



**HAL**  
open science

# Précision de l'évaluation génomique chez la poule pondeuse

David Picard Druet

► **To cite this version:**

David Picard Druet. Précision de l'évaluation génomique chez la poule pondeuse. Biologie animale. Agrocampus Ouest, 2020. Français. NNT : 2020NSARB337 . tel-02916079

**HAL Id: tel-02916079**

**<https://hal.inrae.fr/tel-02916079>**

Submitted on 17 Aug 2020

**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.

# THÈSE DE DOCTORAT DE

AGROCAMPUS OUEST

ECOLE DOCTORALE N° 600

Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation

Spécialité : Biologie et physiologie animales

Par:

**David PICARD DRUET**

## Précision de l'évaluation génomique chez la poule pondeuse

Thèse présentée et soutenue à Rennes, le 22/04/2020

Unité de recherche: Pegase - UMR INRAE/AO, INRA St Gilles

Thèse N°: 2020-7\_B-337

### Rapporteurs avant soutenance :

Laurence MOREAU Directrice de Recherche, INRA UMR Génétique Quantitative et Évolution  
Zulma VITEZICA Maitresse de Conférences, GENPHYSE, INRA Toulouse

### Composition du Jury :

Examineurs : Laurence MOREAU Directrice de Recherche, INRA UMR Génétique Quantitative et Évolution  
Zulma VITEZICA Maitresse de Conférences, GENPHYSE, INRA Toulouse  
Yannick LE COZLER Professeur, Pegase – UMR INRA/AO  
Thierry BURLLOT Directeur R&D, Novogen SAS  
François GUILLAUME PHD, Evolution SAS  
Dir. de thèse : Pascale LE ROY Directrice de Recherche, Pegase – UMR INRA/AO

# Remerciements

Comme pour toutes les thèses, la mienne n'aurait pas été possible sans de nombreuses personnes qui m'ont permis, par leurs conseils, actes ou simplement présence, de boucler ces 3 années intenses et fortes en émotions ! Malgré (ou à cause de ?) la difficulté, j'ai beaucoup de bons souvenirs de ma thèse, et ça restera dans mes souvenirs une période importante de ma vie, de part la quantité astronomique de choses que j'ai pu apprendre, et des gens que j'ai pu rencontrer. N'étant pas expansif de nature, mes remerciements sont courts, mais cependant sincères.

Bien entendu, je me dois de remercier Pascale, sans toi ça n'aurait pas été possible, et je n'aurais pas pu souhaiter meilleur encadrement.

Merci à Laurence Moreau et Zulma Vitezica pour avoir accepté d'être les rapporteuses de ma thèse, ainsi que Yannick Le Cozler, François Guillaume et Thierry Burlot d'avoir accepté de participer à son évaluation.

Merci aux membres de mon comité de thèse, Anne Laperche, Sophie Brard, Llibertat Tussel-Palomero, Vincent Ducrocq et Marc Vandeputte pour les précieux conseils, qui m'ont vraiment beaucoup aidés.

Merci aux personnes de l'équipe et de l'Inrae de St Gilles pour, entre autres, l'accueil, les discussions à bâtons rompus, les mots croisés, les potins, les cafés et les gâteaux. Mes hommages également à la badgeuse, qui relèvera probablement moins d'erreurs dues à un oreiller bien trop confortable pour les horaires de bureau.

Un grand merci aux collègues doctorants (et autres précaires !), qui ont permis de dédramatiser un paquet de situations. Pas mal de craquages, pic-niques au bord de l'étang et divers fou rires idiots, ça n'aurait pas été pareil sans vous.

Merci également aux amis de l'escalade, du jeu de rôle et de plateau, de la reconstitution historique, ou de l'amicale des Anciens Combattants de l'Aubette Volante. Certains également thésards, d'autres absolument pas, mais qui ont permis d'oublier les moments de stress intense.

Et évidemment, merci à ma famille pour le soutien, même à 800 km de distance, je vous aime pareil.  
David

# Table des matières

<b>Introduction</b>	<b>5</b>
<b>1 Contexte bibliographique</b>	<b>7</b>
1.1 La production d'œuf : Aperçu du marché . . . . .	7
1.1.1 La production d'œuf à échelle mondiale . . . . .	7
1.1.2 La production d'œuf à échelle européenne . . . . .	8
1.1.3 Segmentation du marché et adaptabilité . . . . .	9
1.2 Sélectionner en filière poule pondeuse . . . . .	10
1.2.1 Organisation de la sélection . . . . .	10
1.2.2 Élevage des poules pondeuses . . . . .	11
1.2.3 Caractères de qualité d'œuf . . . . .	11
1.2.4 Évaluation génétique . . . . .	14
1.2.5 Sélection génomique . . . . .	18
1.2.6 Application de la sélection génomique . . . . .	23
1.3 Précision de l'évaluation . . . . .	24
1.3.1 Importance de la précision de l'évaluation . . . . .	24
1.3.2 Formules théoriques de la précision de l'évaluation . . . . .	24
1.3.3 Facteurs influant sur la précision de l'évaluation . . . . .	25
1.3.4 Application de la sélection génomique . . . . .	32
1.4 Conclusion du contexte bibliographique . . . . .	33
<b>2 Matériel et méthodes</b>	<b>34</b>
2.1 Animaux . . . . .	34
2.1.1 Animaux de lignée pure . . . . .	34
2.1.2 Animaux croisés . . . . .	36
2.2 Génotypes . . . . .	38
2.3 Caractères . . . . .	39
2.3.1 Mesures des caractères . . . . .	39
2.3.2 Choix des effets environnementaux du modèle et filtration des valeurs aberrantes . . . . .	40
2.4 Évaluations génétiques et génomiques . . . . .	42
2.4.1 Modèle statistique . . . . .	42
2.4.2 Populations candidates . . . . .	43
2.4.3 Précision de l'évaluation . . . . .	43
2.5 Conclusion du matériel et méthodes . . . . .	45

<b>3</b>	<b>Précision de l'évaluation génomique sur les caractères de qualité d'œuf en lignées pures</b>	<b>46</b>
3.1	Présentation de l'article 1 . . . . .	46
3.2	Conclusions de l'article 1 . . . . .	58
<b>4</b>	<b>Apport des informations des descendantes croisées sur la précision de l'évaluation</b>	<b>59</b>
4.1	Présentation de l'article 2 . . . . .	59
4.2	Conclusions de l'article 2 . . . . .	84
<b>5</b>	<b>Genome Wide Association Studies : étude des différences d'architecture génétique des caractères de qualité d'œuf, et effets de l'utilisation conjointe des informations de descendants pures et croisées sur la détection des QTLs.</b>	<b>85</b>
5.1	Introduction . . . . .	85
5.2	Méthodes . . . . .	86
5.2.1	Animaux . . . . .	86
5.2.2	Génotypes . . . . .	87
5.2.3	Caractères . . . . .	88
5.2.4	Réalisation des GWAS . . . . .	89
5.2.5	Estimation des pourcentages de variance expliquée par les SNPs	89
5.3	Résultats . . . . .	90
5.3.1	QTLs significatifs sur les caractères de lignée pure . . . . .	90
5.3.2	QTLs significatifs sur les caractères des descendantes croisées .	93
5.3.3	Différences entre caractère en lignée pure et sur descendantes croisées . . . . .	97
5.4	Discussion . . . . .	99
5.4.1	Résultats obtenus par rapport à la littérature . . . . .	99
5.4.2	Différences d'architecture entre caractères . . . . .	100
5.4.3	Utilisation de la GWAS mixte . . . . .	101
5.5	Conclusion . . . . .	102
<b>6</b>	<b>Discussion générale et perspectives</b>	<b>103</b>
	<b>Annexes</b>	<b>107</b>
6.1	GWAS pure vs GWAS mixte . . . . .	107
6.2	GWAS croisée vs GWAS mixte . . . . .	112
	<b>Bibliographie</b>	<b>116</b>
	<b>Table des figures</b>	<b>128</b>
	<b>Liste des tableaux</b>	<b>132</b>

# Introduction

La sélection est particulièrement importante pour le domaine de la production d'œufs, qui est un marché en expansion au niveau mondial. Celui-ci présente une forte segmentation du marché, c'est à dire que chaque pays aura un cahier des charges particulier, dépendant des attentes des consommateurs. Cela signifie que les entreprises de sélection ont à produire des individus adaptés, ou adaptable, aux conditions d'élevage diverses, ainsi que capables de répondre aux différents besoins des consommateurs. En plus de cela, la sélection des poules pondeuses est un secteur particulièrement compétitif, étant détenue au niveau mondial par 3 grands groupes, Lohman, Hendrix Genetics ainsi que Novogen. Cela fait de la sélection des différentes lignées, qui répondent aux différents critères de production, une étape particulièrement critique. Le progrès génétique des populations sélectionnées est fortement diffusé partout dans le monde, et l'optimiser est un moyen de se distinguer des concurrents.

Chez les poules pondeuses, la sélection des individus se fait en suivant un schéma de sélection pyramidal. Les individus, mâles et femelles, sont sélectionnés en lignées pures, et ces lignées sont ensuite croisées entre elles pour obtenir les individus qui seront ensuite vendus aux marchés cibles. Ce type de schéma de sélection a plusieurs avantages, permettant entre autres de garder en interne les lignées pures. L'évaluation des reproducteurs en elle-même se repose sur l'évaluation génétique pour estimer leur valeur génétique (Estimated Breeding Value, EBV). Celle-ci n'étant pas directement mesurable, elle doit être estimée statistiquement à partir des performances et de la généalogie. Cette estimation se fait classiquement avec BLUP suivant un modèle animal (Henderson, 1973).

Avec l'évaluation, on estime donc une valeur génétique pour chaque individu, et sont choisis comme reproducteurs les candidats dont les EBVs remplissent au mieux l'objectif de sélection. Cet objectif est déterminé par la liste des caractères d'intérêt, chacun pondéré en fonction de son importance économique. Les individus dont la valeur pondérée est la meilleure sont ceux qui seront sélectionnés. L'objectif de cette sélection étant d'obtenir une descendance dont les performances seront en espérance meilleures que celles de leurs parents.

La sélection des poules pondeuses se fait en général sur l'ensemble des données phénotypiques disponibles, sur ancestraux et les individus apparentés de même génération que les candidats à la sélection (collatéraux). Quand les collatéraux ont terminé leurs carrières, leurs performances sont utilisées conjointement avec celles

des ancêtres pour estimer les valeurs génétiques des candidats. Ce type d'évaluation signifie qu'il faut attendre la fin de carrière des individus collatéraux pour évaluer les candidats, ce qui donne un intervalle de génération assez important (18 mois). Élever tous les poussins mâles aurait un coût trop élevé, seul une fraction des poussins, généralement 1 par mère, est conservée, ce qui diminue l'intensité de sélection.

Depuis plusieurs années, la méthodologie de la sélection génomique, théorisée par Meuwissen et al (2001), est mise en place dans différentes filières. Elle a notamment permis d'accélérer le progrès génétique dans la filière bovine, qui a été pionnière dans son utilisation. Cette méthode d'évaluation a été jugée prometteuse pour la filière avicole (Wolc et al., 2016b). Pour estimer précisément l'effet de l'usage de cette méthode d'évaluation sur le progrès génétique des populations sélectionnées, il est nécessaire de connaître la précision de l'évaluation génomique.

L'objectif de ma thèse était d'estimer la précision de l'évaluation génomique sur les caractères de qualité d'œuf chez la poule pondeuse. Le premier chapitre présente en détails la filière de la production d'œufs, le principe de la sélection génomique, ainsi que les différents paramètres influençant la précision de cette méthode d'évaluation. Les différents jeux de données utilisées et les méthodes utilisées pour les évaluations sont détaillées dans le second chapitre. Le chapitre 3 présente les résultats obtenus dans l'estimation de la précision de l'évaluation génomique pour la sélection des reproducteurs mâles et femelles en lignée pure. Les travaux suivants ont porté sur la précision de l'évaluation en cas d'évaluation utilisant conjointement les performances des individus de lignée pure et celles de leur descendance croisée, et sont présentés dans le chapitre 4. Les travaux sur les architectures génétiques des caractères de qualité d'œufs sont ensuite traitées dans le 5<sup>ème</sup> chapitre. Le dernier chapitre reprend l'ensemble des résultats obtenus lors de la thèse, y ajoute des compléments d'informations et présente les perspectives faisant suite à la thèse.

La thèse était co-financée par la région Bretagne, et par le méta-programme INRA SelGen, à travers le projet OptiSeq, et s'appuie sur des données collectées par la société de sélection avicole Novogen, qui a été partenaire tout au long du projet. La société Novogen, filiale du groupe Grimaud, est implantée partout dans le monde, et fait partie des leaders dans le domaine de la sélection des poules pondeuses.

# Chapitre 1

## Contexte bibliographique

### 1.1 La production d'œuf : Aperçu du marché

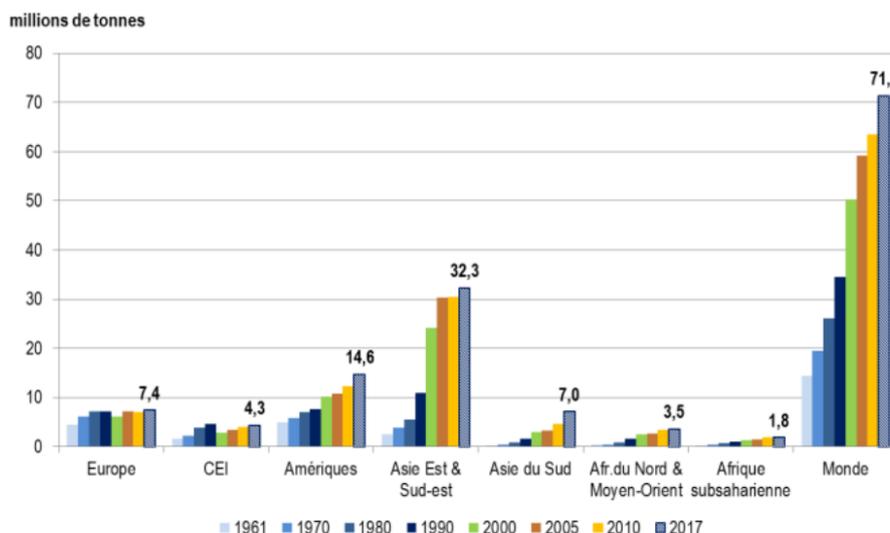
Au niveau de la production d'œufs, il existe deux types de produits différents à considérer. Les œufs coquille, aussi appelés œufs entiers et les ovoproduits, qui sont fabriqués à partir du jaune et/ou du blanc des œufs. Ces ovoproduits peuvent être liquides (e.g. blanc d'œuf en bouteille), ou secs (e.g. œuf en poudre). En matière d'objectifs de sélection, les caractères à améliorer peuvent varier entre ces deux types de produits. La couleur de la coquille, par exemple, est essentielle pour les œufs entiers, alors qu'elle n'est pas pertinente pour les ovoproduits. *A contrario*, la résistance de la coquille à la fracture, la taille et la forme de l'œuf peuvent être des caractères d'intérêt pour l'industrie des ovoproduits, car impliqués dans les processus d'automatisation, e.g. en casserie.

