 2, nous avons vu que cette structuration masquait l'architecture génétique de la résistance à la nodavirose et empêchait donc la détection de QTL. En effet, en supprimant les individus qui engendraient cette structuration, nous avons pu mettre en évidence des QTLs associés à la résistance à la nodavirose. En sélection génomique, avec cette population, la précision obtenue par la puce 1K est inférieure à celle obtenue avec le pedigree, qui est elle-même équivalente à celle obtenue avec la puce 3K. Cette cohorte est la seule qui présente un cas où la sélection génomique est moins performante que la sélection basée sur le pedigree. Et, en effet, la structuration observable dans le chapitre 2 ne l'était qu'à partir de 3K marqueurs. Le pedigree était donc plus performant pour informer de la structure de la population que la matrice d'apparentement génomique estimée avec 1 000 marqueurs.

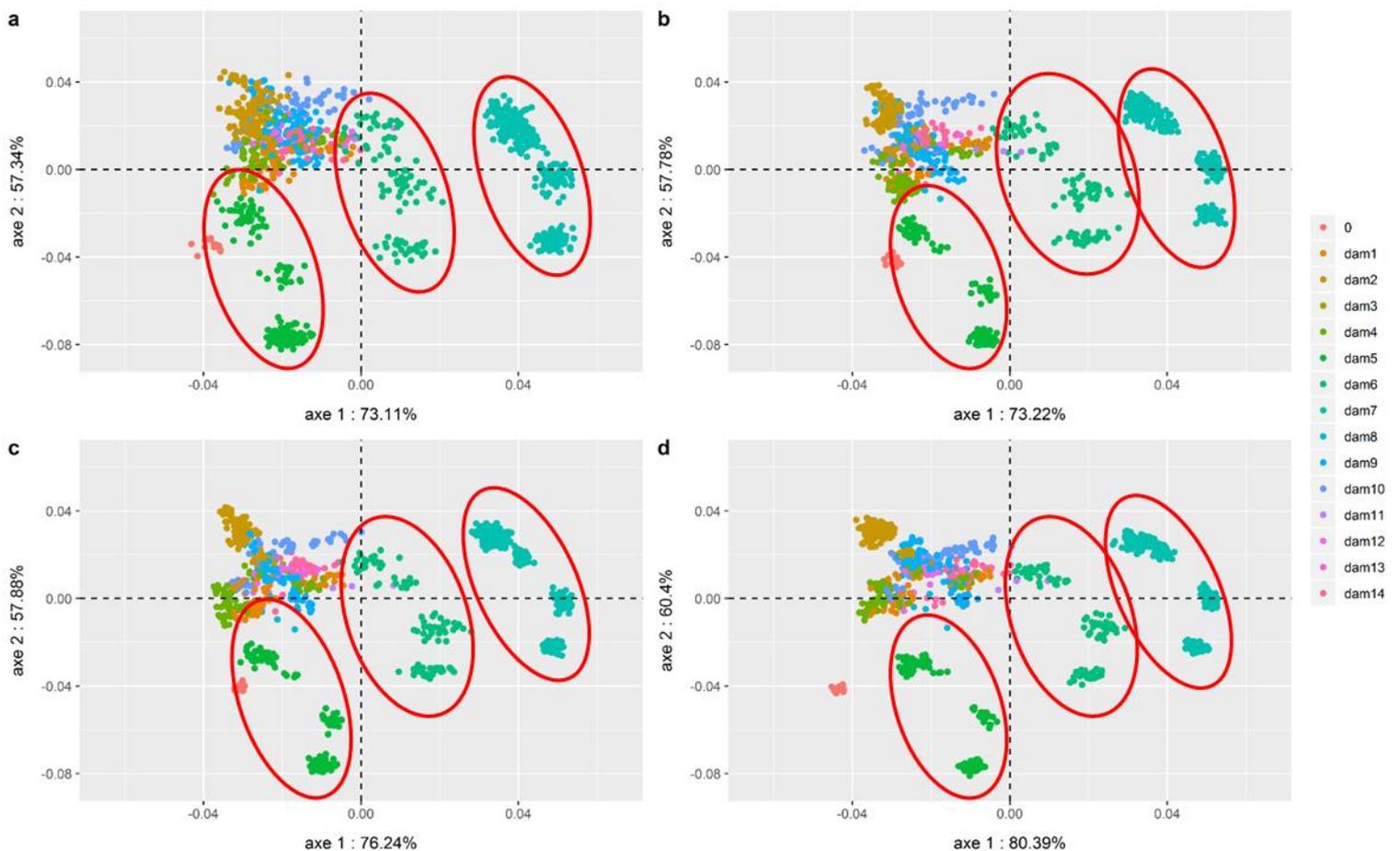


Figure 5.5 : Analyses en Composantes Principales (ACP) de 1 042 bars européens génotypés sur 1K (a), 3K (b), 6K (c) et 44K (d) marqueurs SNPs. La palette de couleurs représente les mères des individus. Les cercles rouges indiquent des clusters différents de la population globale.

Sur la Figure 5.5, on peut voir que les clusters représentant les individus issus de la mère « dam5 » et de la mère « dam7 » se séparent plus nettement du cluster principal à partir de 3K marqueurs (Figure 5.5b, cercle rouge) et sont complètement distincts à partir de 6K marqueurs (Figure 5.5c). Avec 1K marqueurs les clusters formés par les individus provenant de ces deux mères sont plus ou moins

confondus avec la population générale, ce qui doit expliquer en partie la moindre précision obtenue avec cette puce par rapport à celle obtenue avec le pedigree.

5.2.1.3. *Accumuler l'information sur plusieurs générations*

Dans la majorité des espèces sélectionnées, les phénotypes acquis sur les générations précédentes sont utilisés lors de l'estimation des valeurs génétiques de la génération en cours (Lewis and Brotherstone, 2002; VanRaden et al., 2004). Dans cette thèse, nous avons uniquement utilisé les phénotypes acquis sur la génération en cours pour estimer la précision de sélection dans différents scénarios. L'accumulation de phénotypes et de génotypes sur différentes générations pourrait améliorer la précision de sélection. Dans Sonesson and Meuwissen (2009), les auteurs montrent que la précision de sélection augmente avec l'ajout de phénotypes et génotypes pendant quatre générations avant d'atteindre un plateau. Ces résultats sont issus de simulations, dans lesquelles 100 mâles sont croisés avec 100 femelles pour créer 100 familles de plein-frères, les mâles et les femelles étant choisis soit au hasard soit selon leur valeur génétique parmi 3 000 candidats. Ce schéma de croisement, en familles de pleins frères ne ressemble pas au plan factoriel qui est couramment utilisé dans les entreprises de sélection françaises. Les résultats peuvent donc être différent de ceux obtenus dans les plans de croisement factoriel qui sont plus performants pour estimer les paramètres génétiques (Dupont-Nivet et al., 2006) et pour estimer les valeurs génétiques (Sonesson and Ødegård, 2016). Dans Weng et al. (2016), les auteurs montrent des résultats similaires chez la poule pondeuse. La précision de sélection augmente avec le nombre de générations utilisées dans l'estimation des valeurs génétiques et stagne au-delà de six générations. L'utilisation de générations récentes contribue fortement à l'augmentation de la précision tandis que les plus anciennes contribuent peu voire peuvent impacter négativement la précision. Les auteurs concluent que le nombre d'individus dans la population d'entraînement est un paramètre majeur influençant la précision de sélection plus que le nombre de générations. Ainsi, l'objectif premier doit être d'acquérir un maximum de phénotypes et de génotypes afin d'établir une population d'entraînement suffisamment importante pour atteindre une précision élevée. De nombreuses études ont évalué l'impact du nombre de générations incluses dans le pedigree sur la précision de sélection aussi bien dans des jeux de données provenant d'espèces d'élevage ou de simulations. Toutes ont trouvé des résultats semblables. Au plus le pedigree est profond, au plus la précision est élevée (Lourenco et al., 2014; Neuner et al., 2009). Ne garder que la dernière génération ne permet pas d'atteindre la précision maximale car il n'y a pas assez d'individus dans la population de référence. Il est possible de supprimer les générations les plus éloignées de la génération à évaluer car elles apportent moins d'information (plus distantes de la génération en cours et ne représentent qu'un faible proportion de la population de référence globale) et augmentent le temps pour estimer les valeurs génétiques (Lourenco et al., 2014). Bien que ces tendances aient été observées chez différentes espèces, le nombre de générations à garder semble dépendre de l'espèce et du caractère d'intérêt.

5.2.1.4. *Quel phénotype ?*

Tout au long de cette thèse, nous n'avons traité que la survie comme phénotype de résistance à une maladie et ce pour plusieurs raisons. Tout d'abord, c'est le phénotype qui intéresse les sélectionneurs, l'objectif étant de produire des poissons qui survivent à une infection par un pathogène. De plus, c'est un caractère facilement mesurable, largement utilisé dans ce type d'analyse et qui facilite donc les comparaisons de résultats.

Dans les challenges qui ont été réalisés dans le cadre du projet Gènesea, le jour de mort a également été enregistré. Il est mesuré comme l'intervalle de temps entre l'infection et la mort de l'individu. Bien que ce phénotype donne une image plus précise de la réaction des individus challengés au pathogène testé, nous l'avons écarté de l'étude. En effet, l'objectif était de produire des poissons résistants au pathogène et non pas des poissons qui résistent plus longtemps à l'infection avant de mourir. Ensuite, l'analyse de ce caractère suppose l'utilisation de modèles plus complexes, comme le modèle de Cox ou de Weibull, qui ne sont pas utilisés en routine lors des analyses génétiques (Fraslin et al., 2018). Enfin, l'héritabilité du caractère de survie binaire est généralement plus élevée que celle du caractère jour de mort, ce qui encourage son utilisation en sélection (Antonello et al., 2009; Taylor et al., 2009).

Le caractère jour de mort possède néanmoins un avantage sur le caractère survie binaire lorsque la mortalité au challenge est très élevée. Le caractère binaire, ici la survie noté 0 ou 1, fonctionne bien lorsque la mortalité au challenge se situe entre 20% et 80%, avec un optimum à 50% où l'héritabilité sur l'échelle observée et sur l'échelle sous-jacente (continue) sont les plus proches. En général, la mortalité dans un challenge est un paramètre qui peut être modifié. Avant l'épreuve infectieuse, différentes concentrations en pathogène sont testés sur un échantillon des individus qui seront challengés. Cette étape, appelée pré-test, permet de trouver la concentration permettant d'obtenir une mortalité proche de 50% et est appliquée en routine. Cela a été le cas dans les challenges qui ont été étudiés dans cette thèse. Dans certain cas, la mortalité observée lors de challenges infectieux en aquaculture est au-delà de ces limites et l'utilisation de caractère jour de mort permet d'obtenir de meilleures estimations. C'est le cas pour la résistance au WSSV chez la crevette où la survie à l'issue du challenge est généralement inférieure à 15% et où des modèles utilisant le jour de mort donnent de meilleures estimations des paramètres génétiques ($h^2=0,21$ contre 0,01 pour la survie binaire) (Gitterle et al., 2006).

5.2.1.5. Utilisation des QTLs en sélection

La connaissance de l'architecture génétique des caractères permet d'améliorer l'efficacité de la sélection. Lorsqu'un caractère est polygénique, un modèle GBLUP permet d'obtenir une précision élevée (Daetwyler et al., 2010). Cela est principalement dû au fait que le modèle GBLUP est basé sur une hypothèse polygénique, où tous les marqueurs contribuent avec un effet faible à l'estimation de la valeur génétique (Daetwyler et al., 2010). Lorsqu'un ou quelques QTLs à effets modérés ou forts sont associés au caractère sélectionné, alors une sélection assistée par marqueurs ou d'autres modèles peuvent être plus efficaces. Dans le chapitre 2, nous avons montré qu'un QTL à effet fort était impliqué dans la résistance à la nodaviriose chez le bar. Dans le chapitre 3, nous avons montré que la prise en compte du génotype à un marqueur associé à cette résistance permettait d'améliorer la précision de la sélection, démontrant l'efficacité d'une sélection assistée par marqueurs.

Cependant, l'allèle de résistance est en faible fréquence dans les cohortes étudiées aussi bien dans la descendance que chez les parents. Dans la cohorte VNN_1, la fréquence de l'allèle de résistance était de 0,13, de 0,08 chez les pères et 0,2 chez les mères. Dans la cohorte VNN_2, elle était de 0,07 dans la descendance, 0,12 chez les pères et 0,07 chez les mères. Une trop forte pression de sélection sur les individus porteurs de l'allèle de résistance pourrait entraîner une hausse de la consanguinité avec tous les effets indésirables liés (diminution de la variance génétique, dépression de consanguinité, hausse des combinaisons alléliques délétères) (Kincaid, 1983). A l'inverse, dans Villanueva et al. (1999), les auteurs montrent par le biais de simulations, que l'existence d'un QTL à effet fort impliqué dans un caractère, entrainera une hausse de sa fréquence dans la population par sélection, qu'il soit pris en compte ou non dans un modèle d'évaluation génétique. La prise en compte du QTL dans le modèle

accélèrera la fixation du QTL au sein de la population et aussi le gain génétique. Pour éviter une hausse incontrôlée de la consanguinité, les auteurs proposent d'optimiser les croisements pour obtenir un gain génétique maximal avec une consanguinité fixée à une valeur choisie, dans leur cas de 3% par génération.

Pour éviter une baisse des performances à cause d'une hausse de la consanguinité, il convient de maintenir une diversité génétique en évitant des croisements d'apparentés pour augmenter la fréquence du QTL au sein des géniteurs. Une stratégie pouvant être mise en place pourrait s'inspirer de ce qui est fait dans le cas de la sélection assistée par marqueurs pour la résistance à l'IPN chez le saumon. Sur ce caractère, un QTL à effet majeur a été identifié (Houston et al., 2008; Moen et al., 2009). Ce QTL a un effet additif avec un allèle de sensibilité noté *q* et un allèle de résistance *Q* et explique plus de 80% de la variance génétique (Moen et al., 2009). Dès sa découverte, ce QTL a été utilisé en sélection assistée par marqueurs dans les programmes de sélection de saumon Atlantique de la manière suivante. Des mâles ont été sélectionnés par sélection assistée par marqueurs dans le but de créer des individus résistants à l'IPN porteurs d'un génotype *QQ*. Ces mâles sont ensuite croisés avec des femelles qui ne sont pas sélectionnées sur ce QTL et qui peuvent donc présenter les trois génotypes possibles. Dans la descendance, qui sera commercialisée, les individus sont par conséquent soit *qQ* soit *QQ* (Storset et al., 2015), et sont résistants du fait du caractère dominant de l'allèle *Q*. Une stratégie analogue pourrait être mise en place après confirmation ou découverte de marqueurs associés au QTL de résistance à la nodavirose dans les programmes de sélection de bar. Il est important de préciser que le marqueur LG12_8815613, utilisé comme covariable dans nos modèles de prédiction, est le meilleur que nous ayons trouvé mais possède néanmoins de nombreux inconvénients pour une utilisation directe, comme un déséquilibre de liaison imparfait avec le QTL et un effet significatif dans seulement 5 des 8 populations étudiées.

5.2.2. Comment mettre en place la sélection génomique ?

5.2.2.1. Prendre en compte les corrélations génétiques

Dans les programmes de sélection, plusieurs caractères sont sélectionnés. La prise en compte des corrélations génétiques entre les différents caractères est donc essentielle pour estimer le progrès génétique qui peut être attendu. Dans le chapitre 3, nous avons montré que la résistance à la vibriose chez le bar et la résistance à la pasteurellose chez la daurade sont corrélées positivement avec le poids (0,56 pour la corrélation entre la résistance à la vibriose et le poids et 0,89 pour celle entre la résistance à la pasteurellose et le poids). Cela indique qu'une sélection pour la croissance, qui est un objectif majeur des programmes de sélection, entraînera une amélioration indirecte de la résistance à ces maladies. En revanche, la résistance à la nodavirose chez le bar est corrélée négativement au poids (-0,35) ce qui aura l'effet inverse. La sélection sur la croissance aura tendance à diminuer la résistance des candidats. Dans Sonesson et al. (2011), les auteurs ont montré qu'en appliquant une sélection en deux étapes, d'abord une sélection sur la résistance aux maladies par challenge puis une sélection sur la croissance réalisée sur les survivants, il est possible d'améliorer les deux caractères malgré une corrélation négative. Pour cela, l'héritabilité de la résistance à la maladie sélectionnée doit être modérée à forte. C'est ce que nous avons obtenu pour la résistance à la nodavirose, qui a une héritabilité estimée entre 0,24 et 0,38 avec un modèle à seuil. De plus, les auteurs proposent l'utilisation de données génomiques pour estimer les valeurs génétiques et optimiser les plans de croisement dans le but d'améliorer le gain génétique obtenu. Enfin, ils proposent de modifier les index de sélection combinant survie et croissance pour permettre un progrès sur les deux caractères.