#### 1.1.1 La production d'œuf à échelle mondiale

La production d'œuf est un marché en expansion au niveau mondial (cf figure 1.1, source FAOSTAT). Le premier pays producteur au monde (en 2017) est la Chine, avec une production de 24 millions de tonnes (Mt), soit 34% de la production mondiale totale (70.7 Mt). Suivent les 28 pays de l'Union Européenne, avec 7.1 Mt (10%), les États-Unis avec 5.4 Mt (8%), l'Inde avec 4.7 Mt (7%), le Mexique avec 2.8 Mt (4%), la Russie et le Japon avec 2.6 Mt chacun (4%) et le Brésil, avec 2.4 Mt (3%) (Source : ITAVI, Avril 2018).

Les échanges mondiaux (hors intra-UE) concernent à 73% (à la fois en volume et valeurs) les œufs coquille. Les exportations des différents pays sont très hétérogènes. Certains, telle la Chine, exportent assez peu malgré une forte production, qui est donc destinée au marché intérieur. D'autres, comme la Turquie, exportent massivement un type de produit (en l'occurrence, des œufs coquille), et pas l'autre.

Concernant les œufs coquille, en volume, les 3 plus grands exportateurs mondiaux sont la Turquie (35%), les États-Unis (10%), et l'Ukraine (9%), concentrant à eux trois 54% du commerce mondial (sans prendre en compte le commerce intra-UE). En ce qui concerne les ovoproduits, la palme revient à l'UE (38%), les États-Unis



Source : FAOSTAT

FIGURE 1.1 – Production d’œufs entre 1961 et 2017 dans le monde (millions de tonnes)

(28%) et l’Inde (10%), qui concentrent donc 76% du volume du commerce mondial.

### 1.1.2 La production d’œuf à échelle européenne

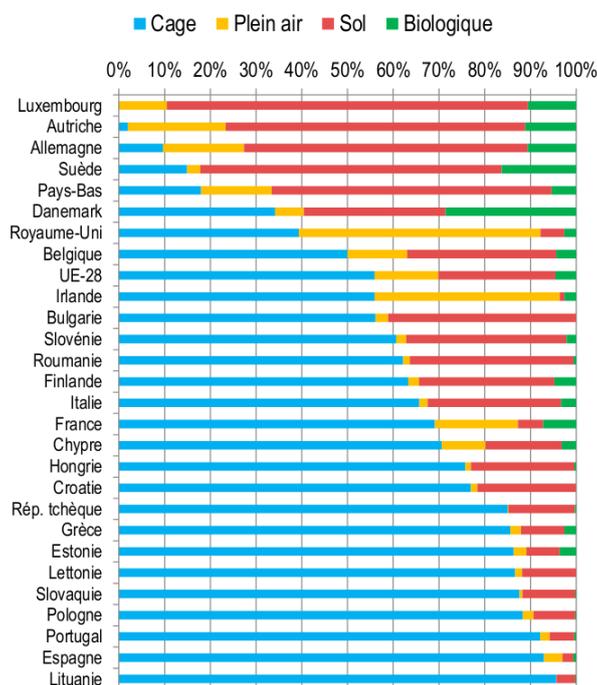
Au sein de l’Union Européenne, la production d’œufs en 2017 a été de 7.09 Mt, stable par rapport à 2016, et en nette croissance depuis les mises aux normes effectuées en 2012 (+12.1%). Ces mises aux normes concernaient l’augmentation de la taille des cages (avec une surface minimale de 750cm<sup>2</sup> par poule), ainsi que l’installation dans les cages de nids, perchoirs et zones de grattage (directive Européenne 1999/74/CE).

Sur la décennie écoulée, l’évolution annuelle moyenne est stable, aux alentours de +0.5%. En 2017, les 4 pays qui produisent le plus d’œufs dans l’UE sont la France (925 kt), suivie par l’Espagne (838 kt), l’Allemagne (820 kt) et l’Italie (800 kt) (Source : ITAVI, Avril 2018).

Le commerce intra-UE est extrêmement dynamique : il représente en effet 54% des volumes, et 59% en valeurs, du commerce mondial. Les ovoproduits représentent 39% du volume des échanges, une proportion plus importante que pour le commerce mondial hors-UE (27%).

Ces dernières années, les consommateurs se sont de plus en plus intéressés aux questions concernant le bien-être animal, ainsi qu’à l’impact de l’élevage sur l’environnement. Cela a eu un effet important sur la production d’œuf, qui a dû se diversifier pour s’adapter aux différentes demandes des consommateurs. Cela s’est traduit, non-seulement par des changements au niveau des cages, tels qu’évoqué précédemment, mais également par une augmentation importante de la part des systèmes alternatifs à la cage. Ces systèmes sont passés de 8% à 44% des effectifs en 20 ans, et la tendance

devrait se poursuivre dans les années futures (Source : ITAVI, avril 2018, cf figure 1.2).



Source : CIRCABC

FIGURE 1.2 – Répartition des différents systèmes d'élevage dans l'UE en 2017 (ITAVI, 2018)

### 1.1.3 Segmentation du marché et adaptabilité

Depuis la domestication des premières poules (aux environs de -6 000 AC), les individus ont été sélectionnés afin de répondre à différents critères de production, produisant différentes lignées avec des phénotypes très différents. Par exemple, les lignées Rhode Island, plutôt optimisées pour la ponte pèsent en moyenne entre 1.7 et 2.2 kg de poids corporel. Une lignée sélectionnée pour la consommation de la chair, telle que la lignée Faverolles, aura des femelles qui pèsent davantage, entre 2.8 et 3.5 kg de poids corporel (Source : Société Centrale d'Aviculture de France).

La sélection est particulièrement importante pour le domaine de la production d'œufs, celui-ci présentant une forte segmentation du marché. C'est à dire que chaque pays aura un cahier des charges particulier (poids de l'œuf, couleur de la coquille, etc...), qui va dépendre des attentes des consommateurs. Par exemple, si en Europe, on consomme majoritairement des œufs bruns, les États-Unis vont plutôt privilégier la consommation d'œufs à coquille blanche. Les individus adaptés aux demandes d'un pays ne le seront pas forcément pour le pays voisin. De plus, les poules pondeuses vont être élevées dans différents systèmes, en conditions variées.

Cela signifie que pour la production d'œufs, les entreprises de sélection doivent fournir un ensemble d'individus adaptés, ou adaptables, aux différentes conditions d'élevage, et capables de répondre aux différentes attentes des consommateurs. Ces individus sont obtenus par le biais de la sélection des reproducteurs de différentes lignées. Chacune de ces lignées pures est sélectionnée pour avoir des caractéristiques précises (taille de l'œuf, couleur de la coquille, etc...). Les différentes lignées sont ensuite croisées entre elles pour obtenir des descendants qui répondent aux différentes demandes des consommateurs ou des transformateurs. Actuellement, la sélection des poules pondeuses est détenue au niveau mondial par 3 grands groupes, avec chacun une clientèle internationale : Lohman, Hendrix Genetics, ainsi que Novogen. Cela en fait un secteur fortement concurrentiel.

## 1.2 Sélectionner en filière poule pondeuse

### 1.2.1 Organisation de la sélection

Classiquement, la sélection avicole se fait via un schéma de sélection pyramidal (cf figure 1.3). La sélection se fait sur des lignées pures, mâles et femelles. Différents croisements, généralement à 4 voies (4 lignées pures différentes), sont ensuite effectués lors de l'étape de multiplication, et les individus obtenus sont vendus aux marchés cibles. L'entreprise pourra décider de vendre les grand-parentaux, les parentaux, ou les croisés finaux, en fonction de ses choix stratégiques.

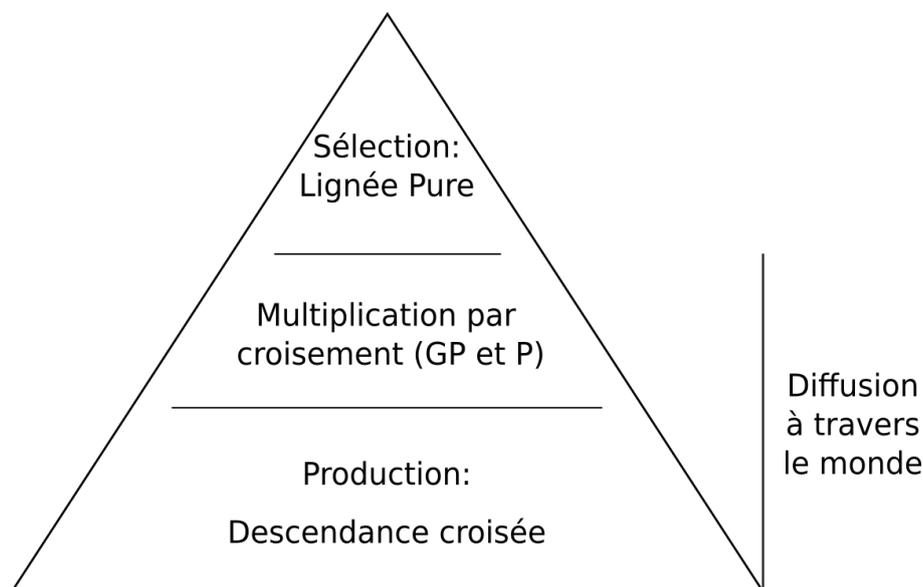


FIGURE 1.3 – Schéma de sélection pyramidal : les lignées pures sont conservées en noyau de sélection, et Grands-parentaux (GP), Parentaux (P) et produits finaux sont diffusés à travers le monde.

L'utilisation d'un schéma de sélection pyramidal a plusieurs avantages. Tout d'abord, étant donné le contexte concurrentiel de la sélection avicole, il permet

de conserver en interne les lignées pures composant le noyau de sélection, et donc sa génétique, en ne commercialisant que des individus croisés. Ensuite, ces lignées pures sont complémentaires : chacune est sélectionnée sur des critères bien précis, et le croisement de ces lignées permet d'obtenir les descendants croisés les plus adaptés aux demandes. De plus, ces descendants profitent des effets d'hétérosis induit par les croisements. C'est à dire que leurs performances sont plus élevées que les performances moyennes des lignées pures parentales du fait des effets de dominances entre les différents allèles des lignées pures. Enfin, du fait de l'étape de multiplication, il est possible d'obtenir beaucoup de descendants à partir d'un seul coq : un individu en lignée pure peut potentiellement avoir 250 000 descendantes croisées, ce qui représente la production de 80 millions d'œufs. Une amélioration au niveau de la sélection aura des répercussions importantes au niveau de la production.

### 1.2.2 Élevage des poules pondeuses

L'organisation de l'élevage est centrée autour de la production d'œufs des poules pondeuses. Cette production est définie par l'intensité de ponte, qui est exprimé en %. Elle est obtenue en divisant le nombre d'œufs pondus sur une période de temps donnée par le nombre de jour de la période de temps, et en multipliant ce ratio par 100. La figure 1.4 présente l'évolution du pourcentage de ponte, ainsi que de la viabilité, le poids d'œuf, le poids corporel et la consommation journalière en fonction de l'âge en semaine pour la souche ISA Brown.

La courbe de ponte peut être découpée en trois phases. La première est une phase ascendante, qui va de l'âge de ponte du premier œuf (entre 16 et 18 semaines) jusqu'à un pic de ponte. Cette phase-ci a lieu entre 18 semaines, âge auquel les poulettes sont intégrées à l'élevage, et 30 semaines d'âge.

Le pic de ponte correspond à la maturité sexuelle des poulettes, qui a lieu entre 20 et 25 semaines, et où l'intensité de ponte est maximale.

La troisième phase est une phase descendante, qui débute juste après le pic de ponte, et dure jusqu'à la réforme de la poule. Elle correspond à une décroissance quasi-linéaire de l'intensité de ponte. En général, les poules sont réformées au-delà de 80 semaines.

### 1.2.3 Caractères de qualité d'œuf

Dans ce manuscrit, les différents caractères seront désignés en accord avec les ontologies des caractères pour l'animal d'élevage (Animal Trait Ontology for Livestock <http://www.atol-ontology.com/>)

#### Définition des caractères étudiés

Poids de l'œuf (Egg Weight, EW) et forme de la coquille (Egg Shell shape index, ESshape) sont des caractères à la fois directement impliqués dans différents processus de commercialisation, et font également partie des attentes des consommateurs. La couleur de la coquille (Egg Shell Color, ESC) est un caractère qui est essentiel

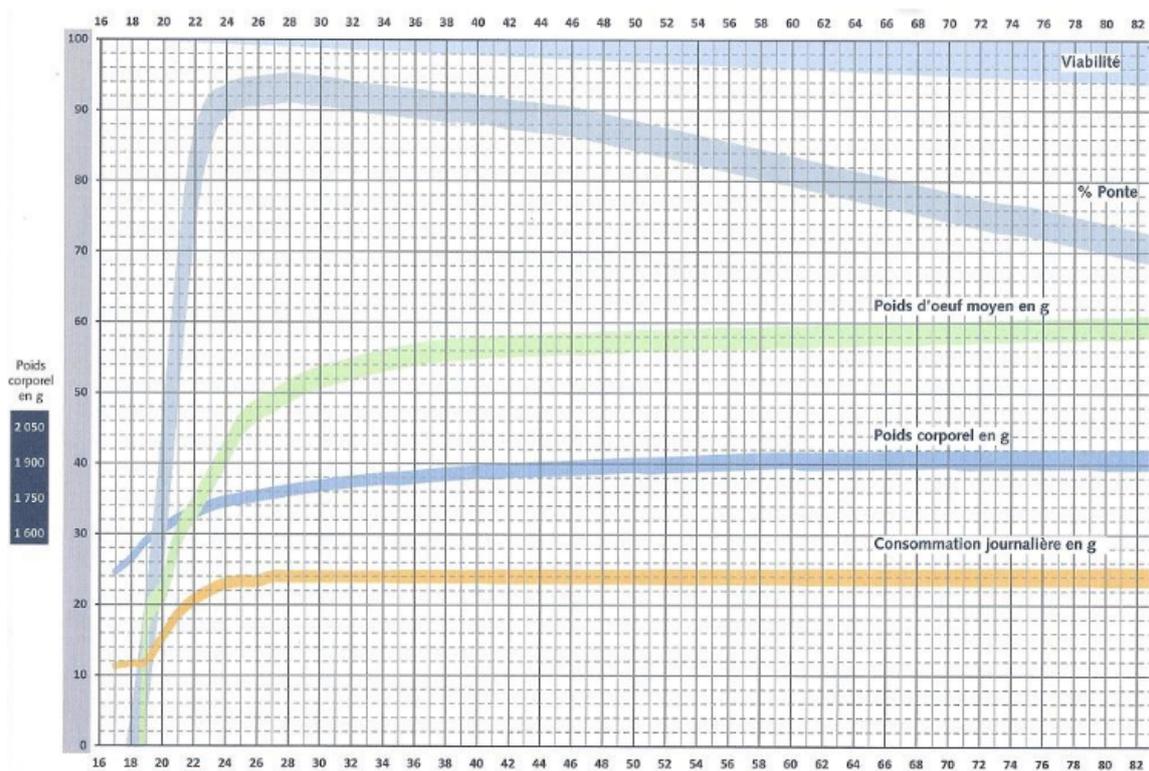


FIGURE 1.4 – Courbes de référence concernant la viabilité, le pourcentage de ponte, le poids d'œuf, le poids corporel et la consommation journalière en fonction de l'âge en semaine pour la souche ISA Brown (Source : Hendrix-ISA)

pour le consommateur, car il touche à ses habitudes de consommation, ce critère de couleur de coquille étant très régionalisé. Certains pays consommeront surtout des œufs bruns, d'autres des œufs blancs. Dans tous les cas, la coquille de l'œuf doit présenter une couleur homogène. La solidité de la coquille (Egg Shell Strength, ESS) est un caractère qui est quant à lui très important pour la commercialisation des œufs. La coquille étant la barrière protégeant le contenu de l'ensemble des pathogènes extérieurs, elle se doit d'être suffisamment résistante pour jouer son rôle protecteur jusqu'à la consommation. La hauteur de l'albumen (Albumen Height, AH) est quant à elle un marqueur de la qualité interne de l'œuf, et notamment de sa vitesse de dégradation, l'albumen composant une part importante du contenu de l'œuf.