Dans le cas de la nodavirose, la présence d'un QTL à effet fort pourrait permettre une introgression de ce QTL dans des lignées sélectionnées pour la croissance pour obtenir des lignées performantes pour les deux caractères. Dans Ødegård et al. (2009), les auteurs ont évalué les performances, par simulation, d'une sélection massale, d'une sélection génomique et d'une sélection assistée par marqueurs dans l'introgression d'un QTL majeur impliqué dans la résistance à une maladie dans une lignée sélectionnée pour un caractère de production par rétro-croisement. Ils ont montré que seule une sélection utilisant l'information génomique, que ce soit la sélection génomique ou la sélection assistée par marqueurs, permettait une introgression du QTL dans la lignée sélectionnée pour le caractère de production. Bien que la différence entre les deux méthodes soit faible, la sélection assistée par marqueurs permettait un traçage du QTL tout au long du processus d'introgression. Ainsi, les auteurs recommandent une sélection génomique couplée avec une sélection assistée par marqueurs pour suivre l'introgression du QTL. Toutefois, il convient de valider que le QTL n'affecte pas négativement la croissance des individus porteurs.

Les corrélations génétiques estimées entre les différentes résistances et le poids interrogent sur la prise en compte du poids dans les analyses génétiques, soit en tant que covariable soit comme deuxième phénotype dans un modèle à deux caractères, pour mieux prédire les valeurs génétiques ou mieux localiser les QTLs impliqués, voire en détecter de nouveaux dans les résistances aux maladies. En général, l'utilisation d'un modèle à deux caractères, dont la corrélation génétique est élevée et dont le caractère à prédire est peu héritable, permet d'augmenter la précision de sélection sur le caractère peu héritable (Calus and Veerkamp, 2010; Jia and Jannink, 2012; Villanueva et al., 1993). En revanche, d'autres études ont montré qu'une évaluation multi-caractère, même lorsque les corrélations entre les caractères sont élevées, n'augmente pas la précision de sélection par rapport à une sélection mono-caractère (Joshi et al., 2020). Le poids avant challenge est mesuré systématiquement, sa prise en compte dans une sélection multi-caractère pourrait permettre d'augmenter la précision de la sélection sur la résistance aux maladies.

En GWAS, l'utilisation d'une approche bi-caractères lorsque les deux caractères sont corrélés permet principalement de détecter des QTLs communs et ainsi de mieux appréhender les causes génétiques de la corrélation observée (Bonnemaijer et al., 2019; Jiang and Zeng, 1995). Dans cette thèse, le choix de ne pas prendre en compte le poids dans les analyses s'est principalement justifié par le fait que le poids n'a pas été pris en compte dans d'autres études réalisées sur l'étude de l'architecture de la résistance à la nodavirose et à la pasteurellose ; et ainsi permettre des comparaisons. De plus, le poids au challenge n'avait pas d'effet significatif sur la résistance à la nodavirose d'après une estimation faite avec un modèle linéaire. Enfin, des GWAS avec ou sans l'ajout du poids en covariable ont été réalisées pour la résistance à la vibriose et à la pasteurellose où le poids avait un effet significatif sur la résistance et aucune différence n'a pu être observée (Figure 5.6).

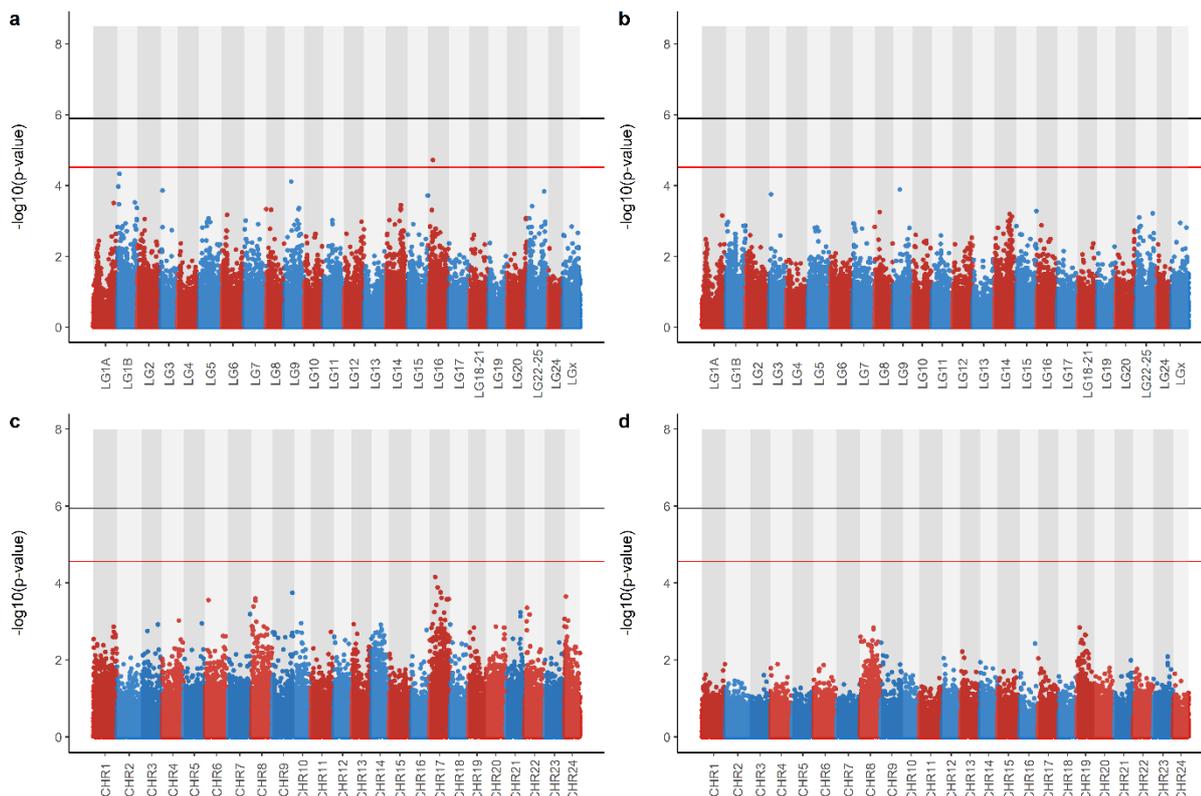


Figure 5.6 : Manhattan plot de GWAS pour la résistance à la vibriose estimée sans (a) ou avec (b) l'effet du poids en covariable dans le modèle dans une cohorte de 1 049 bars européens et pour la résistance à la pasteurellose estimée sans (c) ou avec (d) l'effet du poids en covariable dans un modèle dans une cohorte de 916 daurades. La ligne rouge représente le seuil de significativité chromosomique estimée après correction de Bonferroni et dont la valeur est $-\log_{10}(0,05 / \text{nombre de marqueur moyen par chromosome})$. La ligne noire représente le seuil de significativité génomique estimée après correction de bonferoni et dont la valeur est $-\log_{10}(0,05 / \text{nombre total de marqueurs})$.

Chez le bar, une question se pose : peut-on sélectionner des individus résistants à la fois à la nodavirose et à la vibriose ? Dans les cohortes que nous avons étudiées, il n'a pas été possible d'estimer la corrélation génétique entre les caractères. En revanche, au sein du SYSAAF, cette corrélation a pu être estimée à $-0,54$ sur d'autres jeux de données (Morvezen et al, comm. Pers.). Cela indique qu'une sélection conjointe pour la résistance à la nodavirose et à la vibriose chez le bar sera compliquée à mettre en place. L'obtention d'un bar résistant aux deux maladies se fera par l'utilisation d'outils génomiques permettant d'augmenter le progrès génétique sur les deux caractères malgré la corrélation négative ou en utilisant une sélection assistée par marqueurs pour introgresser le QTL de résistance à la nodavirose dans des lignées auparavant sélectionnées pour la résistance à la vibriose.

5.2.2.2. Quelle rentabilité pour les entreprises ?

Le passage d'une sélection massale ou généalogique sur collatéraux à une sélection génomique ne se fera pas sans se poser la question de la rentabilité de la mise en place de cette technique pour les entreprises. La question peut se poser de plusieurs façons selon le contexte de chaque entreprise : l'augmentation du coût du génotypage et/ou phénotypage sera-t-elle compensée par une augmentation du progrès génétique qui se traduira par une hausse du prix de la génétique vendue (sous forme d'œufs ou d'alevins) ou par une hausse des parts de marché lorsque la génétique est exportée ? Dans le cas d'une production intégrée, sera-t-elle compensée par une baisse de la mortalité dans les cheptels de production ?

Dans le chapitre 3, nous avons montré que, quelle que soit la densité de marqueurs utilisée, la précision obtenue par la sélection génomique est supérieure à celle obtenue par sélection sur pedigree. Dans ce cas, nous pouvons estimer la taille de population d'entraînement, c'est-à-dire le nombre d'individus à phénotyper et génotyper, pour obtenir la même précision que celle obtenue avec une sélection sur pedigree et une population d'entraînement la plus grande possible. En supposant l'utilisation de la puce 57K pour le bar ou 60K pour la daurade, une population d'entraînement de 200, 600, 500 et 350 individus est suffisante chez VNN_A, VNN_B, VIB et PAS respectivement chez qui correspond à une diminution de la population d'entraînement de 25 à 62% selon les cohortes. Cette diminution du coût du phénotypage nécessite cependant le passage d'un génotype sur une dizaine de marqueurs microsatellites ou une centaine de marqueurs SNPs pour l'assignation de parenté à un génotypage sur une puce de plusieurs dizaines de milliers de marqueurs, dont le coût est plus élevé. A nombre de candidats à la sélection constant, la baisse du nombre d'individus à génotyper et phénotyper doit au minimum compenser l'augmentation du coût du génotypage.

Une première alternative serait d'utiliser une puce plus basse densité, de l'ordre de 6K marqueurs, qui offre une précision proche de celle d'une puce 50K marqueurs mais permet une diminution du coût du génotypage. Une autre alternative serait le génotypage de certains individus, principalement les parents, sur des puces 50K marqueurs et le génotypage des candidats et des collatéraux sur des puces très basses densités, de l'ordre de 1K marqueurs ou moins, et d'ensuite imputer leurs génotypes sur la puce 50K. Cette stratégie a montré des avantages chez d'autres espèces (Tribout et al., 2013; Tsai et al., 2017; Tsairidou et al., 2020).

Une autre manière de voir ces résultats est qu'à taille de population d'entraînement constante, la sélection génomique permet d'augmenter la précision de sélection et donc, le gain de survie. Cette augmentation de la survie au sein des individus vendus par les sélectionneurs sera un atout de vente. Des auteurs ont travaillé sur la traduction d'une augmentation de survie par sélection en une valeur économique par l'utilisation du R_0 , qui représente le nombre moyen d'individus infectés par un individu infecté, c'est-à-dire le nombre moyen de cas secondaires contaminés par un individu infecté. D'un point de vue économique, la sélection pour la résistance aux maladies présente un intérêt économique lorsque R_0 est supérieur à 1, condition nécessaire pour qu'une épidémie puisse se développer (Janssen and Bijma, 2020). Dans ce cas en effet, comme un individu infecté va en contaminer plus d'un, le nombre de cas va augmenter dans la population, entraînant une perte pour le producteur. Ainsi, diminuer le R_0 permet de diminuer la consommation de produits thérapeutiques et les pertes (Janssen and Bijma, 2020). Lorsque le R_0 est inférieur à 1, la sélection pour augmenter la résistance aux maladies n'est plus économiquement viable car le nombre d'individus infectés va décroître au cours du temps. Les épidémies sont donc très rares et les conséquences sur la production sont négligeables (Janssen and Bijma, 2020). On peut observer que la valeur économique du R_0 augmente lorsque R_0 diminue puis tombe brutalement à 0 lorsqu'il est inférieur à 1. Cela indique que lorsque le R_0 est très grand, il est moins intéressant économiquement d'améliorer la résistance à la maladie que lorsqu'il est proche de 1. Janssen and Bijma (2020) déclarent que lorsque R_0 est grand, les pertes sont trop nombreuses et le coût en traitements serait trop important pour lutter contre l'épidémie, impliquant que même une baisse de R_0 n'impacterait pas le bilan économique. En revanche, au fur et à mesure que le R_0 se rapproche de 1, chaque gain génétique permet d'augmenter le taux de survie et de diminuer le coût des traitements, ce qui améliore également le bilan économique. Dans Tsairidou et al. (2019), les auteurs montrent que le progrès génétique réalisé sur le R_0 à partir de simulations est de plus en plus faible au fur et à mesure que la valeur de R_0 baisse et parviennent à obtenir un R_0 inférieur à 1 après quatre à six générations de sélection selon le R_0 de départ.

Bien que tous ces résultats dépendent du R_0 , la mesure de sa valeur n'est pas simple à obtenir. Dans les challenges infectieux qui sont réalisés en aquaculture, le phénotype qui est couramment utilisé dans les estimations de valeurs génétiques est la survie mesurée comme un caractère binaire. Ce caractère n'est qu'un indicateur de la sensibilité d'un individu à une maladie alors que l'infectiosité rentre également en compte dans le calcul du R_0 . En effet, dans Anche et al. (2014), les auteurs montrent qu'une sélection basée sur la survie à la fin d'une épidémie peut améliorer le R_0 (le faire diminuer) s'il existe une corrélation génétique entre sensibilité, capturée par la survie, et l'infectiosité (non mesuré). En revanche, ils montrent que lorsqu'il n'y a pas de corrélation génétique entre sensibilité et infectiosité, une sélection à partir de la survie à l'issue du challenge permet de diminuer la sensibilité au fur et à mesure des générations et diminuer la mortalité. De même, dans Tsairidou et al. (2019), les auteurs montrent par simulation qu'une sélection basée sur la sensibilité, donc la survie à l'issue d'un challenge, permet d'atteindre un R_0 inférieur ou égal à 1 après sept générations de sélection, en supposant qu'il n'y a pas de corrélation génétique entre sensibilité et infectiosité. Cela démontre que les résultats de Janssen and Bijma (2020) sont valables même s'il n'est pas possible d'estimer le R_0 d'une population. L'estimation du R_0 dans une population nécessite un challenge particulier, où seule une partie des individus sont infectés et, chaque jour, chaque poisson doit être identifié comme étant infecté ou non (Janssen and Bijma, 2020). La mesure du R_0 dans les populations qui ont été étudiées dans cette thèse n'a donc pas été possible à partir des données qui ont été recueillies.

5.3. Mise en perspective

5.3.1. Virose et bactériose

Dans cette thèse, nous avons étudié une maladie causée par un virus et deux causées par des bactéries. Bien que la résistance à la vibriose et à la pasteurellose ait été mesurées sur deux espèces différentes, nous avons observé des similarités entre les deux caractères. La résistance à la vibriose et la résistance à la pasteurellose ont des héritabilités plus faibles que celle de la résistance à la nodavirose (0,11 et 0,16 contre 0,23). De plus, les résistances aux bactérioses ont une architecture génétique polygénique tandis que la résistance à la nodavirose est au moins en partie oligogénique, du fait de l'existence d'un QTL à effet fort.