Tous ces caractères présentent des variations au cours de la vie d'une poule. Notamment, la coquille devient plus claire, et l'œuf prend une forme plus pointue avec l'augmentation de l'âge de la poule (Ledvinka et al., 2012; Sirri et al., 2018). On observe également une augmentation du poids de l'œuf avec l'âge (cf figure 1.4).

Nous verrons chapitre 2 comment sont effectuées les mesures de ces caractères utilisés dans la présente étude.

## Paramètres génétiques des caractères de qualité d'œuf

De tous les caractères de qualité d'œuf, EW et ESC sont ceux qui présentent l'héritabilité la plus forte, suivi par ESshape et AH, et ESS présente l'héritabilité la plus faible (cf table 1.1).

Caractère	Population	$h^2$	Ref.
EW	Pondeuses œufs bruns naine	0.63	Zhang et al. (2005)
	Rhode Island Red	0.60	Dunn et al. (2005)
	Pondeuses œufs bruns	0.60	Beaumont et al. (2010)
	Pondeuses œufs bruns	0.65	Wolc et al. (2011)
	Pondeuses œufs bruns	0.73	Wolc et al. (2015)
ESC	Pondeuses œufs bruns naine	0.46	Zhang et al. (2005)
	Pondeuses œufs bruns	0.68	Wolc et al. (2011)
	Pondeuses œufs bruns	0.75	Wolc et al. (2015)
ESS	Pondeuses œufs bruns naine	0.24	Zhang et al. (2005)
	Pondeuses œufs bruns	0.32	Beaumont et al. (2010)
	Rhode Island Red	0.24	Kibala et al. (2018)
	Rhode Island White	0.16	Kibala et al. (2018)
AH	Pondeuses œufs bruns naine	0.51	Zhang et al. (2005)
	Pondeuses œufs bruns	0.52	Wolc et al. (2011)
	Pondeuses œufs bruns	0.55	Wolc et al. (2015)
ESshape	Rhode Island Red	0.30	Dunn et al. (2005)
	Pondeuses œufs bruns naine	0.40	Zhang et al. (2005)

TABLE 1.1 – Héritabilité des différents caractères de qualité d'œuf étudiés, dans différentes populations

En termes de corrélations génétiques, il y a relativement peu d'informations disponibles dans la littérature sur les caractères de qualité d'œuf. De ce que l'on peut trouver, EW est un caractère fortement corrélé au diamètre de l'œuf, ainsi que modérément avec AH (0.30) (Zhang et al., 2005; Wolc et al., 2015). ESC n'est corrélé avec aucun autre caractère de qualité d'œuf. ESS est un caractère modérément corrélé avec ESshape (0.20) (Zhang et al., 2005; Beaumont et al., 2010).

## Architecture génétique

À ce jour, 312 études ont été menées, et ont permis de répertorier dans la base de donnée Chicken QTL database un total de 2 486 QTLs associés aux 41 différents caractères de qualité d'œuf chez la poule (Source : Chicken QTL database, version 38, publiée le 24 Avril 2019, Hu et al. (2019)). L'architecture génétique des caractères étudiés a été résumé dans le tableau 1.2.

Caractère	Chromosome	Position	Publications	
EW	1	25Mb	Tuiskula-Haavisto et al. (2002); Kerje et al. (2003); Sasaki et al. (2004); Honkatukia et al. (2005); Schreiweis et al. (2006); Wright et al. (2008); Wolc et al. (2012); Goraga et al. (2012); Wolc et al. (2014); Romé et al. (2015)	
		50Mb		
	2	30Mb		
	4	50Mb		
	Z	32Mb		
ESC	3	80Mb	Wardecka et al. (2002); Schreiweis et al. (2006), Wolc et al. (2014); Romé et al. (2015), Tuiskula-Haavisto et al. (2018); Liu et al. (2018)	
	4	74Mb		
	12	3Mb		
ESS	1	80Mb	Tuiskula-Haavisto et al. (2002); Sasaki et al. (2004); Dunn et al. (2009); Tuiskula-Haavisto et al. (2011); Wolc et al. (2014); Sun et al. (2015); Tuiskula-Haavisto et al. (2018)	
	2	102Mb		
	6	20Mb		
	4	50Mb		
	Z	10Mb		
AH	1	45Mb	Wardecka et al. (2002); Hansen et al. (2005); Liu et al. (2011a); Goto et al. (2014); Wolc et al. (2014)	
		125Mb		
	2	85Mb		
	3	100Mb		
	4	73Mb		
ESshape	2	108Mb	Hansen et al. (2005); Dunn et al. (2009); Yao et al. (2010); Tuiskula-Haavisto et al. (2018)	
		4		50Mb
		8		20Mb

TABLE 1.2 – Régions génomiques affectant les caractères de qualité d’œuf, d’après Romé and Le Roy (2016). La position correspond à la localisation moyenne du pic des QTL.

## 1.2.4 Évaluation génétique

Pour améliorer les lignées aviaires, la sélection des reproducteurs a lieu via l’évaluation génétique. La valeur génétique des individus n’étant pas directement mesurable, elle doit être estimée statistiquement à partir des phénotypes et de la généalogie.

### Principes généraux

Biologiquement, le phénotype d’un individu est le résultat de la combinaison de 2 facteurs : le génotype de l’individu, c’est à dire sa composition allélique sur l’ensemble des gènes de son génome, et son environnement. Cet environnement inclut à la fois les conditions dans lesquelles vit l’individu, et son état physiologique (tel que l’âge).

D’un point de vue mathématique, l’évaluation génétique repose sur un modèle additif, qui considère que ces 2 types d’effets s’additionnent. Par conséquent, le

phénotype  $P$  est décomposé en deux parts : la part génétique  $G$ , dont une part est transmissible aux descendants, et une part due à l'environnement  $E$ . Cela donne un modèle tel que :

$$P = G + E.$$

Où  $P$ ,  $G$ ,  $E$  sont des variables aléatoires suivant des lois normale, de variance  $\sigma_p^2$ ,  $\sigma_g^2$ ,  $\sigma_e^2$  respectivement.

La valeur génétique d'un individu est le résultat de l'expression des allèles qu'il porte, des différents effets de dominance/récessivité inter-alléliques et de l'interaction des gènes entre eux (épistasie). Chaque individu transmettant à ses descendants la moitié de ses allèles, et donc, en espérance, la moitié de sa valeur génétique additive, la valeur génétique d'un individu est l'espérance de celle de ses deux parents, à laquelle s'ajoute un aléa de méiose. Cet aléa de méiose signifie que 2 pleins-frères ou pleines sœurs n'auront pas la même valeur génétique en ayant les mêmes parents. Les autres effets (dominance et épistasie) sont cassés par la méiose, et ne sont pas transmis aux descendants.

La part génétique  $G$  se décompose donc elle-même en 3 groupes d'effets : les effets additifs  $A$ , les effets de dominance  $D$  et les effets d'épistasie  $I$ , ce qui donne :

$$G = A + D + I$$

Où  $A$ ,  $D$ ,  $I$  sont des variables aléatoires suivant des lois normale, de variance  $\sigma_a^2$ ,  $\sigma_d^2$ ,  $\sigma_i^2$  respectivement.

Sachant que la partie  $A$  est transmissible à la descendance, l'héritabilité d'un caractère, c'est à dire la part du phénotype qui est d'origine génétique et additive, est calculée en estimant la part de variance expliquée par  $A$  par rapport à la variance phénotypique totale  $P$ .

$$\text{Soit : } h^2 = \frac{\sigma_A^2}{\sigma_P^2}.$$

Un caractère héritable est améliorable par la sélection. Si l'héritabilité est élevée, cela signifie que l'expression d'un caractère est fortement influencée par la part qui est transmissible aux descendants. Donc, plus l'héritabilité est élevée, plus la sélection sera efficace, car on évalue mieux  $A$  à partir de  $P$ .

Pour l'amélioration génétique des espèces d'élevage, on cherche donc à déterminer les individus avec la meilleure valeur génétique, en étant le plus précis possible.

### **Prédiction de la valeur des reproducteurs : le BLUP**

Pour prédire la valeur génétique additive d'un individu, à partir de ses performances et de celles des animaux apparentés, la méthode statistique de référence est le BLUP (Best Linear Unbiased Predictor, Henderson (1973)). À l'heure actuelle, le modèle le plus communément mis en œuvre est un modèle animal défini comme suit :

$$\mathbf{y} = \mathbf{1}\mu + X\mathbf{b} + W\mathbf{a} + \mathbf{e}$$

Avec  $\mathbf{y}$  le vecteur des observations  $y_i$  ( $i$  étant le nombre de mesures),  $\mathbf{1}$  un vecteur de 1,  $\mu$  une moyenne,  $\mathbf{b}$  est le vecteur contenant les effets fixes. Le vecteur  $\mathbf{a}$  est celui des valeurs génétiques des animaux (effets aléatoires) tel que sa variance :  $V(\mathbf{a}) = A\sigma_A^2$ , avec  $A$  la matrice d'apparentement entre les individus, et  $\sigma_A^2$  la variance génétique additive.  $X$  est la matrice d'incidence reliant les effets fixes aux performances,  $W$  est celle reliant les performances aux animaux, et  $\mathbf{e}$  est le terme résiduel. L'estimation des valeurs génétiques se fait en minimisant la variance résiduelle, pour obtenir le BLUP. Cela se fait par le biais de la résolution d'un système d'équations appelé modèle mixte (Henderson et al., 1959).

Cette méthode d'évaluation permet de prendre en compte l'ensemble des performances mesurées chez tous les animaux apparentés, via l'utilisation de la matrice  $A$ . La valeur génétique de tous les animaux est prédite, quelle que soit la quantité d'information disponible sur chaque animal.

### Utilisation du BLUP en sélection

Avec l'évaluation, on estime donc une valeur génétique pour chaque individu, et sont choisis comme reproducteurs les candidats dont les estimées des valeurs génétiques (ou Estimated Breeding Values, EBV) remplissent au mieux l'objectif de sélection. L'objectif de sélection est déterminé par la liste des caractères sur lesquels une augmentation, réduction ou maintien de performance est souhaité. Chaque caractère est pondéré en fonction de son importance économique, et ce sont les individus pour lesquels la valeur pondérée est la meilleure qui sont sélectionnés. Cette sélection a pour but d'améliorer la valeur génétique moyenne de la population, en lui faisant faire un progrès génétique (cf figure 1.5).

Ce progrès génétique ( $\Delta G$ ), s'exprime suivant l'équation :

$$\Delta G = \frac{i * acc * \sigma_a}{T}$$

où  $i$  est l'intensité de sélection, fonction du taux de sélection (part des individus retenus parmi les candidats) (Falconer D.S., 1996) ;  $acc$  est la précision de l'évaluation, qui correspond à la corrélation entre l'EBV et la valeur génétique vraie de l'individu (True Breeding Value, TBV) ;  $\sigma_A$  la variabilité génétique additive du caractère, et  $T$  l'intervalle de génération.

La sélection des poules pondeuses se fait en général sur l'ensemble des données phénotypiques, via un BLUP utilisant un modèle animal, sur collatéraux et ancestraux. Quand les poules pondeuses apparentées de la même génération (sœurs, demi-sœurs, cousines, etc) que les candidats à la sélection ont terminé leurs carrières, leurs performances, ainsi que celles des ancêtres, parentaux et grand-parentaux, sont utilisées conjointement pour estimer les EBVs candidats. Pour les candidates femelles à la sélection, les performances de leurs apparentées de même génération et de leurs ancêtres sont également utilisées, mais à l'inverse des mâles, elles ont des performances propres qui sont également prises en compte pour

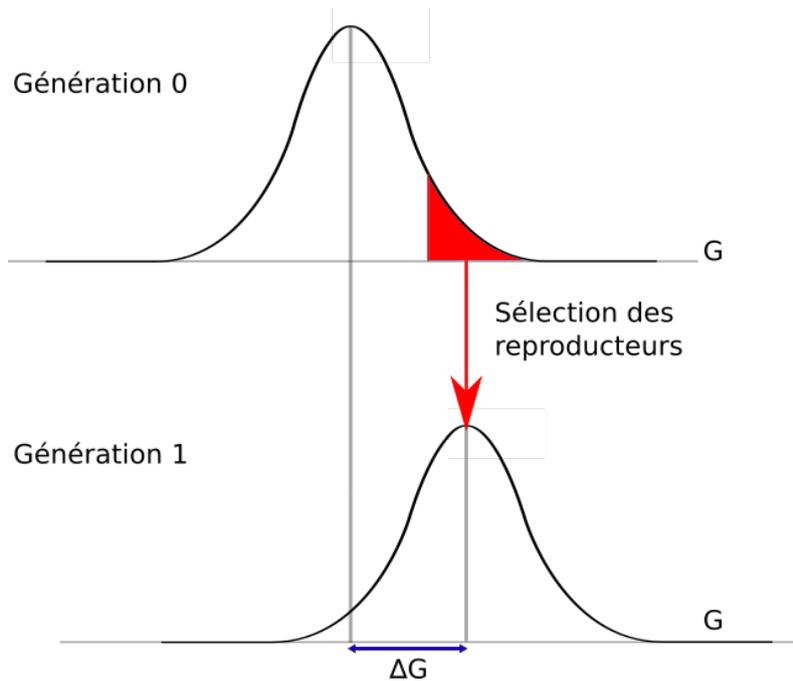


FIGURE 1.5 – Illustration du progrès génétique  $\Delta G$  par la sélection des reproducteurs de la génération 0 à la génération 1.

l'estimation de leurs EBVs.