Cette différence entre résistance à une virose et à une bactériose n'est due qu'aux maladies que nous avons étudiées et non pas à une règle générale. En effet, dans la revue de Ødegård et al. (2011), les auteurs répertorient les héritabilités et QTLs détectés pour de nombreux couples hôte – pathogène en aquaculture. Nous retrouvons aussi bien des résistances à des maladies virales peu hérissables comme la résistance au white spot syndrome virus chez la crevette qui a une héritabilité de 0,04 que des héritabilités élevées pour des résistances à des bactérioses comme celle de la furunculose chez le saumon qui a été estimée entre 0,43 et 0,62. De plus, bien que la majorité des QTLs à effet fort soient impliqués dans des résistances contre des virus (NPI chez le saumon, SHV chez la truite), des QTLs ont pu être détectés pour des résistances à des bactérioses, comme chez la truite arc-en-ciel dans la résistance à *Flavobacterium psychrophilum* (Fraslin et al., 2018; Vallejo et al., 2014) ou chez la daurade dans la résistance à la pasteurellose (Aslam et al., 2018).

5.3.2. Injection et immersion

Le mode de contamination lors du challenge peut également expliquer une partie des différences observées entre la résistance à la nodavirose et les résistances à la vibriose et à la pasteurellose. Lors du challenge au nodavirus, les poissons ont été immergés dans une eau contenant le virus. Par contre, lors des challenges à *V. harveyi* et *Photobacterium damsela subsp. piscicida*, les poissons ont été injectés avec une solution contenant la bactérie. Le mode d'infection au challenge peut avoir un impact dans l'estimation des paramètres génétiques et dans la détection de QTL car il s'agit de deux phénotypes potentiellement différents.

En effet, l'injection de l'agent pathogène directement à l'intérieur de l'organisme permet de passer outre la peau et le mucus qui composent la première barrière de défense. Dans Fraslin et al. (2018), les auteurs ont détecté des QTLs de résistance à *Flavobacterium psychrophilum* chez la truite arc-en-ciel liés à une infection par injection et d'autres liés à une infection par immersion. Au total 15 QTLs ont été détectés, et seulement trois sont communs aux deux modes d'infection. Cela indique l'existence d'un ensemble commun de gènes impliqués dans la résistance à *Flavobacterium psychrophilum*, mais aussi l'existence de mécanismes de défenses immunitaires spécifiques aux voies d'infection du pathogène qui peuvent donc être sélectionnés.

Bien que l'immersion soit plus proche des conditions naturelles, l'injection assure l'entrée du pathogène dans l'organisme et diminue donc l'influence de l'environnement dans la résistance. Dans une situation idéale, il faudrait mener en parallèle un challenge avec une infection par balnéation et un avec une infection par injection si le pathogène utilisé permet les deux types d'inoculation. Cela permettrait dans un premier temps d'évaluer l'architecture génétique des deux caractères et d'en connaître les spécificités et les interactions. Sur le long terme, cela permettrait de sélectionner des individus résistants aux deux types d'infections et donc de multiplier les obstacles contre le pathogène pour assurer un haut potentiel de survie. En pratique, faire deux challenges à chaque génération nécessite la création de deux lots dérivés ce qui double les coûts d'élevage, de phénotypage et de génotypage dans le cadre d'une sélection génomique, ce qui n'est *a priori* pas rentable pour une entreprise. Ainsi, il serait plus avantageux de privilégier une infection par injection lorsque cela est possible puisqu'elle permet d'évaluer la dernière barrière de défense d'un individu (Fraslin et al., 2018). De plus, les résistances aux maladies mesurées avec ces deux modes d'infection sont corrélées génétiquement ce qui indique d'une sélection sur un mode améliorera le second indirectement (Fraslin et al., 2018).

5.4. Perspectives

5.4.1. Vers une augmentation du nombre de marqueurs

Dans cette thèse, nous avons étudié des jeux de données relativement similaires. Ils étaient composés d'environ 1 000 individus génotypés sur environ 45 000 marqueurs. Nous avons évoqué dans le Chapitre 2 l'intérêt d'augmenter la taille des jeux de données, aussi bien en nombre de phénotypes qu'en nombre de marqueurs. Avec le passage récent des sélectionneurs à la sélection génomique, la quantité de génotypes reliés à des phénotypes ne va cesser d'augmenter. Avec une augmentation du nombre de phénotypes, nous avons d'ores et déjà précisé la localisation et l'effet du QTL de résistance à la nodavirose que nous avons détecté initialement avec moins d'individus. Néanmoins, l'intervalle de confiance est resté trop large pour identifier un gène candidat voire la mutation causale. Une augmentation du nombre de marqueurs, par séquençage complet des individus ou de leurs parents, suivi de l'imputation des données de génotypage issues de la puce à SNPs vers la séquence complète

pourrait permettre d'atteindre le gène candidat, comme cela a été fait pour la détection de QTL impliqué dans la masculinisation spontanée chez la truite arc-en-ciel (Fraslin et al., 2020a), dans le nombre de mamelles chez le porc (van den Berg et al., 2019) et dans la production de lait chez les bovins (Daetwyler et al., 2014). Le passage à la séquence complète a en effet souvent été nécessaire pour parvenir à identifier la mutation causale à l'origine de la détection du QTL. Dans le cas du QTL de résistance à l'IPN chez le saumon, le séquençage de 22 individus portant la combinaison d'allèle QQ (double résistant) et de 23 individus portant la combinaison d'allèle qq (double sensible) a permis l'identification de la mutation causale localisée dans un exon d'un gène codant pour une cadhérine épithéliale (Moen et al., 2015). C'est également le cas pour l'identification de gènes candidats impliqués dans la masculinisation spontanée chez la truite arc-en-ciel (Fraslin et al., 2020a). Un total de 1152 individus a été génotypé sur une puce à SNPs de 57K marqueurs puis imputé sur la séquence entière grâce au séquençage complet de 50 mères et de 10 femelles apparentées. Le passage d'une puce à SNPs à la séquence complète a permis de réduire l'intervalle de confiance des QTLs détectés mais aussi d'identifier les QTLs les plus intéressants pour la sélection. Chez les bovins, le passage d'une puce de 50K marqueurs à la séquence a également permis l'identification de mutations causales. Dans Adams et al. (2016), les auteurs ont identifié la mutation causale responsable d'une baisse de la fertilité chez la race Holstein grâce au séquençage de onze individus. Avant le séquençage, les auteurs ont détecté un QTL impliqué dans la baisse de fertilité chez cette race et ont tracé ce QTL dans les bases de données disponibles. Ils ont identifié huit individus, certains porteurs du QTL et d'autres une version recombinée du QTL, qui ont permis de localiser la mutation causale dans un intervalle de 3.4Mb. Le séquençage de trois individus supplémentaires, un taureau porteur du QTL ainsi que deux de ses fils, un porteur et un non porteur, a permis l'identification de la mutation causale responsable de la baisse de fertilité chez la race Holstein. Dans Michot et al. (2017), le séquençage de 54 taureaux Montbéliards dont 14 avaient été identifiés comme porteurs du QTL ainsi que l'ensemble de la base de données du 1 000 bull genome project a permis de mettre en évidence l'impact d'une mutation sur la mortalité embryonnaire. Même si le passage à la séquence ne permet pas d'atteindre la mutation causale, elle permet de mieux identifier le nombre et la position des QTLs. En effet, un QTL identifié avec une puce 50K marqueurs peut être composé en réalité de plusieurs QTLs proches, qui ne sont révélés que lorsque les données de séquençage sont utilisées (Sanchez et al., 2017).

Dans le cas de la résistance à la nodaviriose du bar, une approche par séquençage de certains parents et de l'imputation de descendants sur la séquence complète pourrait être mise en œuvre afin de préciser la position du QTL que nous avons détecté sur le LG12, voire d'en identifier la mutation causale. Si tel est le cas, la mutation causale pourrait être identifiée chez les candidats à la sélection grâce à un ou des marqueurs SNPs et une sélection assistée par marqueurs pourrait être mise en place pour améliorer la résistance à la nodaviriose. Une deuxième possibilité consisterait en l'utilisation de la technologie CRISPR/Cas9 (Doudna and Charpentier, 2014). Elle permettrait dans un premier temps une validation fonctionnelle de l'implication du gène de résistance identifié. Pour cela, une mutation ciblée par CRISPR/Cas9 ou un knockout du gène validerait l'implication du gène identifié dans la résistance à la nodaviriose. Cela a été fait chez le saumon, où le knockout d'un gène impliqué dans la survie des cellules germinales a permis de montrer que la différenciation sexuelle était indépendante de la présence de cellules germinales (Wargelius et al., 2016). La mutation ciblée d'un seul SNP a été réalisée chez le rat pour modifier l'allèle responsable de l'albinisme (Yoshimi et al., 2014). Ensuite, à une plus grande échelle, l'augmentation de la fréquence de l'allèle de résistance pourrait se faire par mutation ciblée plutôt que par sélection assistée par marqueurs, ce qui aurait l'avantage majeur d'accélérer cette augmentation de fréquence (Gratacap et al., 2019).

Dans une perspective d'application à l'échelle du SYSAAF, l'augmentation du nombre de phénotypes peut passer par la mutualisation des phénotypes issus de challenge. Actuellement, l'estimation des

valeurs génétiques pour choisir les candidats à la sélection se fait à l'échelle d'une cohorte. Nous avons déjà évoqué la possibilité d'utiliser des cohortes précédentes dans l'estimation des valeurs génétiques. Ici, nous proposons de mettre en commun les résultats de challenges de différentes cohortes provenant de plusieurs sélectionneurs et d'utiliser cette variabilité pour estimer les valeurs génétiques. Il faut toutefois noter qu'il est nécessaire que l'apparement entre les cohortes soit suffisamment élevé pour permettre un gain de précision (Habier et al., 2010). Chez les bovins, l'utilisation d'une population de référence composée de différentes races permet d'augmenter la précision de sélection lorsque la race des individus à prédire n'est pas représentée dans la population de référence. Lorsque la race des individus à prédire est représentée dans la population de référence, l'utilisation de plusieurs races permet d'atteindre une précision équivalente à celle obtenue en race pure (Pryce et al., 2011). Ainsi, l'utilisation de différentes cohortes, même si elles sont bien moins différentes entre elles que les races bovines, ne devrait pas voir d'effets délétères sur la précision de sélection. En pratique, il serait plus judicieux d'augmenter la taille population d'entraînement par l'ajout d'individus issus de cohortes apparentées ou de générations successives. Ces propositions n'ont cependant pas été démontrées dans notre cas et nécessitent plus de recherches pour établir une recommandation.

L'augmentation du nombre de marqueurs peut également augmenter la précision en sélection génomique. Nous avons montré dans le Chapitre 3 que la précision augmentait avec le nombre de marqueurs. En revanche, nous n'avons pas pu estimer un nombre de marqueurs pour lequel la précision est maximale. D'après les résultats obtenus chez d'autres espèces, l'utilisation des données de séquençage en sélection génomique n'augmente que légèrement la précision, avec un coût du génotypage nettement plus élevé. Ainsi, une densité d'environ 50K marqueurs semble être l'optimum actuel chez les bovins (voir la revue de Misztal et al. (2020)). Une approche mixte, mêlant le séquençage ou le génotypage sur une puce haute densité d'une partie des individus et le génotypage du reste des individus sur une puce basse densité (moins de 1 000 marqueurs), permettrait de combiner l'avantage d'un grand nombre de marqueurs dans l'estimation de la précision et une baisse des coûts de génotypage. Cette stratégie a été évoquée dans le chapitre 3, en se basant sur les études de Tsai et al. (2017) et Tsairidou et al. (2020) et démontre une baisse de 62% du coût du génotypage avec le génotypage des parents sur une puce de 5K marqueurs et candidats et de leurs apparentés sur 200 marqueurs suivi d'une imputation sur la puce 5K marqueurs selon les auteurs de ces publications.

Une augmentation du nombre de marqueurs peut aussi passer par la prise en compte d'autres types de marqueurs comme les marqueurs épigénétiques. Les plus utilisés sont ceux liés à la méthylation de la séquence d'ADN et à la modification chimique des histones responsables du degré de compaction de la molécule d'ADN. La quantification des micro-ARN peut également être un marqueur de la régulation de certains gènes (Chuang and Jones, 2007). L'utilisation de l'épigénétique en sélection n'en est qu'à ses débuts. La principale difficulté réside en la découverte d'une association stable entre un ou des marqueurs épigénétiques et le phénotype d'intérêt, en particulier au cours des générations. Les mécanismes liés à la transmission de l'épigénétique restent encore à préciser (Danchin et al., 2011). En revanche, des nombreuses découvertes de l'impact de l'épigénétique sur des performances d'espèces d'élevages ont été rapportées et ouvrent des perspectives d'utilisation de tels marqueurs en sélection (voir la revue de Ibeagha-Awemu and Zhao, 2015). Pour utiliser des marqueurs épigénétiques en sélection, il faut qu'ils soient transmissibles à la descendance et qu'ils permettent de prédire le même phénotype à chaque génération. Très peu d'études démontrent la transmission de marqueurs épigénétiques chez les espèces d'élevage. Une étude, réalisée chez le porc, montre l'impact d'une alimentation enrichie en micronutriments favorisant la méthylation de l'ADN sur le métabolisme du gras (Braunschweig et al., 2012). A la première génération (F0) trois mâles ont été nourris avec une alimentation contrôlée tandis que trois autres ont été nourris avec une alimentation enrichie en micronutriments favorisant la méthylation de l'ADN. Ces mâles ont été croisés avec des femelles,

nourries avec l'aliment contrôle, pour produire une génération F1. A cette génération, tous les individus ont été nourris avec une alimentation contrôle, puis, cinq mâles, issus du croisement contrôle et trois mâles issus du croisement alimentation enrichie ont été croisés avec des femelles nourries avec l'aliment contrôle pour produire une génération F2. Cette génération a été nourrie avec une alimentation contrôle et a été évalué sur des caractères liés à la qualité de la carcasse. Les auteurs ont montré que la génération F2 issus du croisement de mâles nourris avec une alimentation enrichie présente moins de gras que celle issue du croisement de mâles élevés avec l'aliment contrôle. Ils montrent également une différence de méthylation entre les deux lots au niveau de gènes exprimés dans le foie et les reins. Ils concluent donc à un effet de l'alimentation enrichie en micronutriments favorisant la méthylation sur le métabolisme du gras et à la transmission des marqueurs épigénétiques, établi en F0, jusqu'en F2. Bien que la transmission d'une génération à l'autre soit établie, les auteurs n'ont pas montré s'ils ont observé une différence sur le métabolisme du gras à partir de la génération F1. Du fait du faible nombre d'individus et de l'utilisation d'animaux ayant un fond génétique commun, ils se sont assurés que la transmission d'une part génétique additive n'avait pas d'impact sur les différences de phénotypes observées. De même, ces phénotypes ont été induits par une alimentation particulière, qu'il convient de maintenir ou non selon les objectifs de sélection. Or, il a été montré que des marqueurs épigénétiques, et donc leurs effets sur des caractères, peuvent être supprimés par l'effet de l'environnement (Jammes et al., 2011). Bien que l'épigénétique propose une autre manière d'observer les variations du génome, c'est un champ d'étude récent qui est encore éloigné d'une application dans des programmes d'amélioration.

5.4.2. Vers une sélection assistée par marqueurs ?

Avec la découverte d'un QTL à effet fort impliqué dans la résistance à la nodaviriose chez le bar dans le Chapitre 2, on peut légitimement se poser la question du passage de la sélection génomique à la sélection assistée par marqueurs pour ce caractère. En revanche, il y aurait un risque à la mettre en place à partir des résultats que nous avons obtenus. Premièrement, le QTL explique moins de 10% de la variance génétique totale dans les populations que nous avons étudiées. Ainsi, sélectionner uniquement sur ce QTL ignorerait 90% de la variance génétique à l'origine de la résistance à la nodaviriose. Dans ce cas, la démarche que nous avons suivi dans le Chapitre 3, consistant en un ajout de l'effet du QTL dans un modèle GBLUP permet de combiner une approche polygénique par l'estimation des valeurs génétiques à partir des marqueurs et une approche « QTL » en ajoutant l'information du QTL au modèle. Comme la part de variance génétique expliquée par un QTL dépend de sa fréquence dans la population étudiée, elle pourrait varier aussi bien d'une population à l'autre qu'au cours des générations de sélection. Cela pourrait modifier l'importance de sa prise en compte dans un modèle GBLUP. Si sa fréquence augmente, par exemple par sélection, il pourrait expliquer une part de variance génétique plus importante.

L'utilisation d'une puce à SNPs, permettant le génotypage sur plusieurs milliers de marqueurs, sera bien plus polyvalente qu'une sélection assistée par marqueurs car utilisable pour estimer les valeurs génétiques pour des caractères polygéniques. Dans cette thèse, nous avons déduit que les résistances à la vibriose et à la pasteurellose font partie de cette catégorie. Une architecture polygénique est plus souvent la règle que l'exception. La majorité des caractères de production (croissance, rendement filet, efficacité alimentaire) ou de qualité (teneur en gras, morphologie) sélectionnés aussi bien chez des espèces d'aquaculture ou d'élevage sont polygéniques. Bien que des QTLs aient été détectés pour ces caractères, ils sont souvent nombreux et expliquent une part de variance génétique ou phénotypique très faible, en général de l'ordre de 1% ou 2%. Pour la croissance, aucun QTL n'a été détecté chez la carpe ou chez le saumon (Palaiokostas et al., 2018c; Tsai et al., 2015), et au contraire de nombreux

QTLs à faible effet ont été détectés chez le poulet de chair (Nones et al., 2006). Dans les deux cas, ceci peut être interprété comme une conséquence d'une architecture génétique de type polygénique. Pour le rendement filet chez la truite arc-en-ciel ou chez le poulet de chair, un et deux QTLs ont été identifiés respectivement, expliquant une part de variance faible (Gonzalez-Pena et al., 2016; Pampouille et al., 2018). Seule l'efficacité alimentaire chez le poulet de chair semble impliquer un QTL ayant un effet modéré à fort, expliquant 12% de la variance phénotypique (Yuan et al., 2015). Enfin, la teneur en gras chez le saumon atlantique comme chez le porc est un caractère polygénique (Hérault et al., 2018; Sodeland et al., 2013). L'ensemble de ces caractères a une héritabilité modérée à forte, allant de 0,2 à 0,5. Ces caractères, en particulier la croissance, le rendement filet et la teneur en gras sont sélectionnés dans la majorité des programmes de sélection aquacole. Du fait de leur architecture génétique polygénique, la sélection génomique permettrait d'augmenter la précision de sélection et donc, le gain génétique. C'est en particulier le cas pour le rendement de filet, qui est mesuré sur des apparentés aux candidats à la sélection. Ainsi, l'utilisation d'une puce à SNPs, d'une densité allant de 6K à 57K marqueurs, permettrait une augmentation de la précision sur l'ensemble des caractères sélectionnés et donc une amélioration du progrès génétique.

Il convient néanmoins de confirmer ce résultat avec de nouveaux challenges et cohortes. Pour la résistance à la vibriose chez le bar, aucun autre article n'est paru à ce jour et, pour la résistance à la pasteurellose, un article contredit nos résultats (Aslam et al., 2018). Dans une cohorte de 1 187 daurades provenant de l'écloserie FMD, comme la cohorte de daurade étudiée dans cette thèse, et génotypée sur 28 330 marqueurs SNPs, un QTL impliqué dans la résistance à la pasteurellose a en effet été détecté par ces auteurs sur le chromosome 17 et explique entre 13% et 16% de la variance génétique totale. Cette différence entre ces résultats et ceux obtenus au cours de cette thèse peut être due à des différences de protocoles, car les deux challenges n'ont pas été réalisés au même endroit, avec qui plus est l'utilisation de différentes souches bactériennes (souche 249/ITT/99 chez Aslam et al. (2018), souche PP11787 6/94 dans notre étude). Cela a été démontré chez différentes espèces végétales, comme dans le cas de la pyculariose du riz où des QTLs différents ont été détectés selon la souche du pathogène utilisée (Talukder et al., 2004) ; ou chez l'aubergine où un QTL spécifique à une souche de *Ralstonia solanacearum* a été détecté en plus de deux autres QTLs ayant un effet sur plusieurs souches (Salgon et al., 2017). En revanche, aucune étude de ce type n'a pu être trouvée chez des espèces animales.

5.4.3. Vers un raffinement du phénotype

La qualité du phénotype est un élément essentiel en sélection. Dans cette thèse, le phénotype mesuré était la survie de poisson, infecté par une souche spécifique d'un pathogène donné, dans une condition expérimentale. On peut donc se demander si les résultats que nous avons obtenus sont valables pour d'autres souches du pathogène et dans d'autres conditions d'élevage.

Dans le cas de la vibriose du bar, nous avons uniquement étudié la résistance à une souche de *Vibrio harveyi*. Or, d'autres pathogènes, comme *Vibrio anguillarum*, causent également la vibriose. L'héritabilité de la résistance à *Vibrio anguillarum* a été estimée à 0,32 (Chatziplis et al., 2020). Il est fortement probable que la résistance à *Vibrio harveyi* et à *Vibrio anguillarum* soient deux caractères différents, bien que potentiellement corrélés. Chez l'agneau, il a été montré que les résistances à différents nématodes sont fortement corrélés, allant de 0,5 (Bishop et al., 2004) à des valeurs proches de 1 pour certains couples (Gruner et al., 2004). De même chez le mouton, une corrélation génétique de 0,31 a été trouvée pour la résistance à deux mycotoxines différentes responsable du tournis du ray-grass et de l'eczéma facial (Morris et al., 1995).

Quel que soit l'agent pathogène étudié, une seule souche a été employée pour les challenges. Or, la résistance d'un individu à un pathogène peut être différente selon la souche de pathogène. Pour la résistance à la nodaviriose de bar, Vendramin et al. (2014) ont montré un taux de mortalité allant de 7,5% à 36,25% dans une même cohorte selon la souche de nodavirus employée. Des résultats équivalents ont été montrés chez l'huitre creuse, avec les taux de survie allant d'environ 20% à plus de 50% selon la souche de *V. aestuarianus* utilisée lors du challenge infectieux (Travers et al., 2017). Cette résistance différente selon la souche virale employée montre la nécessité d'adapter la souche utilisée lors du challenge pour quelle corresponde à celle communément trouvée dans le milieu naturel voire d'en utiliser plusieurs afin d'estimer une résistance globale au pathogène et non pas une résistance à une souche spécifique.

Dans le cas de la nodaviriose, les challenges ont été réalisés à une température de 27°C qui est favorable au développement du VNN. Par contre, la température optimale de croissance du bar a été estimée à 26°C (Person-Le Ruyet et al., 2004). Ainsi, les poissons élevés à une température élevée par rapport à l'optimal peuvent subir un stress qui pourrait favoriser le développement de la nodaviriose et donc, augmenter la mortalité. Ainsi, la survie mesurée au cours de challenge pourrait être la résultante à la fois de la sensibilité d'origine génétique des individus à la maladie, mais aussi leur capacité à se développer dans des conditions difficiles. Cette possible interaction entre résistance à la nodaviriose et capacité à se développer dans des conditions stressantes chez le bar a déjà été relevée par Doan et al. (2017). Dans cet article, les auteurs ont montré une résistance différentielle du bar en fonction de son origine géographique. Ils ont montré qu'il existe un gradient de résistance, allant de l'océan Atlantique au sud-est de la Méditerranée. Or, ce gradient géographique se confond avec le gradient de température de l'eau et indique que les individus les plus résistants, originaire de l'est de la Méditerranée, sont plus adaptés à une température de 25°C, température à laquelle tous les poissons ont été challengés, que ceux de l'Atlantique ou de l'ouest de la Méditerranée. Il serait donc recommandé de tester différentes températures de challenge dans une population originaire de la Méditerranée et qui supporte une gamme de température permettant un développement du nodavirus et de mesurer la survie afin de séparer un effet de la température et un effet génétique sur la survie observée.

La mesure de la mortalité lors d'un challenge implique la création d'une cohorte apparentée aux candidats qui servira à estimer les valeurs génétiques des candidats et qui sera sacrifiée à l'issue du challenge. Pour des raisons financières et éthiques, l'utilisation d'un autre phénotype, reflétant la capacité d'un individu à résister à une maladie mais sans l'euthanasier à la fin du challenge est une piste qui peut être envisagée. Pour cela, des challenges infectieux *in vitro* ont été développés, tout d'abord chez les bovins (Price et al., 1990), mais aussi chez le porc (Vincent et al., 2006) et chez différentes espèces d'aquaculture (Quillet et al., 2001; Robinson and Hayes, 2008). Il consiste en un prélèvement de tissus, soit des nageoires soit des organes ciblés par le pathogène, qui seront cultivées sur milieu de culture avant d'être inoculés par l'agent infectieux. Le caractère mesuré peut être la quantité de cellules mortes ou la quantité de pathogène mesurée par qPCR. Cette technique a l'avantage d'être prédictive de la survie de l'individu tout en évitant l'euthanasie de la cohorte challengée. En plus de ses avantages éthiques, elle permet de mesurer la résistance au pathogène directement sur les candidats (Quillet et al., 2001; Verrier et al., 2012). En revanche, la résistance mesurée à partir de challenge *in vitro* ne prédit pas toujours parfaitement ($r^2 = 0,52$ dans Quillet et al. (2001) et $r^2 = 0,99$ dans Verrier et al. (2012)) la survie d'un individu et peut donc créer un biais dans la sélection. Par ailleurs, alors qu'un challenge sur un individu va mettre en évidence la résistance de cet individu au pathogène, le challenge *in vitro* va uniquement montrer la résistance d'un tissu ou d'un type cellulaire au même pathogène. Or, la résistance à un pathogène est un caractère intégratif de multiples mécanismes de défenses d'un individu à un pathogène. Chez les poissons, le système

immunitaire peut être divisé entre trois compartiments différents : les barrières physiques avec le milieu extérieur (mucus, écailles, peau, paroi buccale, paroi du tube digestif), le système humoral qui est constitué de molécules circulantes et le système cellulaire qui est composé des cellules du système immunitaire (phagocytes, leucocytes) (Magnadottir, 2010). Pour illustrer la diversité des mécanismes de défenses immunitaires, prenons l'exemple de la vibriose. Pour commencer, la bactérie doit pénétrer dans l'organisme. S'il n'y a pas de lésions extérieures, la bactérie doit se frayer un passage au travers du mucus et de la peau qui constituent le premier mécanisme de défense (Ina-Salwany et al., 2019). Ensuite, la bactérie va se développer et sécréter des toxines dans l'organisme, capables d'affecter la rate, les reins, le foie et les intestins (Hendrikson and Zenoble, 1983). A ce stade, ce sont les mécanismes internes qui vont lutter contre la bactérie, que se soient les cellules du système immunitaire ou les anticorps (Hendrikson and Zenoble, 1983). Ainsi, la résistance à la vibriose est un phénotype intégrant les défenses présentes à la fois dans le mucus, la peau, les phagocytes et leucocytes mais aussi la synthèse et sécrétion d'anticorps voire d'autres mécanismes encore inconnus. Pour finir, la mise en place d'un challenge sur tissus ou cellules nécessite de maîtriser à la fois la culture des cellules qui seront infectés mais aussi celle du pathogène et requiert plus de temps qu'un challenge sur des individus. Il ne s'agit donc pas d'une voie simple, même si elle est prometteuse et éthique.

5.5. Conclusion

Nous avons montré que l'utilisation de marqueurs SNPs permet non seulement d'obtenir des pedigrees plus fiables mais aussi, par l'utilisation d'une puce moyenne densité, d'étudier l'architecture génétique de la résistance aux maladies chez le bar et la daurade, et de sélectionner plus précisément les candidats à la sélection. Le développement des puces Thermofisher Axiom™ DlabCHIP 57K et SaurCHIP 60K est un atout majeur pour la sélection du bar et de la daurade du fait de leur facilité d'utilisation, de leur polyvalence et de leur efficacité. Nous avons montré que les résistances aux maladies étudiées sont en grande partie des caractères polygéniques, même si un QTL à effet fort a été détecté pour la résistance à la nodavirose. Nous avons également montré que la sélection génomique offre un gain de précision par rapport à une sélection sur pedigree, même lorsque seulement 6 000 marqueurs sont présents sur la puce. Nos résultats appuient l'intérêt d'un passage à la sélection génomique dans les programmes de sélection du bar et de la daurade et proposent des pistes d'amélioration permettant son optimisation et sa rentabilité.

Tout comme cela a été montré chez d'autres espèces et d'autres caractères, la sélection génomique offre un avantage par rapport à une sélection sur pedigree en augmentant la précision de sélection (Hayes et al., 2009; Palaiokostas et al., 2016; Tsai et al., 2015). Nos résultats s'inscrivent dans cette tendance, en présentant également l'impact du nombre d'individus dans la population d'entraînement et du nombre de marqueurs sur la précision qui sont deux aspects importants du coût d'un passage à la sélection génomique. En revanche, l'architecture génétique des caractères étudiés est différente de celle qui avaient déjà été publiées. Alors qu'une publication précédente présentait la résistance à la nodavirose du bar comme un caractère polygénique, nous avons montré qu'un QTL à effet fort était impliqué. A l'inverse, un QTL à effet fort avait été publié pour la résistance à la pasteurellose chez la daurade alors que nous n'avons pu détecter aucun QTL. L'origine de ces différences a été discutée, et d'autres détectations de QTL seront nécessaires pour permettre de valider définitivement les résultats obtenus.

Bien que nous ayons présenté l'avantage de la sélection génomique pour améliorer les résistances aux maladies chez le bar et la daurade, nos résultats doivent être replacés dans un contexte d'amélioration de lignées sur plusieurs caractères, qui comprennent des caractères de croissance et de morphologie

principalement, mais aussi de rendement de découpe et de qualité de la chair. L'estimation des corrélations génétiques entre ces caractères et la résistance aux maladies sera une première étape pour évaluer l'efficacité d'une sélection multi-caractère mais de plus amples travaux doivent être menés pour évaluer l'impact de la sélection génomique dans les schémas de sélection chez le bar et la daurade. Ces futurs travaux peuvent également être menés en parallèle d'une étude plus fine sur l'optimisation de la sélection génomique, en particulier en ce qui concerne l'utilisation d'une puce basse densité suivie d'une imputation sur un panel de marqueurs plus dense, voire d'une modification des plans de croisement pour optimiser la précision de sélection tout en limitant les coûts.

La sélection génomique a été un atout majeur pour les espèces chez lesquelles elle a été mise en place. Le SYSAAF et les sélectionneurs ont d'ores et déjà initié la transition vers la génomique et ces travaux devraient permettre d'en optimiser l'utilisation et consolider l'avantage compétitif des sélectionneurs de bar et daurade français, qui sont de fait les premiers à pouvoir mettre sur le marché des alevins sélectionnés issus d'une sélection génomique.