L'évaluation sur collatéraux implique deux choses : tout d'abord, il faut attendre la fin de carrière des individus collatéraux pour pouvoir évaluer les candidats (aux environs de 70 semaines d'âge), ce qui donne un intervalle de génération d'environ 18 mois. Ce délai induit un second effet : conserver tous les descendants mâles pendant 70 semaine avant d'estimer leurs EBVs aurait un coût trop élevé. Seule une fraction des poussins mâles est donc conservée pour la sélection, en général 1 pris au hasard par mère, le reste étant abattu. Le fait de sélectionner un unique individu par mère est lié au fait qu'en absence d'information sur leur descendance, les pleins frères sont considérés comme ayant tous la même EBV. Cette limitation dans le nombre de candidats conservés, de fait, diminue l'intensité de la sélection.

Depuis quelques années, une nouvelle méthodologie, l'évaluation génomique, théorisée par Meuwissen et al. (2001) et rendue possible par les avancées technologiques, est mise en place dans différentes filières. Cela a notamment été le cas dans la filière bovine, où la structure de la filière était déjà favorable à l'utilisation de cette méthode, et où elle a permis d'accélérer le progrès génétique des populations. Cette méthode d'évaluation peut être un moyen d'optimiser davantage la sélection des reproducteurs, et donc l'amélioration des lignées de poules pondeuses, en jouant sur les paramètres du progrès génétique, tels que décrits dans le paragraphe 1.2.4.

## 1.2.5 Sélection génomique

### Principes généraux de la sélection génomique

Un Quantitative Traits Loci (QTL) est une région du génome dont le polymorphisme cause une partie de la variabilité du caractère quantitatif étudié. La localisation des QTLs permet d'avoir une information sur les régions du génome qui ont une influence sur un caractère, et qui potentiellement contiennent des gènes ayant un effet sur le caractère étudié. La sélection génomique utilise des Single Nucleotide Polymorphisms (SNPs) répartis le long du génome pour capturer l'effet des QTLs. Un SNP peut informer de 2 façons sur les QTLs. Soit en étant une mutation ayant un effet direct sur le caractère, soit en étant en Déséquilibre de Liaison (DL, plus de détails en 1.2.5) avec un polymorphisme ayant un effet sur le caractère.

En sélection génomique, ces informations données un très grand nombre de SNPs sont utilisées pour établir une relation statistique (décrite par une équation de prédiction) entre génotypes obtenus avec les marqueurs et les phénotypes, à partir d'un nombre conséquent d'individus génotypés et phénotypés (cf sous-section 1.3.3 pour plus de détails). Ces individus sur lesquels se base l'équation de prédiction constituent la population de référence.

Une fois l'équation de prédiction établie, il est possible de faire une estimation de la valeur génomique (Genomic Estimated Breeding Value, GEBV) pour de jeunes animaux sur leur génotype. Ces animaux candidats à la sélection composent la population candidate. Cela peut permettre une réduction de l'intervalle de génération, les animaux pouvant dès lors être évalués dès leur naissance (à partir du moment où on peut obtenir de l'ADN), sans attendre qu'ils aient des phénotypes propres.

### Le déséquilibre de liaison (DL)

Prenons deux loci, A et B. A a deux allèles A1 et A2, de fréquences  $P_{A1}$  et  $P_{A2}$ , B les allèles B1 et B2, de fréquences  $P_{B1}$  et  $P_{B2}$ . Quatre haplotypes (combinaison d'allèles) sont donc possibles : A1B1, A1B2, A2B1, A2B2.

En absence de déséquilibre de liaison, on parle d'équilibre d'association gamétique, et la fréquence de chaque haplotype sera le produit de la fréquence des allèles correspondant (cf table 1.3).

	A1	A2	Total
B1	$P_{A1} * P_{B1}$	$P_{A2} * P_{B1}$	$P_{B1}$
B2	$P_{A1} * P_{B2}$	$P_{A2} * P_{B2}$	$P_{B2}$
Total	$P_{A1}$	$P_{A2}$	1

TABLE 1.3 – Fréquence des haplotypes pour 2 loci bi-alléliques en situation d'équilibre d'association gamétique

En cas de déséquilibre de liaison, on observe une déviation  $D$  (cf table 1.4) par rapport à la fréquence attendue de l'haplotype, qui est signe d'une association

préférentielle entre allèles de loci différents. Les zones du génome où le DL est important sont appelées blocs haplotypiques.

Une des mesures de référence du DL est le coefficient de corrélation entre allèles  $r^2$ , qui est défini pour des loci bi-alléliques par la formule  $r^2 = \frac{D^2}{P_{A1} * P_{A2} * P_{B1} * P_{B2}}$ .

	A1	A2	Total
B1	$P_{A1} * P_{B1} + D$	$P_{A2} * P_{B1} - D$	$P_{B1}$
B2	$P_{A1} * P_{B2} - D$	$P_{A2} * P_{B2} + D$	$P_{B2}$
Total	$P_{A1}$	$P_{A2}$	1

TABLE 1.4 – Fréquence des haplotypes pour 2 loci bi-alléliques en situation de déséquilibre de liaison

Le DL est un phénomène lié au support de l'information génétique et à sa méthode de transmission à la descendance. La séquence de nucléotides de l'ADN est un lien physique entre les loci, ceux-ci ne sont pas indépendants. Cependant, au cours de la méiose se produisent des évènements de recombinaison (crossing-over), qui peuvent conduire à des échanges de segments chromosomiques. Ces recombinaisons sont plus fréquentes entre 2 loci éloignés, et plus rares entre 2 loci proches. Le taux de recombinaison entre 2 loci, noté  $r$ , est donc dépendant de la distance génétique entre les marqueurs : plus les marqueurs sont éloignés, plus leurs fréquences de recombinaison auront tendance à être élevées.

Le DL d'une population n'est donc pas statique, mais au contraire évolue avec le temps. Cette évolution est fonction de  $r$ , et du nombre de générations écoulées, et peut être exprimée par cette formule :

$$D_i = D_0 * (1 - r)^i$$

où  $D_0$  est le DL initial et  $i$  le nombre de générations.

Cette formule a permis de tracer la figure 1.6. Il y apparaît que le déséquilibre de liaison se maintient plus longtemps entre locus proches, mais également qu'en cas de DL initial important, celui-ci persiste plus longtemps. En plus de cette évolution due aux recombinaisons, le DL peut aussi évoluer en fonction d'autres évènements, comme le croisement avec une population ayant des fréquences alléliques différentes sur plusieurs loci, des mutations, dérive génétique ou par la sélection (Ytournal, 2008). Ainsi, le DL d'une population est un reflet de son passé, ce qui implique que le DL peut être très différent entre deux populations d'une même espèce.

Le DL est essentiel pour l'évaluation génomique, car avec celui-ci, l'utilisation d'un nombre élevé de SNPs réparti sur le génome permet, *a priori*, de capturer l'ensemble de l'information des QTLs. Étant donné que cette association préférentielle est influencée par l'histoire d'une population du fait des fréquences alléliques et des recombinaisons, il est particulièrement important d'avoir une population de référence pour l'estimation du DL qui soit représentative de la population candidate (Lande

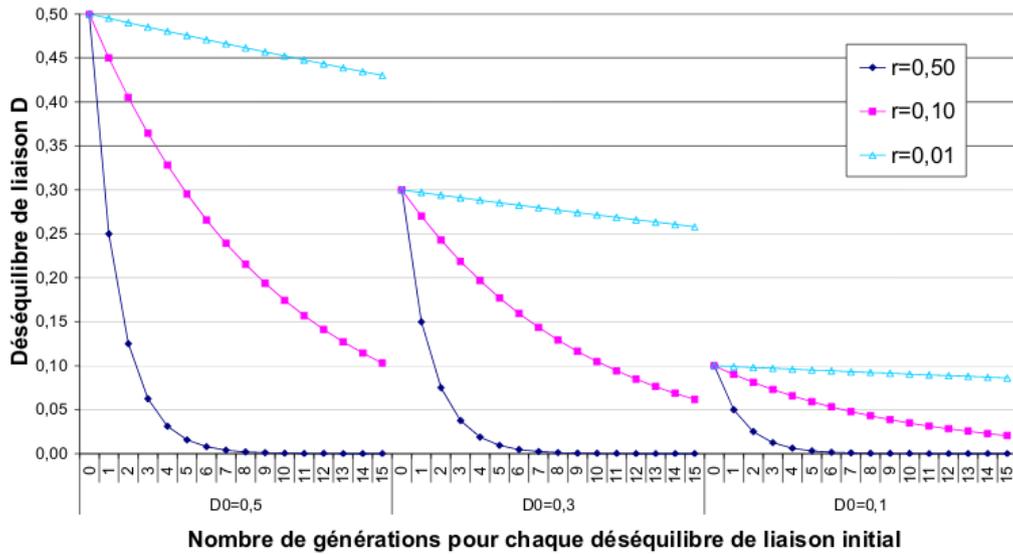


FIGURE 1.6 – Décroissance du DL en fonction du temps, taux de recombinaison ( $r$ ) et DL initial ( $D_0$ ).  
Source : Ytournal (2008)

and Thompson, 1990).

Il existe une spécificité concernant le DL chez les oiseaux. Il s'agit de la présence de microchromosomes (Burt, 2002), qui sont extrêmement denses en gènes, et qui présentent des taux de recombinaison élevés (McQueen et al., 1998). Cela induit que ces micro-chromosomes présentent un DL plus faible que les macro-chromosomes (Megens et al., 2009; Hérault et al., 2018) (cf figure 1.7). Avec un DL plus faible, il est donc plus difficile de capter l'information des QTLs présents sur ces microchromosomes, ce qui pourrait impacter la précision de l'évaluation génomique.

Le DL plus faible a un effet sur la taille des blocs haplotypiques, ceux-ci sont de plus petite taille dans les microchromosomes. Cependant, pour la lignée utilisée au cours de la thèse, cela n'impacte pas la quantité de SNPs dans chacun de ces blocs, qui reste en moyenne la même (cf figure 1.8). On a donc une quantité plus importante de SNPs sur les microchromosomes, ce qui permet de mieux couvrir ces zones riches en gènes et capturer les effets des QTLs. Les fabricants de puces réussissent très bien à optimiser celles-ci pour gérer le taux important de recombinaisons des microchromosomes. Cela a été rendu possible en concevant la puce en tenant compte des distances génétiques. Ainsi, la densité de marqueurs sur les microchromosomes est ajustée au taux de recombinaison, ce qui permet d'obtenir des blocs haplotypiques d'une taille physique variable, mais avec une composition semblable en nombre de SNPs.

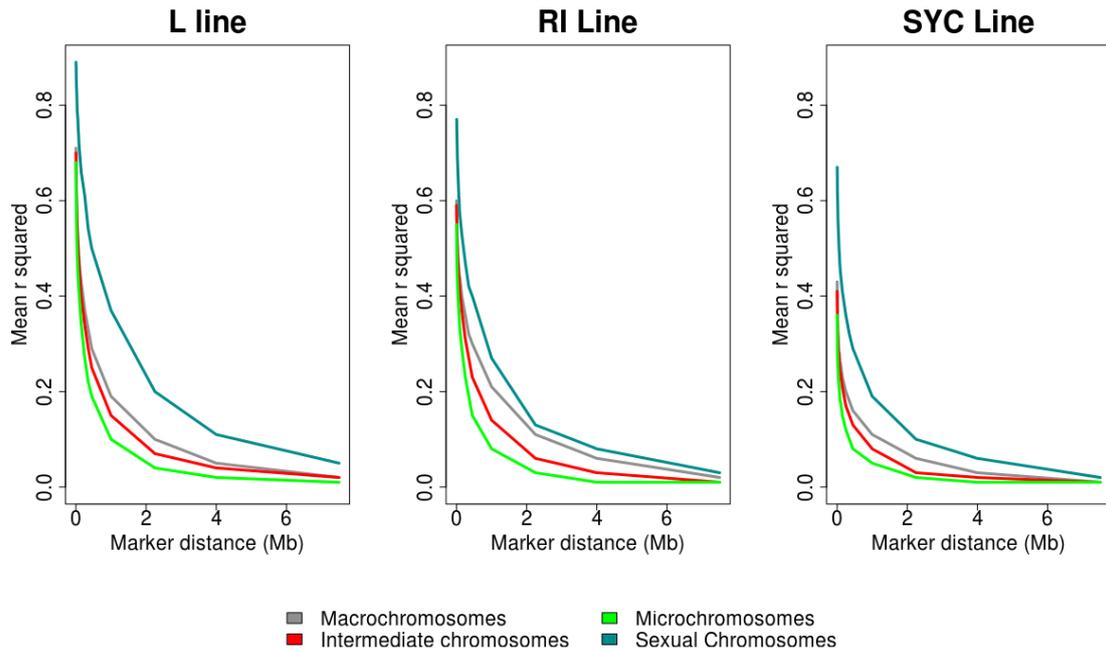


FIGURE 1.7 – Déséquilibre de liaison entre marqueurs en fonction de la distance les séparant, dans 2 lignées de poules pondeuses (L et RI) et une lignée de poulets de chair.

Source : Hérault et al. (2018)

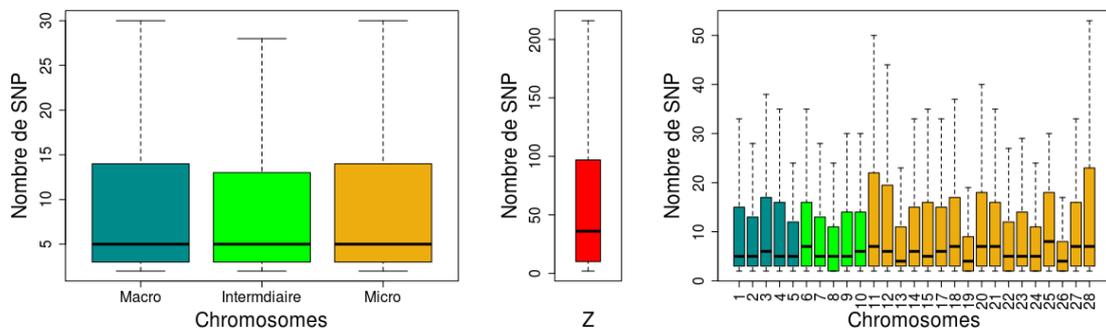


FIGURE 1.8 – Taille des blocs haplotypiques et nombre de SNPs dans ces blocs, en fonction du chromosome, pour la lignée Rhode Island et la puce à SNPs utilisée dans cette thèse.