BIBLIOGRAPHIE

- Årde, J., Moen, T., Santi, N., Korsvoll, S.A., Kjøglum, S., Meuwissen, T.H.E., 2014. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5. <https://doi.org/10.3389/fgene.2014.00402>
- Adams, H.A., Sonstegard, T.S., VanRaden, P.M., Null, D.J., Van Tassell, C.P., Larkin, D.M., Lewin, H.A., 2016. Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science* 99, 6693–6701. <https://doi.org/10.3168/jds.2015-10517>
- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., Misztal, I., 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51, 28. <https://doi.org/10.1186/s12711-019-0469-3>
- Ahmed, N., Thompson, S., Glaser, M., 2019. Global Aquaculture Productivity, Environmental Sustainability, and Climate Change Adaptability. *Environmental Management* 63, 159–172. <https://doi.org/10.1007/s00267-018-1117-3>
- Akdemir, D., Sanchez, J.I., Jannink, J.-L., 2015. Optimization of genomic selection training populations with a genetic algorithm. *Genet Sel Evol* 47, 38. <https://doi.org/10.1186/s12711-015-0116-6>
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anche, M.T., de Jong, M.C.M., Bijma, P., 2014. On the definition and utilization of heritable variation among hosts in reproduction ratio R0 for infectious diseases. *Heredity* 113, 364–374. <https://doi.org/10.1038/hdy.2014.38>
- Anderson, E.C., Garza, J.C., 2005. The Power of Single-Nucleotide Polymorphisms for Large-Scale Parentage Inference. *Genetics* 172, 2567–2582. <https://doi.org/10.1534/genetics.105.048074>
- Andreoni, F., Magnani, M., 2014. Photobacteriosis: Prevention and Diagnosis. *Journal of Immunology Research* 2014, 1–7. <https://doi.org/10.1155/2014/793817>
- Antonello, J., Massault, C., Franch, R., Haley, C., Pellizzari, C., Bovo, G., Patarnello, T., de Koning, D.-J., Bargelloni, L., 2009. Estimates of heritability and genetic correlation for body length and resistance to fish pasteurellosis in the gilthead sea bream (*Sparus aurata* L.). *Aquaculture* 298, 29–35. <https://doi.org/10.1016/j.aquaculture.2009.10.022>
- Arthur, P.F., Herd, R.M., 2005. Efficiency of feed utilisation by livestock — Implications and benefits of genetic improvement. *Can. J. Anim. Sci.* 85, 281–290. <https://doi.org/10.4141/A04-062>
- Aslam, M.L., Carraro, R., Bestin, A., Cariou, S., Sonesson, A.K., Bruant, J.-S., Haffray, P., Bargelloni, L., Meuwissen, T.H.E., 2018. Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *BMC Genet* 19, 43. <https://doi.org/10.1186/s12863-018-0631-x>
- Azémar, P., Lamy, J.-B., Boudry, P., Renault, T., Travers, M.-A., Dégremont, L., 2017. Genetic parameters of resistance to *Vibrio aestuarianus*, and OshV-1 infections in the Pacific oyster, *Crassostrea gigas*, at three different life stages. *Genet Sel Evol* 49, 23. <https://doi.org/10.1186/s12711-017-0297-2>
- Bagni, M., 2005. Cultured Aquatic Species Information Programme. *Dicentrarchus labrax*. FAO Fisheries and Aquaculture Department.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3, e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Bangera, R., Ødegård, J., Præbel, A.K., Mortensen, A., Nielsen, H.M., 2011. Genetic correlations between growth rate and resistance to vibriosis and viral nervous necrosis in Atlantic cod

- (*Gadus morhua* L.). *Aquaculture* 317, 67–73.
<https://doi.org/10.1016/j.aquaculture.2011.04.018>
- Banos, G., Wiggans, G.R., Powell, R.L., 2001. Impact of Paternity Errors in Cow Identification on Genetic Evaluations and International Comparisons. *Journal of Dairy Science* 84, 2523–2529.
[https://doi.org/10.3168/jds.S0022-0302\(01\)74703-0](https://doi.org/10.3168/jds.S0022-0302(01)74703-0)
- Baranski, M., Kettunen Praebel, A., Sommer, A., Haanes, K., Wesmajervi, M., 2010. Major quantitative trait loci for viral nervous necrosis resistance in atlantic cod, in: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production: August 1-6, 2010 Leipzig, Germany*. de.
- Bell, J.D., Watson, R.A., Ye, Y., 2017. Global fishing capacity and fishing effort from 1950 to 2012. *Fish Fish* 18, 489–505. <https://doi.org/10.1111/faf.12187>
- Besson, M., Aubin, J., Komen, H., Poelman, M., Quillet, E., Vandeputte, M., van Arendonk, J.A.M., de Boer, I.J.M., 2016. Environmental impacts of genetic improvement of growth rate and feed conversion ratio in fish farming under rearing density and nitrogen output limitations. *Journal of Cleaner Production* 116, 100–109. <https://doi.org/10.1016/j.jclepro.2015.12.084>
- Bishop, S.C., Jackson, F., Coop, R.L., Stear, M.J., 2004. Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. *Anim. Sci.* 78, 185–194.
<https://doi.org/10.1017/S1357729800053972>
- Boerner, V., Tier, B., 2016. BESSiE: a software for linear model BLUP and Bayesian MCMC analysis of large-scale genomic data. *Genet. Sel. Evol.* 48, 63. <https://doi.org/10.1186/s12711-016-0241-x>
- Boichard, D., Barbotte, L., Genestout, L., 2014. AccurAssign, software for accurate maximum-likelihood parentage assignment. *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*.
- Bonnemaijer, P.W.M., Leeuwen, E.M. van, Iglesias, A.I., Gharahkhani, P., Vitart, V., Khawaja, A.P., Simcoe, M., Höhn, R., Cree, A.J., Igo, R.P., Burdon, K.P., Craig, J.E., Hewitt, A.W., Jonas, J., Khor, C.-C., Pasutto, F., Mackey, D.A., Mitchell, P., Mishra, A., Pang, C., Pasquale, L.R., Springelkamp, H., Thorleifsson, G., Thorsteinsdottir, U., Viswanathan, A.C., Wojciechowski, R., Wong, T., Young, T.L., Zeller, T., Allingham, R., Budenz, D., Bailey, J.C., Fingert, J., Gaasterland, D., Gaasterland, T., Haines, J.L., Hark, L., Hauser, M., Kang, J.H., Kraft, P., Lee, R., Lichter, P., Liu, Y., Moroi, S., Pasquale, L.R., Pericak, M., Realini, A., Rhee, D., Richards, J.R., Ritch, R., Scott, W.K., Singh, K., Sit, A., Vollrath, D., Weinreb, R., Wollstein, G., Wilmer, D.Z., Atan, D., Aslam, T., Barman, S.A., Barrett, J.H., Bishop, P., Blows, P., Bunce, C., Carare, R.O., Chakravarthy, U., Chan, M., Chua, S.Y.L., Crabb, D.P., Cumberland, P.M., Day, A., Desai, P., Dhillon, B., Dick, A.D., Egan, C., Ennis, S., Foster, P., Fruttiger, M., Gallacher, J.E.J., Garway, D.F., Gibson, J., Dan Gore, Guggenheim, J.A., Hardcastle, A., Harding, S.P., Hogg, R.E., Keane, P.A., Khaw, P.T., Lascaratos, G., Macgillivray, T., Mackie, S., Martin, K., McGaughey, M., McGuinness, B., McKay, G.J., McKibbin, M., Mitry, D., Moore, T., Morgan, J.E., Muthy, Z.A., O’Sullivan, E., Owen, C.G., Patel, P., Paterson, E., Peto, T., Petzold, A., Rahi, J.S., Rudnikca, A.R., Self, J., Sivaprasad, S., Steel, D., Stratton, I., Strouthidis, N., Sudlow, C., Thomas, D., Trucco, E., Tufail, A., Vernon, S.A., Viswanathan, A.C., Williams, C., Williams, K., Woodside, J.V., Yates, M.M., Yip, J., Zheng, Y., Gerhold-Ay, A., Nickels, S., Wilson, J.F., Hayward, C., Boutin, T.S., Polašek, O., Aung, T., Khor, C.C., Amin, N., Lotery, A.J., Wiggs, J.L., Cheng, C.-Y., Hysi, P.G., Hammond, C.J., Thiadens, A.A.H.J., MacGregor, S., Klaver, C.C.W., Duijn, C.M. van, International Glaucoma Genetics Consortium, NEIGHBORHOOD consortium, UK Biobank Eye and Vision Consortium, 2019. Multi-trait genome-wide association study identifies new loci associated with optic disc parameters. *Communications Biology* 2, 435.
<https://doi.org/10.1038/s42003-019-0634-9>
- Bradford, H., Pocrnic, I., Fragomeni, B., Lourenco, D., Misztal, I., 2017. Optimum selection of core animals in the efficient inversion of the genomic relationship matrix. *Journal of Animal Science* 95, 90. <https://doi.org/10.2527/asasann.2017.183>

- Brard, S., Ricard, A., 2015. Is the use of formulae a reliable way to predict the accuracy of genomic selection? *J. Anim. Breed. Genet.* 132, 207–217. <https://doi.org/10.1111/jbg.12123>
- Braunschweig, M., Jagannathan, V., Gutzwiller, A., Bee, G., 2012. Investigations on Transgenerational Epigenetic Response Down the Male Line in F2 Pigs. *PLoS ONE* 7, e30583. <https://doi.org/10.1371/journal.pone.0030583>
- Breuil, G., Bonami, J.R., Pepin, J.F., Pichot, Y., 1991. Viral infection (picorna-like virus) associated with mass mortalities in hatchery-reared sea-bass (*Dicentrarchus labrax*) larvae and juveniles. *Aquaculture* 97, 109–116. [https://doi.org/10.1016/0044-8486\(91\)90258-9](https://doi.org/10.1016/0044-8486(91)90258-9)
- Brody, T., Wohlfarth, G., Hulata, G., Moav, R., 1981. Application of electrophoretic genetic markers to fish breeding. *Aquaculture* 24, 175–186. [https://doi.org/10.1016/0044-8486\(81\)90054-5](https://doi.org/10.1016/0044-8486(81)90054-5)
- Broman, K.W., 2001. Review of statistical methods for QTL mapping in experimental crosses. *Lab animal* 30, 44–52.
- Bulmer, M., 1971. The effect of selection on genetic variability. *The American Naturalist* 105, 201–211.
- Calus, M.P.L., Veerkamp, R.F., 2010. Increase in accuracy using multi-trait genomic breeding value estimation. *Adv. anim. biosci.* 1, 60–60. <https://doi.org/10.1017/S2040470010002037>
- Chakraborty, R., Meagher, T.R., Smouse, P.E., 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* 118, 527.
- Chatziplis, D., Oikonomou, S., Loukovitis, D., Tsiokos, D., Samaras, A., Dimitroglou, A., Kottaras, L., Papanna, K., Papaharisis, L., Tsigenopoulos, C., Pavlidis, M., 2020. QTL for Stress and Disease Resistance in European Sea Bass, *Dicentrarchus labrax* L. *Animals* 10, 1668. <https://doi.org/10.3390/ani10091668>
- Chuang, J.C., Jones, P.A., 2007. Epigenetics and MicroRNAs. *Pediatr Res* 61, 24R-29R. <https://doi.org/10.1203/pdr.0b013e3180457684>
- Churchill, G.A., Doerge, R.W., 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963.
- Cleveland, M.A., Hickey, J.M., 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. *Journal of Animal Science* 91, 3583–3592. <https://doi.org/10.2527/jas.2013-6270>
- Colloca, F., Cerasi, S., 2005. Cultured Aquatic Species Information Programme. *Sparus aurata*. FAO Fisheries and Aquaculture Department.
- Costa, J.Z., Thompson, K.D., 2016. Understanding the interaction between Betanodavirus and its host for the development of prophylactic measures for viral encephalopathy and retinopathy. *Fish & Shellfish Immunology* 53, 35–49. <https://doi.org/10.1016/j.fsi.2016.03.033>
- Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.J., Anderson, C., VanTassell, C.P., Hulsege, I., Goddard, M.E., Guldbbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R., Hayes, B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet* 46, 858–865. <https://doi.org/10.1038/ng.3034>
- Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185, 1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- Damir, K., Irena, V.S., Damir, V., Emin, T., 2013. Occurrence, characterization and antimicrobial susceptibility of *Vibrio alginolyticus* in the Eastern Adriatic Sea. *Marine Pollution Bulletin* 75, 46–52. <https://doi.org/10.1016/j.marpolbul.2013.08.008>
- Danchin, É., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B., Blanchet, S., 2011. Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet* 12, 475–486. <https://doi.org/10.1038/nrg3028>

- De Woody, J.A., 2005. Molecular approaches to the study of parentage, relatedness and fitness: practical application for wild animals. *Journal of Wildlife Management* 69, 1400–1418. [https://doi.org/10.2193/0022-541X\(2005\)69\[1400:MATTSO\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)69[1400:MATTSO]2.0.CO;2)
- Dempster, E.R., Lerner, I.M., 1950. Heritability of Threshold Characters. *Genetics* 35, 212–236.
- DePoy, E., Gitlin, L.N., 2016. Statistical Analysis for Experimental-Type Designs, in: *Introduction to Research*. Elsevier, pp. 282–310. <https://doi.org/10.1016/B978-0-323-26171-5.00020-3>
- Doan, Vandeputte, M., Chatain, B., Haffray, P., Vergnet, A., Breuil, G., Allal, F., 2017a. Genetic variation of resistance to Viral Nervous Necrosis and genetic correlations with production traits in wild populations of the European sea bass (*Dicentrarchus labrax*). *Aquaculture* 478, 1–8. <https://doi.org/10.1016/j.aquaculture.2017.05.011>
- Doan, Vandeputte, M., Chatain, B., Morin, T., Allal, F., 2017b. Viral encephalopathy and retinopathy in aquaculture: a review. *J. Fish Dis.* 40, 717–742. <https://doi.org/10.1111/jfd.12541>
- Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. <https://doi.org/10.1126/science.1258096>
- Dufflocq, P., Pérez-Enciso, M., Lhorente, J.P., Yáñez, J.M., 2019. Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. *Aquaculture* 503, 225–230. <https://doi.org/10.1016/j.aquaculture.2018.12.061>
- Dupont-Nivet, M., Vandeputte, M., Haffray, P., Chevassus, B., 2006. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture* 252, 161–170.
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., Gagnaire, P.-A., 2018. The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nat. Commun.* 9, 2518. <https://doi.org/10.1038/s41467-018-04963-6>
- Durel, C.E., Laurens, F., Fouillet, A., Lespinasse, Y., 1998. Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. *TAG Theoretical and Applied Genetics* 96, 1077–1085. <https://doi.org/10.1007/s001220050842>
- Embury, G.C., Hayford, C.O., 1925. The Advantage of Rearing Brook Trout Fingerlings from Selected Breeders. *Transactions of the American Fisheries Society* 55, 135–148. [https://doi.org/10.1577/1548-8659\(1925\)55\[135:TAORBT\]2.0.CO;2](https://doi.org/10.1577/1548-8659(1925)55[135:TAORBT]2.0.CO;2)
- Falconer, D., Mackay, T., 1996. *Introduction to quantitative genetics*, Ed 4. ed. Longmans Green, Harlow, Essex, UK.
- FAO, 2020. *La situation mondiale des pêches et de l'aquaculture 2020*. FAO. <https://doi.org/10.4060/ca9229fr>
- FAO, 1997. *FAOSTAT statistical database*. FAO, Rome.
- FEAP, 2017. *Annual Report 2017 [WWW Document]*. FEAP – The Federation of European Aquaculture Producers. URL https://issuu.com/feapsec/docs/feap_ar2017 (accessed 4.16.20).
- Fjalestad, K.T., Gjedrem, T., Gjerde, B., 1993. Genetic improvement of disease resistance in fish: an overview, in: *Genetics in Aquaculture*. Elsevier, pp. 65–74. <https://doi.org/10.1016/B978-0-444-81527-9.50011-7>
- Foerster, K., Coulson, T., Sheldon, B.C., Pemberton, J.M., Clutton-Brock, T.H., Kruuk, L.E.B., 2007. Sexually antagonistic genetic variation for fitness in red deer. *Nature* 447, 1107–1110. <https://doi.org/10.1038/nature05912>
- Fraslin, C., Brard-Fudulea, S., D'Ambrosio, J., Bestin, A., Charles, M., Haffray, P., Quillet, E., Phocas, F., 2019. Rainbow trout resistance to bacterial cold water disease: two new quantitative trait loci identified after a natural disease outbreak on a French farm. *Anim Genet* 50, 293–297. <https://doi.org/10.1111/age.12777>
- Fraslin, C., Dechamp, N., Bernard, M., Krieg, F., Hervet, C., Guyomard, R., Esquerré, D., Barbieri, J., Kuchly, C., Duchaud, E., Boudinot, P., Rochat, T., Bernardet, J.-F., Quillet, E., 2018. Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. *Genet Sel Evol* 50, 60. <https://doi.org/10.1186/s12711-018-0431-9>

- Fraslin, C., Phocas, F., Bestin, A., Charles, M., Bernard, M., Krieg, F., Dechamp, N., Ciobotaru, C., Hozé, C., Petitprez, F., Milhes, M., Lluch, J., Bouchez, O., Poncet, C., Hocdé, P., Haffray, P., Guiguen, Y., Quillet, E., 2020a. Genetic determinism of spontaneous masculinisation in XX female rainbow trout: new insights using medium throughput genotyping and whole-genome sequencing. *Sci Rep* 10, 17693. <https://doi.org/10.1038/s41598-020-74757-8>
- Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.-F., Collet, B., Lallias, D., Boudinot, P., 2020b. Combining Multiple Approaches and Models to Dissect the Genetic Architecture of Resistance to Infections in Fish. *Front. Genet.* 11, 677. <https://doi.org/10.3389/fgene.2020.00677>
- Georges, M., Dietz, A.B., Mishra, A., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., Leipold, H., Womack, J.E., 1993. Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc. Natl. Acad. Sci.* 90, 1058–1062. <https://doi.org/10.1073/pnas.90.3.1058>
- Gitterle, T., Ødegård, J., Gjerde, B., Rye, M., Salte, R., 2006. Genetic parameters and accuracy of selection for resistance to White Spot Syndrome Virus (WSSV) in *Penaeus (Litopenaeus) vannamei* using different statistical models. *Aquaculture* 251, 210–218. <https://doi.org/10.1016/j.aquaculture.2005.06.008>
- Gjedrem, T., 2015. Disease Resistant Fish and Shellfish Are within Reach: A Review. *JMSE* 3, 146–153. <https://doi.org/10.3390/jmse3010146>
- Gjedrem, T., 1985. Improvement of productivity through breeding schemes. *GeoJournal* 10, 233–241.
- Gjedrem, T., Gjøen, H.M., 1995. Genetic variation in susceptibility of Atlantic salmon, *Salmo salar* L., to furunculosis, BKD and cold water vibriosis. *Aquaculture Res* 26, 129–134. <https://doi.org/10.1111/j.1365-2109.1995.tb00892.x>
- Gjedrem, T., Robinson, N., 2014. Advances by Selective Breeding for Aquatic Species: A Review. *Agric. Sci.* 05, 1152–1158. <https://doi.org/10.4236/as.2014.512125>
- Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 350–353, 117–129. <https://doi.org/10.1016/j.aquaculture.2012.04.008>
- Glaubitz, J.C., Rhodes, O.E., Dewoody, J.A., 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology* 12, 1039–1047. <https://doi.org/10.1046/j.1365-294X.2003.01790.x>
- Goddard, M., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- Gomez-Casado, E., Estepa, A., Coll, J.M., 2011. A comparative review on European-farmed finfish RNA viruses and their vaccines. *Vaccine* 29, 2657–2671. <https://doi.org/10.1016/j.vaccine.2011.01.097>
- Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B.M., Kenney, P.B., Vallejo, R.L., Palti, Y., Leeds, T.D., 2016. Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). *Front. Genet.* 7. <https://doi.org/10.3389/fgene.2016.00203>
- Grashei, K.E., Ødegård, J., Meuwissen, T.H.E., 2018. Using genomic relationship likelihood for parentage assignment. *Genetics Selection Evolution* 50. <https://doi.org/10.1186/s12711-018-0397-7>
- Gratacap, R.L., Wargelius, A., Edvardsen, R.B., Houston, R.D., 2019. Potential of Genome Editing to Improve Aquaculture Breeding and Production. *Trends in Genetics* 35, 672–684. <https://doi.org/10.1016/j.tig.2019.06.006>
- Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., Vandeputte, M., 2020. APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Mol. Ecol. Resour.* 20, 579–590. <https://doi.org/10.1111/1755-0998.13103>
- Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin, T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.-S., Peyrou, B., Gagnaire, P.-A., Vandeputte, M., 2021. Genome-wide association studies for resistance to viral nervous

- necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. *Aquaculture* 530, 735930.
<https://doi.org/10.1016/j.aquaculture.2020.735930>
- Gruner, L., Bouix, J., Brunel, J.C., 2004. High genetic correlation between resistance to *Haemonchus contortus* and to *Trichostrongylus colubriformis* in INRA 401 sheep. *Veterinary Parasitology* 119, 51–58. <https://doi.org/10.1016/j.vetpar.2003.10.014>
- Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J., 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186. <https://doi.org/10.1186/1471-2105-12-186>
- Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., Thaller, G., 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* 42, 5.
- Haley, C.S., Knott, S.A., 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315–324. <https://doi.org/10.1038/hdy.1992.131>
- Havenstein, G., Ferket, P., Qureshi, M., 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poultry Science* 82, 1500–1508. <https://doi.org/10.1093/ps/82.10.1500>
- Hayes, B., 2013. Overview of Statistical Methods for Genome-Wide Association Studies (GWAS), in: Gondro, C., van der Werf, J., Hayes, B. (Eds.), *Genome-Wide Association Studies and Genomic Prediction, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 149–169.
https://doi.org/10.1007/978-1-62703-447-0_6
- Hayes, B., Daetwyler, H., Bowman, P., Moser, G., Tier, B., Crump, R., Khatkar, M., Raadsma, H., Goddard, M., others, 2009. Accuracy of genomic selection: comparing theory and results, in: *Proc Assoc Advmt Anim Breed Genet*. pp. 34–37.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443.
<https://doi.org/10.3168/jds.2008-1646>
- Hendrikson, R.G., Zenoble, R.D., 1983. Vibriosis in fish: a review. *Iowa state university Veterinarian* 45, 5.
- Hérault, F., Damon, M., Cherel, P., Le Roy, P., 2018. Combined GWAS and LDLA approaches to improve genome-wide quantitative trait loci detection affecting carcass and meat quality traits in pig. *Meat Science* 135, 148–158. <https://doi.org/10.1016/j.meatsci.2017.09.015>
- Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M., Cook, D., 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245–256.
[https://doi.org/10.1016/0044-8486\(95\)01109-9](https://doi.org/10.1016/0044-8486(95)01109-9)
- Hjeltnes, B., 2014. Fish health report 2013. Norwegian Veterinary Institute, Oslo.
- Hocking, P.M., 2005. Review of QTL mapping results in chickens. *World's Poultry Science Journal* 61, 215–226. <https://doi.org/10.1079/WPS200461>
- Houston, R.D., 2017. Future directions in breeding for disease resistance in aquaculture species. *R. Bras. Zootec.* 46, 545–551. <https://doi.org/10.1590/s1806-92902017000600010>
- Houston, R.D., Gheyas, A., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J., Haley, C.S., Bishop, S.C., 2008. Detection and Confirmation of a Major QTL Affecting Resistance to Infectious Pancreatic Necrosis (IPN) in Atlantic Salmon (*Salmo Salar*), in: Pinard, M.-H., Gay, C., Pastoret, P.-P., Dodet, B. (Eds.), *Animal Genomics for Animal Health*. Development Biological, Basel, pp. 199–204. <https://doi.org/10.1159/000317160>
- Ibeagha-Awemu, E.M., Zhao, X., 2015. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Front. Genet.* 6. <https://doi.org/10.3389/fgene.2015.00302>
- Ina-Salwany, M.Y., Al-saari, N., Mohamad, A., Mursidi, F., Mohd-Aris, A., Amal, M.N.A., Kasai, H., Mino, S., Sawabe, T., Zamri-Saad, M., 2019. Vibriosis in Fish: A Review on Disease Development and Prevention. *J Aquat Anim Health* 31, 3–22.
<https://doi.org/10.1002/aah.10045>

- International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945. <https://doi.org/10.1038/nature03001>
- Jamieson, A., C S Taylor, St., 1997. Comparisons of three probability formulae for parentage exclusion. *Animal Genetics* 28, 397–400. <https://doi.org/10.1111/j.1365-2052.1997.00186.x>
- Jammes, H., Junien, C., Chavatte-Palmer, P., 2011. Epigenetic control of development and expression of quantitative traits. *Reprod. Fertil. Dev.* 23, 64. <https://doi.org/10.1071/RD10259>
- Janssen, K., Bijma, P., 2020. The economic value of R0 for selective breeding against microparasitic diseases. *Genet Sel Evol* 52, 3. <https://doi.org/10.1186/s12711-020-0526-y>
- Janssen, K., Chavanne, H., Berentsen, P., Komen, H., 2017. Impact of selective breeding on European aquaculture. *Aquaculture* 472, 8–16. <https://doi.org/10.1016/j.aquaculture.2016.03.012>
- Jia, Y., Jannink, J.-L., 2012. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* 192, 1513–1522. <https://doi.org/10.1534/genetics.112.144246>
- Jiang, C., Zeng, Z.B., 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140, 1111–1127.
- Jones, A.G., Small, C.M., Paczolt, K.A., Ratterman, N.L., 2010. A practical guide to methods of parentage analysis: TECHNICAL REVIEW. *Molecular Ecology Resources* 10, 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
- Joshi, R., Skaarud, A., de Vera, M., Alvarez, A.T., Ødegård, J., 2020. Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516, 734641. <https://doi.org/10.1016/j.aquaculture.2019.734641>
- Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment: CERVUS LIKELIHOOD MODEL. *Molecular Ecology* 16, 1099–1106. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>
- Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *J. Am. Stat. Assoc.* 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kennedy, B.W., Quinton, M., van Arendonk, J.A.M., 1992. Estimation of effects of single genes on quantitative traits. *Journal of Animal Science* 70, 2000–2012. <https://doi.org/10.2527/1992.7072000x>
- Kincaid, H.L., 1983. Inbreeding in fish populations used for aquaculture. *Aquaculture* 33, 215–227. [https://doi.org/10.1016/0044-8486\(83\)90402-7](https://doi.org/10.1016/0044-8486(83)90402-7)
- Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 29. <https://doi.org/10.1186/1746-4811-9-29>
- Kriaridou, C., Tsairidou, S., Houston, R.D., Robledo, D., 2020. Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. *Front. Genet.* 11, 124. <https://doi.org/10.3389/fgene.2020.00124>
- Kruuk, L.E.B., 2004. Estimating genetic parameters in natural populations using the “animal model.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, 873–890. <https://doi.org/10.1098/rstb.2003.1437>
- Kubota, S., Kimura, M., Egusa, S., 1970. Studies of a bacterial tuberculoidosis of the yellowtail-I Symptomatology and histopathology. *Fish Pathol.* 4, 111–118. <https://doi.org/10.3147/jsfp.4.111>
- LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research* 37, 4181–4193. <https://doi.org/10.1093/nar/gkp552>
- Lande, R., Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.
- Le Breton, A., Grisez, L., Sweetman, J., Ollevier, F., 1997. Viral nervous necrosis (VNN) associated with mass mortalities in cage-reared sea bass, *Dicentrarchus labrax* (L.). *J. Fish Diseases* 20, 145–151. <https://doi.org/10.1046/j.1365-2761.1997.00284.x>
- Legarra, A., Croiseau, P., Sanchez, M., Teyssèdre, S., Sallé, G., Allais, S., Fritz, S., Moreno, C., Ricard, A., Elsen, J.-M., 2015. A comparison of methods for whole-genome QTL mapping using dense

- markers in four livestock species. *Genet. Sel. Evol.* 47, 6. <https://doi.org/10.1186/s12711-015-0087-7>
- Legarra, A., Robert-Granié, C., Manfredi, E., Elsen, J.-M., 2008. Performance of Genomic Selection in Mice. *Genetics* 180, 611–618. <https://doi.org/10.1534/genetics.108.088575>
- Lewis, R., Brotherstone, S., 2002. A genetic evaluation of growth in sheep using random regression techniques.
- Li, Y., Wang, L., Yang, Y., Li, X., Dai, H., Chen, S., 2019. Genetic analysis of disease resistance to *Vibrio harveyi* by challenge test in Chinese tongue sole (*Cynoglossus semilaevis*). *Aquaculture* 503, 430–435. <https://doi.org/10.1016/j.aquaculture.2019.01.011>
- Lin, Z., Hayes, B.J., Daetwyler, H.D., 2014. Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 65, 1177. <https://doi.org/10.1071/CP13363>
- Liu, Z., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G., Reents, R., 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol* 43, 19. <https://doi.org/10.1186/1297-9686-43-19>
- Lourenco, D.A.L., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T.J., Forni, S., Weller, J.I., 2014. Are evaluations on young genotyped animals benefiting from the past generations? *Journal of Dairy Science* 97, 3930–3942. <https://doi.org/10.3168/jds.2013-7769>
- Lyu, D., Yu, Y., Zhang, Q., Luo, Z., Wang, Q., Xiang, J., Li, F., 2020. Estimating genetic parameters for resistance to *Vibrio parahaemolyticus* with molecular markers in Pacific white shrimp. *Aquaculture* 527, 735439. <https://doi.org/10.1016/j.aquaculture.2020.735439>
- Magnadottir, B., 2010. Immunological Control of Fish Diseases. *Mar Biotechnol* 12, 361–379. <https://doi.org/10.1007/s10126-010-9279-x>
- Marco-Noales, E., Milán, M., Fouz, B., Sanjuán, E., Amaro, C., 2001. Transmission to Eels, Portals of Entry, and Putative Reservoirs of *Vibrio vulnificus* Serovar E (Biotype 2). *Appl. Environ. Microbiol.* 67, 4717–4725. <https://doi.org/10.1128/AEM.67.10.4717-4725.2001>
- Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7, 639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>
- McCarthy, D.H., 1976. *Vibrio* disease in eels. *J Fish Biology* 8, 317–320. <https://doi.org/10.1111/j.1095-8649.1976.tb03955.x>
- Meuwissen, T., Hayes, B., Goddard, M., 2016. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* 6, 6–14. <https://doi.org/10.2527/af.2016-0002>
- Meuwissen, T., Hayes, B., Goddard, M., 2013. Accelerating Improvement of Livestock with Genomic Selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. <https://doi.org/10.1146/annurev-animal-031412-103705>
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157, 1819.
- Miccoli, A., Saraceni, P.R., Scapigliati, G., 2019. Vaccines and immune protection of principal Mediterranean marine fish species. *Fish Shellfish Immun.* 94, 800–809. <https://doi.org/10.1016/j.fsi.2019.09.065>
- Michenet, A., Barbat, M., Saintilan, R., Venot, E., Phocas, F., 2016. Detection of quantitative trait loci for maternal traits using high-density genotypes of Blonde d'Aquitaine beef cattle. *BMC Genet.* 17, 88. <https://doi.org/10.1186/s12863-016-0397-y>
- Michot, P., Fritz, S., Barbat, A., Boussaha, M., Deloche, M.-C., Grohs, C., Hoze, C., Le Berre, L., Le Bourhis, D., Desnoes, O., Salvetti, P., Schibler, L., Boichard, D., Capitan, A., 2017. A missense mutation in PFAS (phosphoribosylformylglycinamide synthase) is likely causal for embryonic lethality associated with the MH1 haplotype in Montbéliarde dairy cattle. *Journal of Dairy Science* 100, 8176–8187. <https://doi.org/10.3168/jds.2017-12579>
- Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17, 240–248. <https://doi.org/10.1101/gr.5681207>