Source : Hérault et al. (2018)

### Utilisation des SNPs dans l'évaluation

Plusieurs approches ont été développées pour intégrer les effets des SNPs en évaluation via différentes méthodes d'évaluation, afin d'obtenir la Genomic Estimated Breeding Value (GEBV) des candidats à la sélection (Robert-Granié et al., 2011). Cette partie se concentrera sur la méthode du GBLUP, et son extension ssGBLUP qui a été la méthode utilisée pour les évaluations au cours de la thèse.

## Le BLUP génomique (GBLUP)

Le GBLUP présenté par Meuwissen et al. (2001) est équivalent à un BLUP classique (cf sous-section 1.2.4 pour plus de détails), où la matrice de parenté  $A$  calculée à partir du pedigree est remplacée par une matrice de parenté génomique  $G$  (Habier et al., 2007; Goddard, 2009). Cela revient à remplacer dans le BLUP la matrice de parenté espérée par une matrice de parenté réalisée, calculée d'après les génotypes des individus.

La matrice de parenté génomique  $G$  est définie par :

$$G = \frac{ZZ'}{2 \sum p_i(1 - p_i)}$$

avec  $Z$  la matrice contenant dans chaque colonne le génotype de chaque SNP et en ligne les différents individus ; et  $p_i$  la fréquence d'un des allèles du SNP  $i$  (VanRaden, 2008).

Comme pour le BLUP, l'estimation des valeurs génétiques se fait en minimisant la variance résiduelle, en résolvant les équations du modèle mixte.

## GBLUP en une étape : ssGBLUP

Pour mettre en place l'évaluation des candidats avec un GBLUP, il est nécessaire d'avoir accès au génotype d'individus phénotypés. Comme le génotypage est une technique encore coûteuse, tous les animaux phénotypés ne sont pas systématiquement génotypés.

Le modèle single step (Aguilar et al., 2010) permet de prendre en compte, en plus des individus génotypés et phénotypés et des individus seulement génotypés, les individus qui sont seulement phénotypés, soit l'ensemble des informations disponibles pour la population étudiée.

Cela se fait par l'utilisation d'une matrice  $H$ , qui remplace la matrice  $G$  qui elle-même remplaçait la matrice  $A$  du BLUP (Legarra et al., 2009). Du fait d'une quantité importante de phénotypes d'individus non-génotypés, j'ai utilisé le modèle ssGBLUP au cours de ma thèse pour les évaluations génomiques.

## Avantages et limites de la sélection génomique

Il y a plusieurs intérêts à utiliser la sélection génomique. Tout d'abord, l'utilisation des informations génomiques doit permettre d'améliorer la précision des valeurs génétiques estimées, notamment pour différencier les plein-frères et pleine-sœurs entre eux. En évaluation génétique, tant que ces individus n'ont pas de performances propres ou de descendance avec performances, ils ont la même valeur génétique estimée, là où la connaissance de leur génotype permet de les différencier en estimant l'aléa de méiose (Hayes et al., 2009).

Ensuite, il est attendu que la sélection génomique améliore la précision de l'évaluation quand les animaux évalués ne sont pas ceux qui ont les performances (Muir, 2007), comme un coq pour des caractères de qualité d'œuf, ou un taureau pour la production de lait. De même, une augmentation de la précision des valeurs génétiques estimées est également attendue sur des caractères mesurés une fois l'animal abattu, comme la qualité de la viande (Liu et al., 2014), ou qu'il n'est pas envisageable de tester directement sur les candidats à la sélection, comme les caractères de robustesse ou de santé.

Enfin, l'ADN pouvant être obtenu dès la naissance (voir avant), il est possible d'estimer la valeur génétique d'un individu très jeune, et ainsi diminuer l'intervalle de génération en mettant à la reproduction dès leur maturité sexuelle les individus sélectionnés sans avoir à attendre des performances. Cela permet également d'augmenter l'intensité de sélection si tous les individus n'étaient pas conservables comme candidats, notamment pour des raisons de coût d'entretien. De plus, la sélection à la naissance a l'avantage de permettre au sélectionneur de n'élever que les reproducteurs. Par contre, diminuer fortement l'intervalle de génération peut être dommageable sur la précision de l'évaluation, étant donné que l'on dispose de moins de performances pour évaluer les individus.

Il y a cependant des limitations à la sélection génomique. Tout d'abord, elle est influencée par la taille de la population de référence (VanRaden, 2008), ce qui peut limiter les gains liés à l'utilisation de cette méthode sur des populations de petite taille (Aguilar et al., 2010). De plus, si le génotypage des individus permet une meilleure évaluation des reproducteurs, il s'agit également d'une technique assez coûteuse (e.g. une puce de 600K SNP revient actuellement à environ 150€), ce qui peut être rédhibitoire quand le nombre d'individus à génotyper est élevé.

Ensuite, si la sélection génomique permet de sélectionner plus rapidement, ce n'est pas sans désavantage. Un intervalle de génération réduit signifie qu'avec une augmentation de la consanguinité donnée à chaque génération, dans un intervalle de temps  $T$ , la consanguinité augmentera davantage avec la sélection génomique (Meuwissen et al., 2013). Cet effet peut être limité en utilisant des règles supplémentaires pour la sélection des reproducteurs, en limitant par exemple le nombre d'individus sélectionnés par famille, ou constituer des couples d'animaux avec le moins d'apparentement possible.

Une sélection plus rapide signifie également des changements dans la structure du DL (Calus, 2010). Cela implique qu'il faut mettre à jour régulièrement la population de référence, sous peine de voir la qualité de l'évaluation se dégrader (Wolc et al., 2011).

### 1.2.6 Application de la sélection génomique

Peu d'études ont été publiées sur l'utilisation de la sélection génomique chez la poule pondeuse. Mais celles existantes ont confirmé le potentiel de la sélection génomique pour la filière avicole (Wolc et al., 2016b). Il a effectivement été montré

que l'utilisation de la sélection génomique permettait d'obtenir dans le même laps de temps que la sélection classique une meilleure réponse des individus à la sélection sur 16 caractères de production, avec un progrès génétique estimé jusqu'à 2 fois plus important pour la sélection génomique (Wolc et al., 2015).

Concernant la précision de l'évaluation génomique, sur les poulets de chair, l'utilisation du GBLUP a permis une augmentation moyenne de 5% de précision par rapport à un BLUP, sur 5 critères d'abattage (Liu et al., 2014). Sur le poids de l'œuf et le poids de carcasse, il a été montré que les méthodes génomiques permettaient une précision moyenne de 0.25 pour GBLUP là où les évaluations BLUP obtiennent en moyenne une précision de 0.07 (Wolc et al., 2016a).

## 1.3 Précision de l'évaluation

### 1.3.1 Importance de la précision de l'évaluation

Comme nous l'avons vu, la précision de l'évaluation est un paramètre clé pour optimiser le progrès génétique dans les populations sélectionnées.

En fonction du schéma de sélection utilisé, ainsi que du type d'évaluation effectué, le progrès génétique que l'on peut attendre à la génération suivante peut fortement varier. Pour être capable de choisir le schéma de sélection assurant un progrès génétique optimal, il est donc important de connaître la précision de l'évaluation dans les différents cas de figure, et par exemple déterminer le meilleur moment pour choisir les futurs reproducteurs.

La précision d'une évaluation est la corrélation entre la vraie valeur génétique des individus (True Breeding Value, TBV) et les valeurs génétiques estimées, (G)EBVs. Or, cette TBV n'est jamais connue quand on travaille avec des données réelles. Cet écueil peut être contourné par des approches de validation croisée, en comparant les (G)EBVs soit aux phenotypes observés (Legarra et al., 2008), soit à des EBVs provenant d'une évaluation sur descendance (Daetwyler et al., 2013). Il est également possible d'utiliser la valeur moyenne des descendants, ajustée pour les différents effets du modèle, fixes et aléatoires non-génétiques, (Daughter Yield Deviation, DYD) pour approcher la TBV (VanRaden, 2008).

### 1.3.2 Formules théoriques de la précision de l'évaluation

Depuis le début de la sélection génomique, des formules ont été développées pour en estimer la précision (Daetwyler et al., 2008; Goddard, 2009; Goddard et al., 2011; Meuwissen et al., 2013). Ces formules estiment la précision attendue de l'évaluation à partir de plusieurs paramètres : l'héritabilité du caractère, le nombre d'individus dans la population de référence, le nombre de SNPs et le nombre de blocs haplotypiques dans le génome. Ces formules permettant de calculer par exemple la précision à attendre pour une évaluation sur un certain nombre d'individus, ou le nombre d'individus nécessaires pour atteindre une précision donnée.

Cependant, la précision estimée par ce moyen peut être assez éloignée de celle observée. Brard and Ricard (2015) ont notamment observé (sur le cheval) que la précision estimée par les formules théoriques était souvent surestimée par rapport à celle observée. Les écarts entre précision attendue et observée varient également en fonction des caractères, une formule pouvant prédire correctement la précision observée sur un caractère et surestimer fortement celle attendue sur un autre. Les précisions estimées par ces formules sont donc à utiliser avec prudence, la précision attendue pouvant être assez éloignée de celle observée. C'est pour cette raison que les précisions des évaluations effectuées lors de cette thèse ont été estimées à partir des (G)EBVs.

### 1.3.3 Facteurs influant sur la précision de l'évaluation

La précision d'une évaluation est influencée par plusieurs facteurs. Ces facteurs peuvent être regroupés en deux grands groupes. Le premier contient les facteurs qui ne peuvent pas être ajustés, comme l'héritabilité du caractère évalué. Le second regroupe de fait les facteurs qui peuvent être modifiés, comme la méthode d'évaluation et, dans le cas d'une évaluation génomique, la quantité de SNPs utilisé pour le génotypage. Ce sont ces paramètres, qui peuvent être ajustés pour optimiser la qualité de l'évaluation, qui vont être détaillés dans cette partie.

#### Population de référence

La composition de la population de référence et le nombre d'individus qui la compose, sont des paramètres clés pour obtenir une bonne précision de l'évaluation génomique, notamment sur les caractères les moins héritable.

#### Taille de la population de référence

Il est reconnu depuis le début de l'évaluation génomique que cette population de référence devait être d'une taille suffisante (Meuwissen et al., 2001; Daetwyler et al., 2008), notamment pour les caractères peu héritable (par simulation, Hayes et al. (2009); et chez les bovins, Luan et al. (2009)). D'après Meuwissen (2009), sur simulation, le nombre d'individus dans la population de référence devrait être le résultat de  $2N_eL$ ,  $N_e$  étant la taille efficace de la population, et  $L$  la longueur de génome en Morgan.

Il a été effectivement montré sur données réelles (cf figure 1.9) qu'une augmentation de la taille de la population de référence avait un effet positif sur la précision de l'évaluation, et ce chez plusieurs espèces, comme les bovins (VanRaden et al., 2009) ou les caprins (Daetwyler et al., 2010a).

Il a cependant également été constaté que ce gain n'est pas linéaire, et finit par atteindre un palier où l'ajout d'animaux dans la population de référence ne permet plus de gain de précision. Sur données réelles, chez le bovin, au-delà de 5 000 individus, Erbe et al. (2013) ne trouvent plus d'amélioration de la précision (cf figure 1.10).

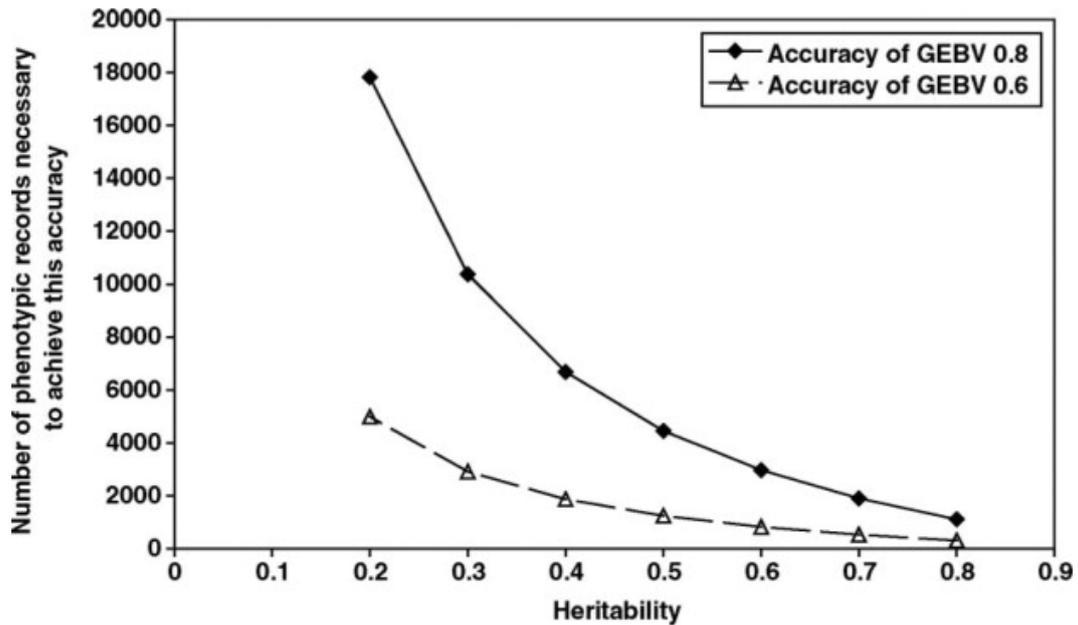


FIGURE 1.9 – Nombre d’individus nécessaires (en ordonnée), pour atteindre une précision donnée en fonction de l’héritabilité du caractère (en abscisse), d’après les formules de Goddard and Hayes (2007).

Source : Hayes et al. (2009)

À l’inverse, un faible nombre d’animaux dans la population de référence limite la précision de l’évaluation des GEBVs des candidats (Brito et al., 2011). D’après Liu et al. (2014), une faible variabilité génétique de la population permettrait de minimiser l’effet d’une petite population de référence : le nombre d’haplotypes à intégrer au modèle serait plus faible, et donc plus facilement géré avec un nombre réduit d’individus.

La taille minimale de la population de référence dépend également de la structure génétique des caractères étudiés, ainsi que du nombre d’informations disponibles par individu : les caractères influencés par de nombreux QTLs à effets faibles nécessitent beaucoup de performances pour que l’estimation des effets des SNPs soit effectuée de manière optimale (Hayes et al., 2009).

### Composition de la population de référence

Le nombre d’individus qui composent la population de référence est important, mais l’apparentement de la population de référence avec la population candidate a également un effet sur la précision de l’évaluation.

En effet, l’évaluation est d’autant plus précise si la population de référence est apparentée à la population candidate (Liu et al., 2011b; Habier et al., 2013; Elsen, 2016). Chez le poulet de chair, en validation croisée, Liu et al. (2014) ont montré que la précision de l’évaluation génomique est plus importante quand les animaux sont répartis dans des groupes aléatoires (précision GBLUP moyenne : 0.430) en comparaison de celle observée quand les groupes sont constitués en minimisant

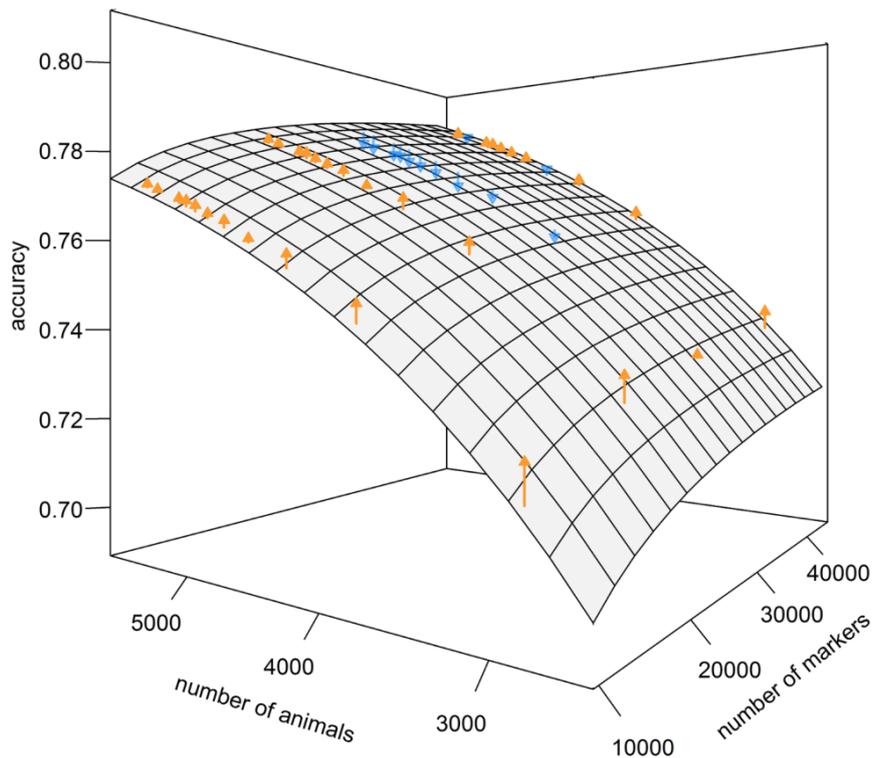


FIGURE 1.10 – Précision de l'évaluation, estimée et observée, en fonction du nombre d'animaux dans la population de référence, et du nombre de marqueurs utilisés pour le génotypage. Le plan est le résultat théorique obtenue à partir d'une équation modifiée de Daetwyler et al. (2010b), les flèches représentent les résultats obtenus sur les données réelles des taureaux.

Source : Erbe et al. (2013)

l'apparentement (précision GBLUP moyenne : 0.318). L'explication avancée est que la répartition aléatoire dans les groupes permet d'avoir des pleins- et/ou demi-frères et sœurs dans la population de référence, cet apparentement permettant une meilleure estimation des valeurs génétiques.

L'apparentement entre population référence et candidate permet donc d'estimer plus précisément les valeurs génétiques, les SNPs capturant les relations de parenté.

L'effet de l'apparentement est lui-même affecté par la taille de la population de référence : celui ci sera plus faible dans une population de référence de grande taille, et réciproquement, il sera plus important en cas de population de référence de petite taille (Wientjes et al., 2013) (sur simulation, cf figure 1.11).

Plutôt que d'envisager la population comme un groupe d'individus isolés, une autre approche est de considérer le nombre de générations d'individus à inclure dans la population. D'après Muir (2007), par simulation d'une population animale, la précision de l'évaluation serait meilleure quand on utilise plusieurs générations plutôt qu'une seule génération. Un caractère avec une héritabilité à 0.5, estimée avec une précision de 0.2 avec une seule génération dans la population de référence

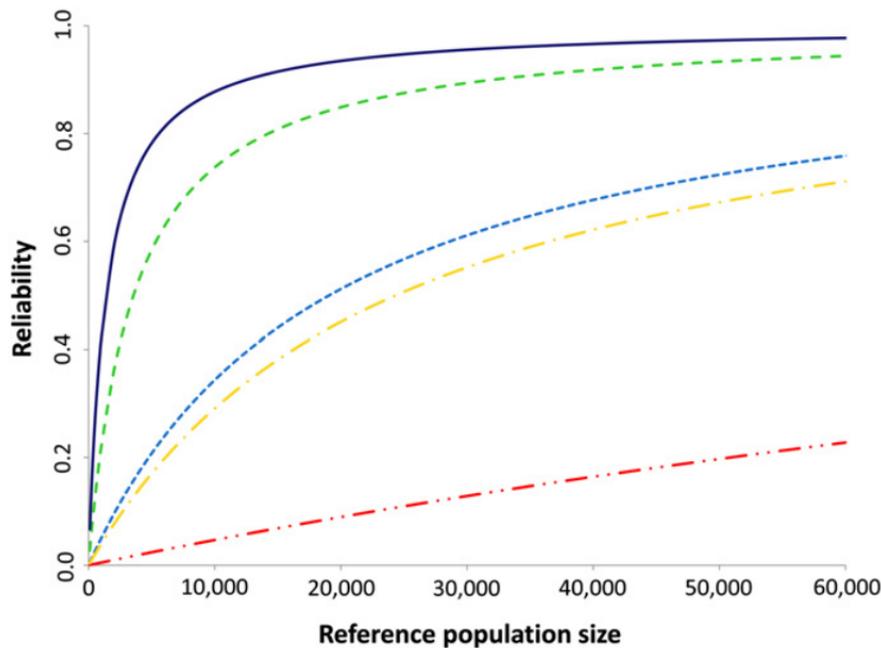


FIGURE 1.11 – Précision des résultats du GBLUP en fonction de la taille de la population de référence et du type de relation entre population de référence et candidate. Résultats obtenus par simulation.

La ligne continue bleu foncé présente les précisions obtenues avec des candidats apparentés aux individus dans la population de référence (individus réels). La ligne hachurée en bleu clair présente les précisions obtenues par des candidats simulés en utilisant le DL et les fréquences alléliques de la population de référence.

En rouge sont les résultats quand la population candidate est simulée à partir des fréquences alléliques de la référence. En jaune, quand les candidats sont simulés en combinant les haplotypes de la population de référence. En vert, quand les candidats sont simulés par combinaison de chromosomes de la population de référence.

Source : Wientjes et al. (2013)

verra la précision de l'évaluation augmenter à 0.5, 0.65 et 0.75 quand respectivement 2, 3 et 4 générations sont utilisées dans la population de référence, avec donc une augmentation de la taille de la population de référence.

Par contre, avec une population de référence de même taille, il a été montré en simulation qu'il n'y avait aucune différence de précision entre les GEBVs obtenues en GBLUP en utilisant une population de référence comprenant 1 seule génération et celles obtenues avec une population de référence de 10 générations (Bastiaansen et al., 2012).

Les études sur l'avoine (Asoro et al., 2011), la vache et le cochon (Lourenco et al., 2014) ainsi que le poulet de chair (Weng et al., 2016), ont montré que sur données réelles, augmenter le nombre de générations inclus dans la population de référence a un effet positif sur la précision de l'évaluation. Cependant, à partir d'un certain seuil, il n'y a plus de gain, voir une perte de précision (Lourenco et al., 2014; Weng et al., 2016) (cf figure 1.12).

Weng et al. (2016) montrent également un lien entre nombre optimal de générations et héritabilité du caractère. Un caractère très héritable tel que la couleur de l'œuf (héritabilité estimée à 0.68 dans leur lignée) a un nombre optimal de générations de 6, alors que pour le caractère de production d'œuf, avec une héritabilité estimée à 0.25, ce nombre optimal de générations est de 3. Aux vues des résultats obtenus, Weng et al. (2016) ont conclu que 4 générations était le nombre de générations optimale à utiliser pour maximiser la précision de l'évaluation sur le maximum de caractères.

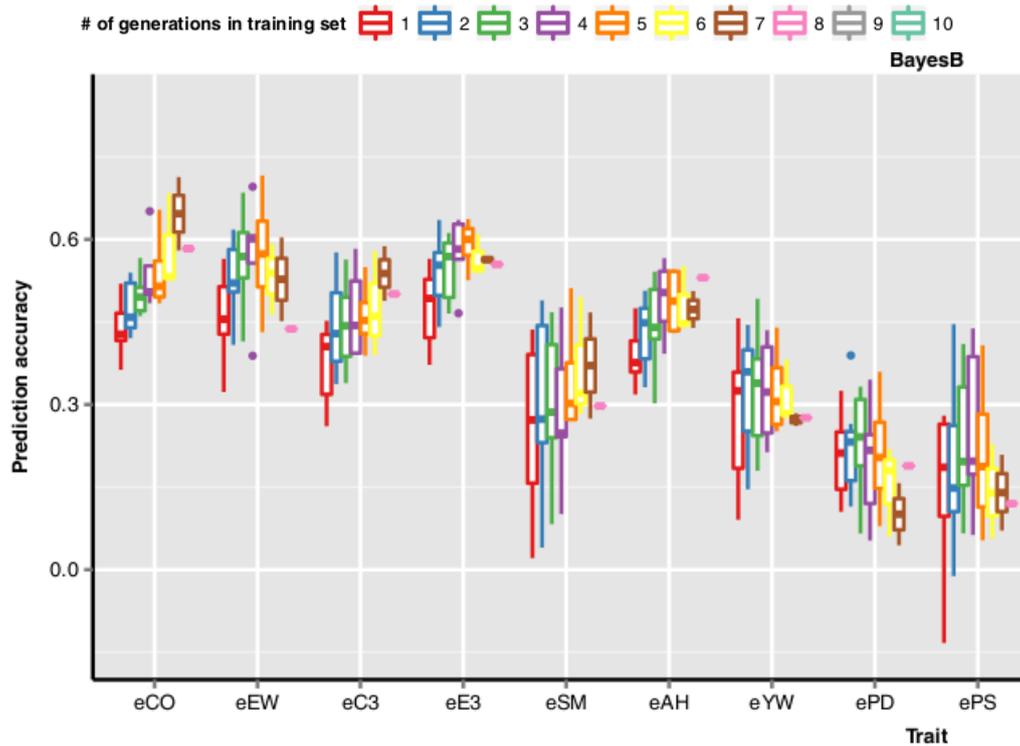


FIGURE 1.12 – Évolution de la précision de l'évaluation sur plusieurs caractères de production d'œuf chez la poule pondeuse, en fonction du nombre de générations prises en compte.

Caractères : eCO : couleur de la coquille, eEW : poids de l'œuf, eC3 : couleur des 3 premiers œufs, eSM : age de maturité sexuelle, eAH : hauteur d'albumen, eYW : poids du jaune, ePD : taux ponte, ePS : perforation de la coquille. Performances récoltées entre 26 et 28 semaines.

Source : Weng et al. (2016)

### Enrichissement de la population de référence

L'enrichissement de la population de référence au cours du temps est également un point important. En effet, il a été montré que la précision de l'évaluation génomique décroît quand des générations successives sont évaluées à partir d'une même population de référence (Muir, 2007; Zhong et al., 2009). Cette perte de précision est particulièrement importante au premier saut de génération, et se poursuit de manière moins marquée sur les suivants.

Un caractère estimé avec une précision de 0.43 en génération 1 avec une population de référence composée de la génération 0, est estimé en génération 2 avec une précision de 0.33 si la même population de référence est conservée (Wolc et al., 2011). Cependant, si la population de référence est constituée par les générations 0 et 1, la précision d'évaluation du caractère en génération 2 est de 0.49. Cela est lié au DL (cf section 1.2.5). Le DL d'une population se modifiant à chaque génération, si la même population de référence est conservée, la relation entre génotype et phénotypes faite par l'équation de prédiction sera de moins en moins pertinente pour les individus candidats. Il est donc recommandé de maintenir à jour la population de référence en ajoutant les données des nouveaux animaux pour garder une précision optimale de l'évaluation.

## Méthodes d'évaluation

Les modèles présentés par Meuwissen et al. (2001) suivent des hypothèses sur la distribution de l'effet des SNPs. Le modèle infinitésimal du GBLUP suppose que chacun des SNPs explique une part de la variance additive, là où les méthodes bayésiennes (BayesA, BayesB, etc) supposent qu'une faible part des SNPs a un effet non nul sur les performances, et introduisent (ou infèrent à partir des données) cet *a priori* dans le modèle.

D'après plusieurs études, le GBLUP permettrait des estimations aussi bonnes ou meilleures que les méthodes bayésiennes sur les caractères polygéniques. Cela a été montré notamment en simulation (Daetwyler et al., 2010b), pour l'orge (Zhong et al., 2009), et les bovins (Verbyla et al., 2009). Cependant, les modèles bayésiens permettraient des évaluations plus précises sur les caractères influencés par quelques QTLs ayant un effet important, la distribution *a priori* des méthodes bayésiennes permettant une meilleure approximation de la réalité sur ces caractères, comme montré sur simulations par Goddard and Hayes (2007).

Daetwyler et al. (2010a) ont notamment trouvé que la précision des évaluations en utilisant la méthode BayesB était supérieure à celle obtenue avec le GBLUP, 0.75 vs 0.60 respectivement, dans le cas où le nombre de QTLs à affecter le caractère était réduit (300). À l'inverse avec un nombre de QTLs plus élevé (2 000), la précision de GBLUP était toujours estimée à 0.60, et celle du BayesB à 0.55.

Sur des données réelles, les valeurs génétiques des caractères connus pour être influencés par peu de QTLs à forts effets sont effectivement mieux estimées avec les modèles bayésiens (Zhong et al., 2009; Daetwyler et al., 2010a). Wolc et al. (2016a) ont montré que sur les caractères du poids d'œuf et du poids corporel, la précision de l'évaluation était de 0.32 avec le BayesB là où elle est de 0.25 en utilisant le GBLUP.

Ces différences sont cependant à nuancer. Les valeurs génétiques estimées par ces différentes méthodes peuvent être très fortement corrélées entre elles (VanRaden et al., 2009), en ayant une précision équivalente (Habier et al., 2011). Sur des caractères de carcasses de poulet de chair, la précision moyenne des GEBVs obtenue avec le GBLUP est de 0.318, et de 0.320 avec la méthode BayesMix4 (Liu et al., 2014).

Les changements de distribution *a priori* des différents modèles bayésiens seraient également peu impactants sur la précision de l'évaluation (Verbyla et al., 2010), les GEBVs obtenues par les différentes méthodes bayésiennes étant toutes fortement corrélées, entre 0.99 et 0.89.