- Misztal, I., Lawlor, T.J., Short, T.H., VanRaden, P.M., 1992. Multiple-Trait Estimation of Variance Components of Yield and Type Traits Using an Animal Model. *Journal of Dairy Science* 75, 544–551. [https://doi.org/10.3168/jds.S0022-0302\(92\)77791-1](https://doi.org/10.3168/jds.S0022-0302(92)77791-1)
- Misztal, I., Lourenco, D., Legarra, A., 2020. Current status of genomic evaluation. *Journal of Animal Science* 98, skaa101. <https://doi.org/10.1093/jas/skaa101>
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D., 2002. BLUPF90 and related programs (BGF90). *Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Communication 28*, 743–744.
- Moen, T., Baranski, M., Sonesson, A.K., Kjøglum, S., 2009. 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* 10, 368. <https://doi.org/10.1186/1471-2164-10-368>
- Moen, T., Torgersen, J., Santi, N., Davidson, W.S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., Lubieniecki, K.P., Isdal, E., Lien, S., 2015. Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics* 200, 1313–1326. <https://doi.org/10.1534/genetics.115.175406>
- Moghaddar, N., Swan, A.A., van der Werf, J.H.J., 2017. Genomic prediction from observed and imputed high-density ovine genotypes. *Genet Sel Evol* 49, 40. <https://doi.org/10.1186/s12711-017-0315-4>
- Mori, K.-I., Nakai, T., Muroga, K., Arimoto, M., Mushiake, K., Furusawa, I., 1992. Properties of a new virus belonging to nodaviridae found in larval striped jack (*Pseudocaranx dentex*) with nervous necrosis. *Virology* 187, 368–371. [https://doi.org/10.1016/0042-6822\(92\)90329-N](https://doi.org/10.1016/0042-6822(92)90329-N)
- Morin, P.A., Luikart, G., Wayne, R.K., the SNP workshop group, 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* 19, 208–216. <https://doi.org/10.1016/j.tree.2004.01.009>
- Morris, C.A., Towers, N.R., Wheeler, M., Amyes, N.C., 1995. A note on the genetics of resistance or susceptibility to ryegrass staggers in sheep. *New Zealand Journal of Agricultural Research* 38, 367–371. <https://doi.org/10.1080/00288233.1995.9513138>
- Nash, C.E., 2011. The History of Aquaculture. <https://doi.org/10.1002/9780470958971>
- Neuner, S., Edel, C., Emmerling, R., Thaller, G., Götz, K.-U., 2009. Precision of genetic parameters and breeding values estimated in marker assisted BLUP genetic evaluation. *Genet Sel Evol* 41, 26. <https://doi.org/10.1186/1297-9686-41-26>
- Nielsen, H.M., Sonesson, A.K., Meuwissen, T.H.E., 2011. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *Journal of Animal Science* 89, 630–638. <https://doi.org/10.2527/jas.2009-2731>
- Nones, K., Ledur, M.C., Ruy, D.C., Baron, E.E., Melo, C.M.R., Moura, A.S.A.M.T., Zanella, E.L., Burt, D.W., Coutinho, L.L., 2006. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. *Animal Genetics* 37, 95–100. <https://doi.org/10.1111/j.1365-2052.2005.01387.x>
- Oddou-Muratorio, S., Houot, M.L., Demesure-Musch, B., Austerlitz, F., 2003. Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. *Molecular Ecology* 12, 3427–3439. <https://doi.org/10.1046/j.1365-294X.2003.01989.x>
- Ødegård, J., Baranski, M., Gjerde, B., Gjedrem, T., 2011. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects: Genetic evaluation of disease resistance. *Aquac. Res.* 42, 103–114. <https://doi.org/10.1111/j.1365-2109.2010.02669.x>
- Ødegård, J., Meuwissen, T.H., 2012. Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* 44, 16. <https://doi.org/10.1186/1297-9686-44-16>

- Ødegård, J., Sonesson, A.K., Yazdi, M.H., Meuwissen, T.H., 2009. Introgression of a major QTL from an inferior into a superior population using genomic selection. *Genet Sel Evol* 41, 38. <https://doi.org/10.1186/1297-9686-41-38>
- Oliehoek, P.A., Bijma, P., 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genetics Selection Evolution* 41, 9. <https://doi.org/10.1186/1297-9686-41-9>
- Palaiokostas, C., Bekaert, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Chatain, B., Penman, D.J., Vandeputte, M., 2015. A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genet. Sel. Evol.* 47, 68. <https://doi.org/10.1186/s12711-015-0148-y>
- Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.-S., Haffray, P., Morin, T., Cabon, J., Allal, F., Vandeputte, M., Houston, R.D., 2018a. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet. Sel. Evol.* 50, 30. <https://doi.org/10.1186/s12711-018-0401-2>
- Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.-S., Haffray, P., Morin, T., Cabon, J., Allal, F., Vandeputte, M., Houston, R.D., 2018b. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genetics Selection Evolution* 50. <https://doi.org/10.1186/s12711-018-0401-2>
- Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R.D., Bargelloni, L., 2016. Genomic Prediction of Resistance to Pasteurellosis in Gilthead Sea Bream (*Sparus aurata*) Using 2b-RAD Sequencing. *G3* 6, 3693–3700. <https://doi.org/10.1534/g3.116.035220>
- Palaiokostas, C., Kocour, M., Prchal, M., Houston, R.D., 2018c. Accuracy of Genomic Evaluations of Juvenile Growth Rate in Common Carp (*Cyprinus carpio*) Using Genotyping by Sequencing. *Front. Genet.* 9, 82. <https://doi.org/10.3389/fgene.2018.00082>
- Pampouille, E., Berri, C., Boitard, S., Hennequet-Antier, C., Beauclercq, S.A., Godet, E., Praud, C., Jégo, Y., Le Bihan-Duval, E., 2018. Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. *BMC Genomics* 19, 202. <https://doi.org/10.1186/s12864-018-4598-9>
- Pauletto, M., Manousaki, T., Ferraresso, S., Babbucci, M., Tsakogiannis, A., Louro, B., Vitulo, N., Quoc, V.H., Carraro, R., Bertotto, D., Franch, R., Maroso, F., Aslam, M.L., Sonesson, A.K., Simionati, B., Malacrida, G., Cestaro, A., Caberlotto, S., Sarropoulou, E., Mylonas, C.C., Power, D.M., Patarnello, T., Canario, A.V.M., Tsigenopoulos, C., Bargelloni, L., 2018. Genomic analysis of *Sparus aurata* reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. *Commun Biol* 1, 119. <https://doi.org/10.1038/s42003-018-0122-7>
- Pemberton, J.M., Slate, J., Bancroft, D.R., Barrett, J.A., 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* 4, 249–252. <https://doi.org/10.1111/j.1365-294X.1995.tb00214.x>
- Person-Le Ruyet, J., Mahé, K., Le Bayon, N., Le Delliou, H., 2004. Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass, *Dicentrarchus labrax*. *Aquaculture* 237, 269–280. <https://doi.org/10.1016/j.aquaculture.2004.04.021>
- Porto-Neto, L.R., Sonstegard, T.S., Liu, G.E., Bickhart, D.M., Da Silva, M.V., Machado, M.A., Utsunomiya, Y.T., Garcia, J.F., Gondro, C., Van Tassell, C.P., 2013. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14, 876. <https://doi.org/10.1186/1471-2164-14-876>
- Pousao-Ferreira, P., Morais, S., Dores, E., Narciso, L., 1999. Eggs of gilthead seabream *Sparus aurata* L. as a potential enrichment product of *Brachionus* sp. in the larval rearing of gilthead seabream *Sparus aurata* L. *Aquac Research* 30, 751–758. <https://doi.org/10.1046/j.1365-2109.1999.00394.x>
- Prayitno, S.B., Latchford, J.W., 1995. Experimental infections of crustaceans with luminous bacteria related to *Photobacterium* and *Vibrio*. Effect of salinity and pH on infectiosity. *Aquaculture* 132, 105–112. [https://doi.org/10.1016/0044-8486\(94\)00374-W](https://doi.org/10.1016/0044-8486(94)00374-W)

- Price, R., Templeton, J., Smith, R., Adams, L., 1990. Ability of mononuclear phagocytes from cattle naturally resistant or susceptible to brucellosis to control in vitro intracellular survival of *Brucella abortus*. *Infection and immunity* 58, 879–886.
- Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P., 2000. Association Mapping in Structured Populations. *Am. J. Hum. Genet.* 67, 170–181. <https://doi.org/10.1086/302959>
- Pryce, J.E., Gredler, B., Bolormaa, S., Bowman, P.J., Egger-Danner, C., Fuerst, C., Emmerling, R., Sölkner, J., Goddard, M.E., Hayes, B.J., 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. *Journal of Dairy Science* 94, 2625–2630. <https://doi.org/10.3168/jds.2010-3719>
- Purcell, S., Chang, C., 2015. PLINK 1.9. URL <https://www.cog-genomics.org/plink2>.
- Quillet, E., Dorson, M., Aubard, G., Torhy, C., 2001. In vitro viral haemorrhagic septicaemia virus replication in excised fins of rainbow trout: correlation with resistance to waterborne challenge and genetic variation. *Dis. Aquat. Org.* 45, 171–182. <https://doi.org/10.3354/dao045171>
- Rajsic, P., Weersink, A., Navabi, A., Peter Pauls, K., 2016. Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. *Euphytica* 210, 259–276. <https://doi.org/10.1007/s10681-016-1716-0>
- Ranade, K., Chang, M.S., Ting, C.T., Pei, D., Hsiao, C.F., Olivier, M., Pesich, R., Hebert, J., Chen, Y.D., Dzau, V.J., Curb, D., Olshen, R., Risch, N., Cox, D.R., Botstein, D., 2001. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res.* 11, 1262–1268. <https://doi.org/10.1101/gr.157801>
- Rastas, P., 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. <https://doi.org/10.1093/bioinformatics/btx494>
- Rauw, W.M., Kanis, E., Noordhuizen-Stassen, E.N., Grommers, F.J., 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science* 56, 15–33. [https://doi.org/10.1016/S0301-6226\(98\)00147-X](https://doi.org/10.1016/S0301-6226(98)00147-X)
- Reno, P.W., 1998. Factors Involved in the Dissemination of Disease in Fish Populations. *Journal of Aquatic Animal Health* 10, 160–171. [https://doi.org/10.1577/1548-8667\(1998\)010<0160:FIITDO>2.0.CO;2](https://doi.org/10.1577/1548-8667(1998)010<0160:FIITDO>2.0.CO;2)
- Reverter, M., Sarter, S., Caruso, D., Avarre, J.-C., Combe, M., Pepey, E., Pouyaud, L., Vega-Heredía, S., de Verdal, H., Gozlan, R.E., 2020. Aquaculture at the crossroads of global warming and antimicrobial resistance. *Nat Commun* 11, 1870. <https://doi.org/10.1038/s41467-020-15735-6>
- Riedelsheimer, C., Melchinger, A.E., 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor Appl Genet* 126, 2835–2848. <https://doi.org/10.1007/s00122-013-2175-9>
- Rincint, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., Rodríguez, V.M., Moreno-Gonzalez, J., Melchinger, A., Bauer, E., Schoen, C.-C., Meyer, N., Giauffret, C., Bauland, C., Jamin, P., Laborde, J., Monod, H., Flament, P., Charcosset, A., Moreau, L., 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). *Genetics* 192, 715–728. <https://doi.org/10.1534/genetics.112.141473>
- Robinson, N., Hayes, B., 2008. Modelling the use of gene expression profiles with selective breeding for improved disease resistance in Atlantic salmon (*Salmo salar*). *Aquaculture* 285, 38–46. <https://doi.org/10.1016/j.aquaculture.2008.08.016>
- Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev Aquacult* 10, 670–682. <https://doi.org/10.1111/raq.12193>
- Romalde, J.L., 2002. *Photobacterium damsela* subsp. *piscicida*: an integrated view of a bacterial fish pathogen. *Int Microbiol* 5, 3–9. <https://doi.org/10.1007/s10123-002-0051-6>

- Rothschild, M.F., Hu, Z., Jiang, Z., 2007. Advances in QTL Mapping in Pigs. *Int. J. Biol. Sci.* 192–197. <https://doi.org/10.7150/ijbs.3.192>
- Salgon, S., Jourda, C., Sauvage, C., Daunay, M.-C., Reynaud, B., Wicker, E., Dintinger, J., 2017. Eggplant Resistance to the *Ralstonia solanacearum* Species Complex Involves Both Broad-Spectrum and Strain-Specific Quantitative Trait Loci. *Front. Plant Sci.* 8, 828. <https://doi.org/10.3389/fpls.2017.00828>
- Sanchez, M.-P., Govignon-Gion, A., Croiseau, P., Fritz, S., Hozé, C., Miranda, G., Martin, P., Barbat-Leterrier, A., Letaïef, R., Rocha, D., Brochard, M., Boussaha, M., Boichard, D., 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genet Sel Evol* 49, 68. <https://doi.org/10.1186/s12711-017-0344-z>
- San Cristobal, M., Chevalet, C., 1997. Error tolerant parent identification from a finite set of individuals. *Genetical Research* 70, 53–62. <https://doi.org/10.1017/S0016672397002851>
- Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15, 478. <https://doi.org/10.1186/1471-2164-15-478>
- Smith, B.J., 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *J. Stat. Softw.* 21, 1–37.
- Snieszko, S.F., Bullock, G.L., Hollis, E., Boone, J.G., 1964. PASTEURILLA SP. FROM AN EPIZOOTIC OF WHITE PERCH (*ROCCUS AMERICANUS*) IN CHESAPEAKE BAY TIDEWATER AREAS. *J. Bacteriol.* 88, 1814–1815. <https://doi.org/10.1128/JB.88.6.1814-1815.1964>
- Sodeland, M., Gaarder, M., Moen, T., Thomassen, M., Kjøglum, S., Kent, M., Lien, S., 2013. Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon. *Aquaculture* 408–409, 169–174. <https://doi.org/10.1016/j.aquaculture.2013.05.029>
- Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Ødegard, J., Meuwissen, T.H., 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.* 41, 53. <https://doi.org/10.1186/1297-9686-41-53>
- Sommerset, I., Krossøy, B., Biering, E., Frost, P., 2005. Vaccines for fish in aquaculture. *Expert Rev. Vaccines* 4, 89–101. <https://doi.org/10.1586/14760584.4.1.89>
- Sonesson, A.K., Gjerde, B., Robinson, N., 2011. A simple selection scheme to improve disease resistance and growth. *Aquaculture* 319, 337–341. <https://doi.org/10.1016/j.aquaculture.2011.07.009>
- Sonesson, A.K., Meuwissen, T.H., 2009a. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol* 41, 37. <https://doi.org/10.1186/1297-9686-41-37>
- Sonesson, A.K., Meuwissen, T.H., 2009b. Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol* 41, 37. <https://doi.org/10.1186/1297-9686-41-37>
- Sonesson, A.K., Ødegård, J., 2016. Mating structures for genomic selection breeding programs in aquaculture. *Genet Sel Evol* 48, 46. <https://doi.org/10.1186/s12711-016-0224-y>
- Storset, A., Moen, T., Ødegard, J., Santi, N., 2015. Pioneering in the implementation of genomic information in fish breeding.
- Tacon, A.G.J., Metian, M., 2009. Fishing for Aquaculture: Non-Food Use of Small Pelagic Forage Fish—A Global Perspective. *Reviews in Fisheries Science* 17, 305–317. <https://doi.org/10.1080/10641260802677074>
- Tacon, A.G.J., Metian, M., 2008. Global overview on the use of fish meal and fish oil in industrially compounded aquafeeds: Trends and future prospects. *Aquaculture* 285, 146–158. <https://doi.org/10.1016/j.aquaculture.2008.08.015>
- Talukder, Z.I., Tharreau, D., Price, A.H., 2004. Quantitative trait loci analysis suggests that partial resistance to rice blast is mostly determined by race-specific interactions. *New Phytologist* 162, 197–209.

- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., Meyre, D., 2019. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. <https://doi.org/10.1038/s41576-019-0127-1>
- Taylor, J.F., 2014. Implementation and accuracy of genomic selection. *Aquaculture* 420–421, S8–S14. <https://doi.org/10.1016/j.aquaculture.2013.02.017>
- Taylor, R.S., Kube, P.D., Muller, W.J., Elliott, N.G., 2009. Genetic variation of gross gill pathology and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease challenge. *Aquaculture* 294, 172–179. <https://doi.org/10.1016/j.aquaculture.2009.06.007>
- Thiéry, R., Cozien, J., de Boissésou, C., Kerbart-Boscher, S., Névarez, L., 2004. Genomic classification of new betanodavirus isolates by phylogenetic analysis of the coat protein gene suggests a low host-fish species specificity. *J. Gen. Virol.* 85, 3079–3087. <https://doi.org/10.1099/vir.0.80264-0>
- Thompson, F.L., Iida, T., Swings, J., 2004. Biodiversity of *Vibrios*. *MMBR* 68, 403–431. <https://doi.org/10.1128/MMBR.68.3.403-431.2004>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* 5, 5770. <https://doi.org/10.1038/ncomms6770>
- Toranzo, A.E., Barreiro, S., Casal, J.F., Figueras, A., Magarinos, B., Barja, J.L., 1991. Pasteurellosis in cultured gilthead seabream (*Sparus aurata*): first report in Spain. *Aquaculture* 99, 1–15. [https://doi.org/10.1016/0044-8486\(91\)90284-E](https://doi.org/10.1016/0044-8486(91)90284-E)
- Travers, M.-A., Tourbiez, D., Parizadeh, L., Haffner, P., Kozic-Djellouli, A., Aboubaker, M., Koken, M., Dégremont, L., Lupo, C., 2017. Several strains, one disease: experimental investigation of *Vibrio aestuarianus* infection parameters in the Pacific oyster, *Crassostrea gigas*. *Vet Res* 48, 32. <https://doi.org/10.1186/s13567-017-0438-1>
- Tribout, T., Larzul, C., Phocas, F., 2013. Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genet Sel Evol* 45, 40. <https://doi.org/10.1186/1297-9686-45-40>
- Tsai, H.-Y., Hamilton, A., Tinch, A.E., Guy, D.R., Bron, J.E., Taggart, J.B., Gharbi, K., Stear, M., Matika, O., Pong-Wong, R., Bishop, S.C., Houston, R.D., 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet Sel Evol* 48, 47. <https://doi.org/10.1186/s12711-016-0226-9>
- Tsai, H.-Y., Hamilton, A., Tinch, A.E., Guy, D.R., Gharbi, K., Stear, M.J., Matika, O., Bishop, S.C., Houston, R.D., 2015. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16, 969. <https://doi.org/10.1186/s12864-015-2117-9>
- Tsai, H.-Y., Matika, O., Edwards, S.M., Antolín-Sánchez, R., Hamilton, A., Guy, D.R., Tinch, A.E., Gharbi, K., Stear, M.J., Taggart, J.B., Bron, J.E., Hickey, J.M., Houston, R.D., 2017. Genotype Imputation To Improve the Cost-Efficiency of Genomic Selection in Farmed Atlantic Salmon. *G3* 7, 1377–1383. <https://doi.org/10.1534/g3.117.040717>
- Tsairidou, S., Anacleto, O., Woolliams, J.A., Doeschl-Wilson, A., 2019. Enhancing genetic disease control by selecting for lower host infectivity and susceptibility. *Heredity* 122, 742–758. <https://doi.org/10.1038/s41437-018-0176-9>
- Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E., Houston, R.D., 2020. Optimizing Low-Cost Genotyping and Imputation Strategies for Genomic Selection in Atlantic Salmon. *G3* 10, 581–590. <https://doi.org/10.1534/g3.119.400800>
- Tsuruta, S., Misztal, I., 2006. THRGIBBS1F90 for estimation of variance components with threshold-linear models. Presented at the Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil. Instituto Prociencia, Belo Horizonte, Brazil., pp. 27–31.
- Tsuruta, S., Misztal, I., 2006. THRGIBBS1F90 for estimation of variance components with threshold-linear models. *Journal of Animal Science* 27–31.

- Ulmer, J.B., Valley, U., Rappuoli, R., 2006. Vaccine manufacturing: challenges and solutions. *Nat. Biotechnol.* 24, 1377–1383. <https://doi.org/10.1038/nbt1261>
- Vallejo, R.L., Palti, Y., Liu, S., Evenhuis, J.P., Gao, G., Rexroad, C.E., Wiens, G.D., 2014. Detection of QTL in Rainbow Trout Affecting Survival When Challenged with *Flavobacterium psychrophilum*. *Mar Biotechnol* 16, 349–360. <https://doi.org/10.1007/s10126-013-9553-9>
- van den Berg, S., Vandenplas, J., van Eeuwijk, F.A., Bouwman, A.C., Lopes, M.S., Veerkamp, R.F., 2019. Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies. *Genet Sel Evol* 51, 2. <https://doi.org/10.1186/s12711-019-0445-y>
- Vandeputte, M., 2012. An accurate formula to calculate exclusion power of marker sets in parentage assignment. *Genetics Selection Evolution* 44, 36. <https://doi.org/10.1186/1297-9686-44-36>
- Vandeputte, M., Baroiller, J.-F., Haffray, P., Quillet, E., 2009. Amélioration génétique des poissons: quelles réalisations et quels défis pour demain? *Agriculture* 18, 262–269. <https://doi.org/10.1684/agr.2009.0291>
- Vandeputte, M., Gagnaire, P. -A., Allal, F., 2019. The European sea bass: a key marine fish model in the wild and in aquaculture. *Anim Genet* 50, 195–206. <https://doi.org/10.1111/age.12779>
- Vandeputte, Marc, Haffray, P., 2014. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Front. Genet.* 5. <https://doi.org/10.3389/fgene.2014.00432>
- Vandeputte, M., Haffray, P., 2014. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics* 5. <https://doi.org/10.3389/fgene.2014.00432>
- Vandeputte, M., Mauger, S., Dupont-Nivet, M., 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes* 6, 265–267. <https://doi.org/10.1111/j.1471-8286.2005.01167.x>
- VanRaden, Paul M., O’Connell, Jeffrey R., Wiggans, G.R., Weigel, K.A., 2011. Genomic evaluations with many more genotypes. *Genetics Selection Evolution* 43, 10. <https://doi.org/10.1186/1297-9686-43-10>
- VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- VanRaden, P.M., Sanders, A.H., Tooker, M.E., Miller, R.H., Norman, H.D., Kuhn, M.T., Wiggans, G.R., 2004. Development of a National Genetic Evaluation for Cow Fertility. *Journal of Dairy Science* 87, 2285–2292. [https://doi.org/10.3168/jds.S0022-0302\(04\)70049-1](https://doi.org/10.3168/jds.S0022-0302(04)70049-1)
- Vendramin, N., Toffan, A., Mancin, M., Cappelozza, E., Panzarin, V., Bovo, G., Cattoli, G., Capua, I., Terregino, C., 2014. Comparative pathogenicity study of ten different betanodavirus strains in experimentally infected European sea bass, *Dicentrarchus labrax* (L.). *J Fish Dis* 37, 371–383. <https://doi.org/10.1111/jfd.12117>
- Vendramin, N., Zrncic, S., Padros, F., Oraic, D., Le Breton, A., Zarza, C., Olesen, N.J., 2016. Fish health in Mediterranean Aquaculture, past mistakes and future challenges. *Bulletin of The European Association of Fish Pathologists* 36, 38–45.
- Verrier, E., Langevin, C., Tohy, C., Houel, A., Ducrocq, V., Benmansour, A., Quillet, E., Boudinot, P., 2012. Genetic resistance to rhabdovirus infection in teleost fish is paralleled to the derived cell resistance status. *PLoS One* 7, e33935.
- Verrier, E.R., Dorson, M., Mauger, S., Torhy, C., Ciobotaru, C., Hervet, C., Dechamp, N., Genet, C., Boudinot, P., Quillet, E., 2013. Resistance to a Rhabdovirus (VHSV) in Rainbow Trout: Identification of a Major QTL Related to Innate Mechanisms. *PLoS ONE* 8, e55302. <https://doi.org/10.1371/journal.pone.0055302>
- Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34, 275. <https://doi.org/10.1186/1297-9686-34-3-275>
- Vij, S., Kuhl, H., Kuznetsova, I.S., Komissarov, A., Yurchenko, A.A., Van Heusden, P., Singh, S., Thevasagayam, N.M., Prakki, S.R.S., Purushothaman, K., Saju, J.M., Jiang, J., Mbandi, S.K.,

- Jonas, M., Hin Yan Tong, A., Mwangi, S., Lau, D., Ngoh, S.Y., Liew, W.C., Shen, X., Hon, L.S., Drake, J.P., Boitano, M., Hall, R., Chin, C.-S., Lachumanan, R., Korlach, J., Trifonov, V., Kabilov, M., Tupikin, A., Green, D., Moxon, S., Garvin, T., Sedlazeck, F.J., Vurture, G.W., Gopalapillai, G., Kumar Katneni, V., Noble, T.H., Scaria, V., Sivasubbu, S., Jerry, D.R., O'Brien, S.J., Schatz, M.C., Dalmay, T., Turner, S.W., Lok, S., Christoffels, A., Orbán, L., 2016. Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding. *PLoS Genet.* 12, e1005954. <https://doi.org/10.1371/journal.pgen.1005954>
- Villanueva, B., Pong-Wong, R., Grundy, B., Woolliams, J.A., 1999. Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. *Genet Sel Evol* 31, 115. <https://doi.org/10.1186/1297-9686-31-2-115>
- Villanueva, B., Wray, N.R., Thompson, R., 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim. Sci.* 57, 1–13. <https://doi.org/10.1017/S0003356100006541>
- Vincent, A.L., Thacker, B.J., Halbur, P.G., Rothschild, M.F., Thacker, E.L., 2006. An investigation of susceptibility to porcine reproductive and respiratory syndrome virus between two genetically diverse commercial lines of pigs¹. *Journal of Animal Science* 84, 49–57. <https://doi.org/10.2527/2006.84149x>
- Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five Years of GWAS Discovery. *The American Journal of Human Genetics* 90, 7–24. <https://doi.org/10.1016/j.ajhg.2011.11.029>
- Visscher, P.M., Woolliams, J.A., Smith, D., Williams, J.L., 2002. Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. *Journal of Dairy Science* 85, 2368–2375. [https://doi.org/10.3168/jds.S0022-0302\(02\)74317-8](https://doi.org/10.3168/jds.S0022-0302(02)74317-8)
- Wang, L., Liu, P., Huang, S., Ye, B., Chua, E., Wan, Z.Y., Yue, G.H., 2017. Genome-Wide Association Study Identifies Loci Associated with Resistance to Viral Nervous Necrosis Disease in Asian Seabass. *Mar. Biotechnol.* 19, 255–265. <https://doi.org/10.1007/s10126-017-9747-7>
- Wargelius, A., Leininger, S., Skaftnesmo, K.O., Kleppe, L., Andersson, E., Taranger, G.L., Schulz, R.W., Edvardsen, R.B., 2016. Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Scientific reports* 6, 1–8.
- Weller, J.I., Ezra, E., Ron, M., 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. *Journal of Dairy Science* 100, 8633–8644. <https://doi.org/10.3168/jds.2017-12879>
- Weng, Z., Wolc, A., Shen, X., Fernando, R.L., Dekkers, J.C.M., Arango, J., Settar, P., Fulton, J.E., O'Sullivan, N.P., Garrick, D.J., 2016. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genet Sel Evol* 48, 22. <https://doi.org/10.1186/s12711-016-0198-9>
- Whittaker, J.C., Thompson, R., Denham, M.C., 2000. Marker-assisted selection using ridge regression. *Genet. Res.* 75, 249–252. <https://doi.org/10.1017/S0016672399004462>
- Wilson, A., Réale, D., Clements, M.N., Morrissey, M., Postma, E., Walling, C.A., Kruuk, L.E.B., Nussey, D.H., 2010. An ecologist's guide to the animal model. *Journal of Animal Ecology* 79, 13–26. <https://doi.org/10.1111/j.1365-2656.2009.01639.x>
- Yoshikoshi, K., Inoue, K., 1990. Viral nervous necrosis in hatchery-reared larvae and juveniles of Japanese parrotfish, *Oplegnathus fasciatus* (Temminck & Schlegel). *J Fish Diseases* 13, 69–77. <https://doi.org/10.1111/j.1365-2761.1990.tb00758.x>
- Yoshimi, K., Kaneko, T., Voigt, B., Mashimo, T., 2014. Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR–Cas platform. *Nature communications* 5, 1–9.
- Yuan, J., Wang, K., Yi, G., Ma, M., Dou, T., Sun, C., Qu, L.-J., Shen, M., Qu, L., Yang, N., 2015. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. *Genet Sel Evol* 47, 82. <https://doi.org/10.1186/s12711-015-0161-1>
- Zenger, K.R., Khatkar, M.S., Jones, D.B., Khalilisamani, N., Jerry, D.R., Raadsma, H.W., 2019. Genomic Selection in Aquaculture: Application, Limitations and Opportunities With Special Reference

to Marine Shrimp and Pearl Oysters. *Front. Genet.* 9, 693.
<https://doi.org/10.3389/fgene.2018.00693>

Titre : Développement d'outils et de méthodes de sélection génomique chez le bar et la daurade

Mots clés : Bar européen, daurade royale, aquaculture, résistance aux maladies, sélection génomique, détection de QTL

Résumé : Le bar (*Dicentrarchus labrax*) et la daurade (*Sparus aurata*) sont deux espèces majeures de l'aquaculture méditerranéenne. Comme tout élevage, l'aquaculture doit faire face à de nombreuses épidémies, que l'on peut tenter de contrôler par sélection génétique. Avec la facilité d'accès aux technologies de séquençage du génome et aux outils génomiques, la question de l'utilisation de données génomiques en sélection se pose. L'objectif de cette thèse était de développer des outils et des méthodes pour mettre en place la sélection génomique pour améliorer la résistance à la nodaviriose et à la vibriose chez le bar et à la pasteurellose chez la daurade.

Dans un premier temps, nous avons développé un outil simple et opérationnel d'assignation de parenté basé sur une méthode permettant d'assigner des individus à leurs parents à partir des probabilités mendéliennes de transmission estimées sur la population à assigner. Ensuite, l'architecture génétique des caractères a été étudiée par l'estimation des composantes de la variance et par détection de QTL. Nous avons montré que la résistance à la nodaviriose chez le bar est un caractère oligogénique,

en partie contrôlé par un QTL à effet fort et avec une héritabilité modérée. Nous avons également pu montrer que la résistance à la vibriose chez le bar et de la résistance à la pasteurellose chez la daurade sont des caractères polygéniques dont les héritabilités sont modérées. Enfin, nous avons évalué la précision de la sélection génomique avec différentes densités de marqueurs et différentes tailles de populations d'entraînement, en utilisant ou non l'information sur le QTL de résistance à la nodaviriose. Nous avons montré que la sélection génomique permet un gain de précision compris entre 8.9% et 24.5% pour les espèces et les caractères étudiés. De plus, la prise en compte de l'information du QTL de résistance à la nodaviriose chez le bar permet d'augmenter la précision de 10.5% à 26.3%.

Cette thèse a permis d'évaluer l'efficacité de la sélection génomique chez le bar et la daurade, de développer des outils facilitant l'utilisation des données génomiques dans les schémas de sélection. Nous disposons ainsi d'un cadre opérationnel pour mettre en place et optimiser la sélection génomique chez le bar et la daurade.

Title : Development of tools and methods for genomic selection in the European seabass and gilthead seabream

Keywords : European seabass, gilthead seabream, aquaculture, disease resistance, genomic selection, QTL detection

Abstract : European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*) are two major species of Mediterranean aquaculture. As any animal production, aquaculture must face many disease outbreaks, that can be controlled by selective breeding. With easy access to whole genome sequencing technologies and genomic tools, the use of genomic data in selective breeding has to be considered. The purpose of this thesis was to develop tools and methods to implement genomic selection to improve resistance to viral nervous necrosis and vibriosis in seabass and to pasteurellosis in seabream.

First, we developed a simple and efficient parentage assignment tool based on a method using Mendelian transmission probabilities, estimated from the population to assign. Then, we studied the genetic architecture of the traits by variance components estimation and QTL detection. We showed that the viral nervous necrosis resistance in seabass is an oligogenic trait controlled

by a strong effect QTL with a moderate heritability. We also showed that resistance to vibriosis in seabass and to pasteurellosis in seabream are two polygenic traits with moderate heritability. Finally, we evaluated the accuracy of genomic selection with different marker densities and different training population sizes, using or not the information on the viral nervous necrosis resistance QTL in seabass. We showed that genomic selection increased selection accuracy by 8.9% to 24.5% in the species and traits we studied. Then, we showed that accounting of the viral nervous necrosis resistance QTL information in seabass increased selection accuracy by 10.5% to 26.3%. This thesis evaluated the efficiency of the genomic selection in seabass and seabream, and develop tools making the use of genomic data in breeding schemes easier. Thus, we now have a framework to implement and optimize the genomic selection in seabass and seabream.