Concernant le GBLUP, Daetwyler et al. (2010a,b) ont montré que celui-ci ne serait en fait que peu sensible à l'architecture génétique du caractère, cette méthode ne présentant que peu de différences en termes de précision d'évaluation en fonction du nombre de QTLs expliquant la variabilité du caractère. Dans Daetwyler et al. (2010b), la précision de l'évaluation par le GBLUP reste de 0.60, que le nombre de QTL soit de 300 ou 2300.

Selon Meuwissen et al. (2013), les faibles différences entre le GBLUP et les méthodes bayésiennes observées en pratique peuvent s'expliquer par le fait que le nombre de QTL à faible effet est important. Cela fait que l'assertion que chaque SNP est en DL avec un QTL est vrai, et que chaque SNP a donc un effet non-nul est également vrai. Dans la pratique, si les modèles bayésiens sont sensés mieux décrire la réalité biologique des caractères, le modèle du GBLUP permet le plus souvent d'obtenir une précision convenable.

### **Nombre de SNPs utilisés**

La sélection génomique reposant sur les SNPs pour capturer l'effet des QTL grâce au DL, il faut un nombre suffisant de marqueurs répartis sur le génome pour que chaque QTL soit en DL avec un marqueur.

L'augmentation du nombre de SNPs a un effet généralement positif sur la précision de l'évaluation (VanRaden et al., 2009; Brito et al., 2011; Habier et al., 2013), cf figure 1.10. En densifiant la carte génétique en SNPs, on augmente la probabilité qu'ils soient en DL élevé avec un QTL. La densité optimale va varier en fonction de la taille du génome, et du DL dans la population évaluée (Goddard and Hayes, 2007).

Un nombre important de SNPs permet également de limiter l'importance de la répartition des SNPs sur le génome (Habier et al., 2009; Moser et al., 2010). Quand le nombre de SNPs est limité, la précision de l'évaluation peut être impactée par leur répartition, car ils ne seront pas capables de capter l'ensemble des effets des QTLs si le DL n'est pas suffisant.

Cependant, passé un certain nombre de SNPs, il n'y aura plus d'amélioration de la précision de l'évaluation avec l'ajout de SNPs (Jensen et al., 2012; Erbe et al., 2013). D'après (Erbe et al., 2012), chez le bovin, entre une puce 20k SNPs et une 40k SNPs, la part de variance génétique capturée par les SNPs augmente très peu (74.5% vs 76.6%) alors que la densité est multipliée par 2. Un nombre important de SNPs permet de capturer la majorité de la variance génétique, les SNPs excédentaires au delà ne permettant pas d'obtenir significativement plus d'information.

Les coûts de génotypage peuvent être un obstacle financier important, qui peut être un frein à une utilisation optimale de l'évaluation génomique en limitant le

nombre d'animaux génotypés (cf partie 1.2.5). Dans le but de réduire ces coûts, il est possible de s'orienter vers l'imputation pour génotyper les individus avec des puces moyenne ou basse densité, moins coûteuses que les puces haut débit.

L'imputation implique de prédire le génotype haute densité des individus à partir de leur génotypage basse densité, et du génotype haute densité d'une population de référence (Marchini et al., 2007). Dans Herry et al. (2018), qui porte sur les mêmes populations que celles évaluées dans cette thèse, il a été montré que cette approche fonctionne très bien chez la poule pondeuse. Les imputations permettent d'obtenir des génotypes Haute Densité (HD) imputés dont la corrélation avec les véritables génotypes HD variait entre 0.86 et 0.99, en fonction de la puce de départ, de la méthode de sélection des SNPs sur le génome et de la population étudiée.

L'imputation a l'avantage de permettre de génotyper des individus à moindre coût, mais nécessite tout de même de régulièrement génotyper des individus à haut-débit, afin de garder des imputations de qualité (Habier et al., 2009; Dasonneville et al., 2011)

### 1.3.4 Application de la sélection génomique

En plus du cadre théorique, la sélection génomique a été permise par les progrès technologiques effectués ces dernières années, avec la fabrication de puces à SNPs à haut débit, telle que la puce Affymetrix<sup>®</sup> Axiom<sup>®</sup> 600K SNPs pour le poulet (Kranis et al., 2013), permettant un génotypage à haut débit à un coût raisonnable. Cette combinaison entre les aspects théoriques et la technique a permis de mettre en pratique la sélection génomique.

Dans la filière bovine, qui a été précurseur dans son utilisation, la sélection génomique a permis des gains substantiels de progrès génétiques (VanRaden et al., 2009; Weber et al., 2012). Les résultats observés chez les bovins ne sont cependant pas directement transférables à la sélection des poules pondeuses.

Ainsi, au premier rang des différences, l'utilisation systématique du croisement dans les schémas de sélection pyramidaux.

Ensuite, si l'utilisation des DYD permettent d'obtenir chez les bovins une estimation de la précision sans connaître la TBV des animaux, ces valeurs ont cependant quelques limitations. Il faut corriger les phénotypes et il peut être difficile d'obtenir des DYDs fiables quand des animaux avec peu ou pas de descendance sont évalués, ce qui est le cas chez la poule pondeuse. Cela peut provoquer certaines incohérences, telles que des précisions estimées supérieures à 1 (Legarra and Reverter, 2017). Une autre approche, la méthode LR (Legarra and Reverter, 2018), a récemment été développée dans l'objectif de dépasser ces limitations et de compléter l'approche de validation croisée avec des composants semi-paramétriques, basés sur les modèles classiques des évaluations génétiques. L'intérêt de procédures semi-paramétriques basées sur les équations du modèle mixte est de combiner l'aspect théorique, que l'on sait représentatif du fonctionnement biologique, avec des indicateurs qui ne sont pas influencés par un modèle.

## 1.4 Conclusion du contexte bibliographique

La précision de l'évaluation est donc un élément central pour pouvoir estimer le progrès génétique des populations sélectionnées en utilisant l'évaluation génomique. Cette précision est affectée par de nombreux facteurs, qui interagissent entre eux, rendant difficile l'élaboration d'une formule permettant d'estimer la précision d'une évaluation à partir de quelques paramètres. Il n'existe donc pas de règles simples pour définir les conditions pour lesquels la précision de l'évaluation sera optimale. Estimer celle-ci sur données réelles est le moyen le plus sûr d'avoir des résultats qui soient proches de ceux qui seront obtenus en production.

Le chapitre suivant présente en détails les données utilisées lors de cette thèse, ainsi que les méthodes utilisées pour effectuer les évaluations et en analyser les résultats.

# Chapitre 2

## Matériel et méthodes

### 2.1 Animaux

Toutes les données étudiées proviennent d'animaux sélectionnés et élevés à but commercial par la société de sélection Novogen (Plédran, France), filiale du Groupe Grimaud. Les mesures de qualité d'œufs ont été réalisées par la société Zootests (Ploufragan, France).

En sélection avicole, les poules sont élevées en lot, qui regroupent des individus d'une même génération. La conduite de l'élevage se fait au niveau du lot, les individus le composant traversant tous les mêmes étapes au même moment.

#### 2.1.1 Animaux de lignée pure

Les poules pondeuses qui composent la lignée pure ont été sélectionnées à partir d'une lignée Rhode Island Red, qui pond des œufs bruns.

#### Élevage et contrôle des performances

Dans chaque lot, les poulettes sont gardées en poussinières jusqu'à l'âge de 18 semaines, où elles sont transférées en cages collectives de 5 pleines-sœurs (mêmes père et mère, environ 2 000 poules/lot). Les poussins mâles sont également transférés à 18 semaines dans le même bâtiment, mais en cages individuelles (200 mâles/lot, sur 2 000 poussins). Les animaux resteront dans ce bâtiment jusqu'à ce qu'ils aient atteint l'âge de 60 semaines.

La qualité des œufs est mesurée 2 fois pendant cette période, à 30 et 50 semaines d'âge.

À 60 semaines d'âge, une première sélection a lieu, dans l'objectif principal d'éliminer les familles dans lesquelles il y a eu un comportement de picage lors de cette étape d'élevage en cages collectives. Une partie (150 mâles et 650 femelles) des individus est alors transféré en cages individuelles. Ces individus sont conservés jusqu'en fin de carrière, aux environs de 80 semaines. À cette étape, la qualité des œufs est mesurée chaque semaine sur 1 œuf/poule.

## Structure des données analysées

L'ensemble des données de performances de qualité d'œufs concerne 4 générations de poules pondeuses élevées entre 2008 et 2015. Comme détaillé précédemment, ces poules sont élevées dans 2 milieux différents, le premier en cages collectives quand elles sont jeunes, puis, après une étape de sélection, en cages individuelles pour leur fin de carrière (cf figure 2.1).

Les performances mesurées ont donc été séparées en 2 ensembles de données, en fonction de l'étape de sélection concernée.

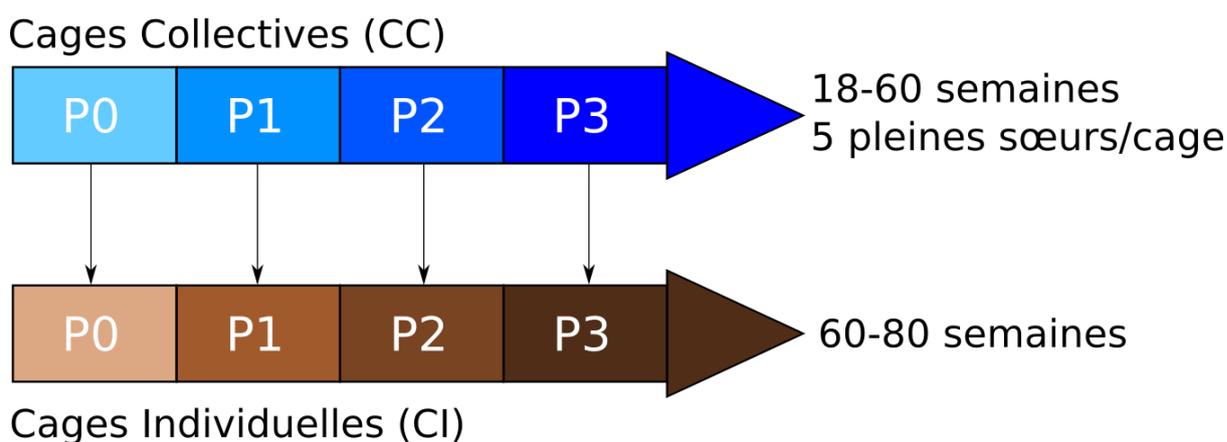


FIGURE 2.1 – Structure des populations de la lignée pure.

Le jeu de donnée Cages Collectives (CC) regroupe toutes les performances mesurées en cages collectives, sur la période d'âge 18-60 semaines, avec mesures de qualité d'œufs à 30 et 50 semaines. Ce sont des poules jeunes, avec des performances relativement homogènes (cf section 2.3). Ce jeu contient des informations sur 2 997 cages, soit 14 985 poules, avec 27 915 œufs mesurés.

Les données concernant la seconde étape d'élevage sont regroupées dans le jeu de données Cages Individuelles (CI). Il s'agit de poules plus âgées, en fin de carrière, chez qui on observe une variabilité plus importante des caractères (cf section 2.3). Ce jeu de données regroupe 74 976 performances, récoltées sur 7 982 poules.

Étaient à disposition dans ces jeux de données, outre les caractères de qualité d'œufs détaillés en section 2.3, le lot à l'origine de la mesure, la batterie dans laquelle était la cage ainsi que la position de la cage dans la batterie. Étaient également disponibles l'âge de la poule à l'origine de la ponte (en semaines), le temps écoulé entre la ponte et la mesure de l'œuf.

La généalogie de tous les individus était enregistrée dans un fichier de pedigree, qui concernait un total de 2 273 reproducteurs (514 mâles et 1 759 femelles), remontant à la génération P0. La généalogie détaillée des différents lots est représentée en figure 2.2.

## Cages Collectives (CC)

## Cages individuelles (CI)

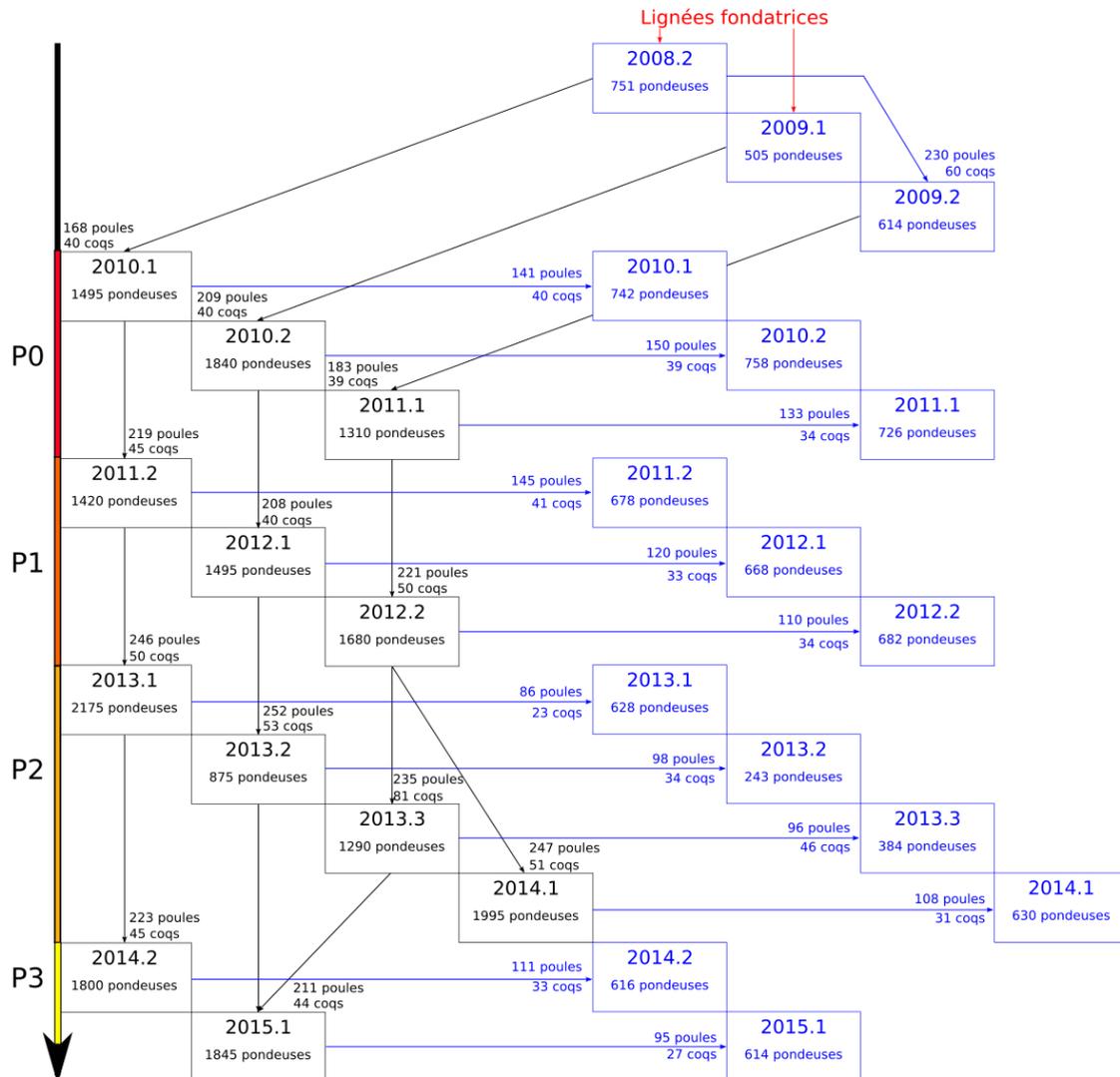


FIGURE 2.2 – Généalogie détaillée des différents lots composant la lignée pure. Pour chaque lot est indiqué le nombre de poules le composant, ainsi que le nombre de parents, mâles et femelles.

### 2.1.2 Animaux croisés

#### Élevage et contrôle des performances

Afin de disposer de données sur des poules de type "commercial" (produit terminal du schéma de sélection), Novogen place des poules issus de différents croisements dans des fermes sous contrat. Il s'agit pour le département R&D de l'entreprise de vérifier les qualités des poules commerciales. Les données acquises lors de ces contrôles ne sont actuellement pas utilisées pour la sélection des reproducteurs de lignée pure.

Les poules croisées utilisées pour nos travaux sont les descendantes de coqs reproducteurs

(1 384 au total) de la lignée pure présentée précédemment, croisés avec des femelles d'autres lignées. Ces poules ont été élevées de 2010 à 2016, sur 3 générations (cf figure 2.3).

Chaque coq avait en moyenne 45 filles obtenues par croisement avec une poule d'un autre type génétique. Ce type génétique n'était pas forcément le même pour tous les lots. Le pedigree de la voie maternelle des croisées était inconnu.

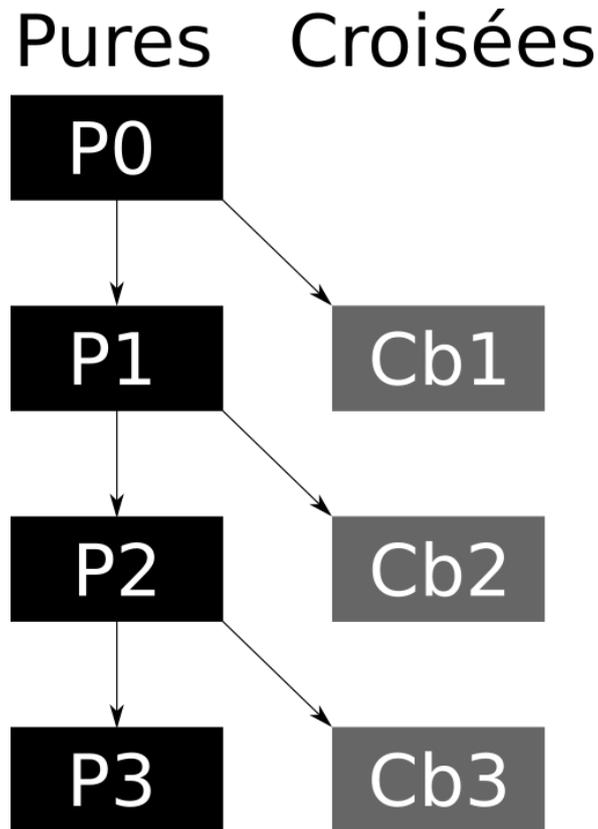


FIGURE 2.3 – Relations de parenté entre lignée pure et cheptel de poules croisées.

Tous ces individus ont été élevés en cage collective regroupant 12 demi-sœurs de père, de 18 semaines d'âge à leur fin de carrière, à environ 80 semaines d'âge. Pour chaque lot, la qualité des œufs a été mesurée 2 fois à 50 semaines et 70 semaines.

### Structure des données

Les performances des descendantes croisées ont été regroupées en fonction de l'âge auquel avait lieu la mesure. Le premier jeu de données sur les poules âgées de 50 semaines (50sem) contenait 41 343 œufs mesurés, celui sur les poules âgées de 70 semaines (70sem) en contenait 37 232.

Les mêmes informations que pour les lignées pures étaient disponibles dans ces jeux de données, à l'exception de l'identifiant de la mère, celle-ci étant inconnue.

Les jeux de données "pures" et "croisées" ont également été fusionnés en fichiers combinés, pour permettre des évaluations mixtes, utilisant à la fois les performances obtenues en lignée pure et celles en croisées. Le jeu de données CC a été fusionné avec le jeu de données 50sem, pour former un regroupement de poules jeunes (YH), tandis que le CI a été fusionné avec 70sem pour former un groupe de performances issues de poules âgées (OH).

## 2.2 Génotypes

Seuls des animaux de lignée pure ont été génotypés. Sur cette population, 2 374 individus ont été génotypés en utilisant la puce haute densité 600K SNPs Affymetrix<sup>®</sup> Axiom<sup>®</sup> (Kranis et al., 2013). Les échantillons de sang ont été collectés à partir de la veine brachiale des animaux et l'ADN extrait.

Les mâles candidats à la sélection des deux premières générations (447 mâles en P0, et 580 en P1) ont été génotypés par Ark-Genomics (Edinburgh, UK) pendant le projet UtopIGe (Romé et al., 2015). À partir de la génération P2, les mâles reproducteurs (132 en P2 et 55 en P3) et femelles candidates en CI (722 en P2 et 466 en P3) ont été génotypés à la plateforme de génotypage haut-débit Gentyane (Clermont-Ferrand, France) pour les besoins de l'évaluation génomique de la lignée. Les deux plateformes ont utilisé la suite d'analyse Axiom pour créer les fichiers de génotype au format PED (Purcell et al., 2007). L'uniformité des génotypes a été testé avec PLINK, qui a identifié 870 SNPs avec une inversion de brin visible, et Génotype Harmonizer (Deelen et al., 2014), qui a identifié 148 SNPs avec une inversion de brin invisible. PLINK a été utilisé pour corriger ces erreurs.

Chaque animal a été génotypé sur 580 961 marqueurs SNPs. En accord avec la 5<sup>ème</sup> annotation du génome de *Gallus gallus* (Warren et al., 2017), ces SNPs sont répartis sur les macrochromosomes 1 à 5, les chromosomes intermédiaires 6 à 10 et les microchromosomes 11 à 28 et 33, un groupe de liaison (LGE64), les deux chromosomes sexuels Z et W, ainsi qu'un groupe de SNPs de localisation inconnue (3 724 SNPs).

Les génotypes obtenus ont ensuite été filtrés suivant 4 étapes successives. Ont été retirés :

- Les individus avec un call rate < à 95% (aucun individu enlevé)
- Les SNPs avec une fréquence de l'allèle mineur (MAF) < 0.05
- Les SNPs avec des fréquences qui déviaient très significativement de l'équilibre d'Hardy-Weinberg ( $P < 10^{-4}$ )
- Les animaux avec des incompatibilités de pedigree (12 individus exclus)

La plupart des SNPs ont été enlevés car ils montraient une MAF de 0. Cela était attendu, étant donné que la puce à SNP a été conçue à la fois pour les poules pondeuses et les poulets de chair, avec un ratio de 1 :2.

Au final, 302 102 SNPs (cf table 2.1), concernant 1 214 mâles et 1 148 femelles ont été conservés.

Filtre	Nombre de SNPs enlevés
MAF <0.05	258 772
SNP Call Rate <0.95	7 549
Équilibre d'Hardy-Weinberg ( $P < 10^{-4}$ )	12 538
SNPs conservés	302 102

TABLE 2.1 – Résumé des différentes étapes de filtration du contrôle qualité des SNPs

## 2.3 Caractères

Les différents caractères sont nommés selon l'Animal Trait Ontology for Livestock (<http://www.atol-ontology.com/>). Cinq caractères de qualité d'œufs ont été étudiés au cours de la thèse : le poids de l'œuf (Egg Weight, EW), la couleur de la coquille (Egg Shell Color, ESC), la résistance de la coquille à la pression (Egg Shell Strength, ESS), la hauteur de l'albumen (Albumen Height, AH) et la forme de l'œuf (Egg Shell shape, ESshape).

### 2.3.1 Mesures des caractères

La première étape dans la récolte des performances de qualité d'œuf est la mesure du diamètre de l'œuf (SLE), et de EW (en grammes). Avec ces deux mesures est calculé l'ESshape, suivant la formule :  $ESshape = (SLE/10)/(EW^{1/3}/10)$ . Ensuite, la couleur de la coquille (cf image 2.4) est mesurée avec un chromamètre Minolta, et 3 paramètres sont enregistrés : la teneur en rouge  $a^*$ , la teneur en jaune  $b^*$  et la clarté  $L^*$  de la coquille. L'ESC est ensuite calculée telle que :  $ESC = 100 - (L^* - a^* - b^*)$ .

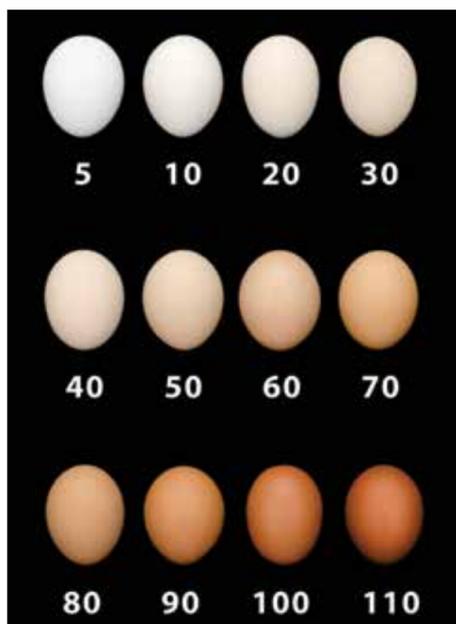


FIGURE 2.4 – Différentes couleurs de coquille d'œuf

Puis, la solidité de la coquille est mesurée par une machine de compression. L'œuf est placé entre 2 plaques se déplaçant à vitesse constante. L'ESS est la force maximale enregistrée avant rupture de la coquille (en Newton, multipliée par 100). Enfin, l'œuf est cassé, et AH est mesuré (en millimètres) en utilisant un tripode. Cette mesure se fait à mi-chemin entre le bord du jaune, et la partie épaisse du blanc d'œuf, au moyen d'un capteur électronique (cf image 2.5).



FIGURE 2.5 – Mesure de la hauteur de l'albumen

### 2.3.2 Choix des effets environnementaux du modèle et filtration des valeurs aberrantes

Pour chaque caractère, les mesures de qualité d'œuf ont été ajusté pour les effets environnementaux en utilisant la procédure GLM de SAS<sup>®</sup>. L'objectif était de vérifier quelle condition environnementale a un effet significatif sur le caractère, et enlever les données aberrantes.

Cet ajustement suivait le modèle linéaire suivant :

$$Y_{ijklm} = \mu + s_i + h_j + b_{jk} + p_{jkl} + \beta_1 W_{ijklm} + \beta_2 A_{ijklm} + E_{ijklm}$$

Avec  $Y_{ijklm}$  la performance,  $\mu$  la moyenne,  $s_i$  l'effet fixé du père,  $h_j$  l'effet fixé du lot,  $b_{jk}$  l'effet fixé de la batterie  $k$  dans le lot  $j$ , et  $p_{jkl}$  la localisation  $l$  de la cage dans la batterie ;  $W_{ijklm}$  est la covariable temps d'attente entre la ponte et la mesure de l'œuf (en jours), et  $A_{ijklm}$  l'âge de la poule (en semaines).  $E_{ijklm}$  est la variable aléatoire résiduelle.

Pour tous les caractères, les effets étaient significatifs, au seuil établi à  $p < 0.05$ . Les données brutes ont ensuite été corrigées en utilisant les estimées de tous ces effets, à l'exception de l'effet paternel. Les valeurs aberrantes, considérées comme telles quand présentant un écart à la moyenne supérieur à 5 écarts-types, ont été supprimées. Le processus complet a été renouvelé jusqu'à ce qu'il n'y ai plus de

valeurs aberrantes, 2 cycles maximum ont été suffisants pour tous les caractères. Ce seuil assez élevé a été choisi afin de n'éliminer que les données résultant d'une erreur lors de la collecte des performances, et conserver le maximum de performances.

Les statistiques des performances retenues après cette filtration sont détaillées dans les tableaux 2.2, 2.3, 2.4 et 2.5.

Caractère	EW	ESC	ESS	AH	ESShape
Nombre de performances	27 915	27 932	25 578	26 447	25 538
Moyenne	59.87	81.19	3785.29	5.48	1.11
Écart-type	4.56	9.12	691.50	1.24	0.02
Min	41	35.36	870	1.1	1.038
Max	82	109.98	6620	11.7	1.175

TABLE 2.2 – Statistiques élémentaires des performances de CC

Caractère	EW	ESC	ESS	AH	ESShape
Nombre de performances	41 307	41 305	38 047	30 430	38 053
Moyenne	62.30	81.50	4109.3	6.21	1.10
Écart-type	4.97	10.55	749.73	1.15	0.01
Min	38.50	29.80	810	1.00	1.03
Max	89.40	114.50	7090	11.80	1.17

TABLE 2.3 – Statistiques élémentaires des performances de 50sem

Caractère	EW	ESC	ESS	AH	ESShape
Nombre de performances	74 976	73 033	65 890	72 107	67 308
Moyenne	60.01	78.37	3671.70	4.70	1.11
Écart-type	4.79	9.82	741.73	1.28	0.02
Min	41.7	30.55	682	0.9	1.021
Max	82.8	109.29	7187	9	1.173

TABLE 2.4 – Statistiques élémentaires des performances en CI

Caractère	EW	ESC	ESS	AH	ESShape
Nombre de performances	37 212	36 702	32 904	25 777	32 892
Moyenne	63.10	80.04	3786.10	5.52	1.10
Écart-type	5.49	11.09	763.85	1.23	0.02
Min	37.70	28.86	640	1.40	1.02
Max	91.60	111.48	6923	11.60	1.17

TABLE 2.5 – Statistiques élémentaires des performances en 70sem

## 2.4 Évaluations génétiques et génomiques

Pour permettre des évaluations multi-caractères sans problèmes liés aux différences d'échelle, toutes les performances ont été centrées et réduites avant évaluation. Les évaluations génétiques ont été faites par BLUP, et le ssGBLUP a été utilisé pour les évaluations génomiques (Misztal et al., 2009; Aguilar et al., 2010). La suite de programmes BLUPF90 a été utilisée pour effectuer ces évaluations (Misztal et al., 2002).

Les matrices de variances-covariances ont été estimées dans un premier temps avec REMLF90 afin d'obtenir un point de départ pour AIREMLF90. Ces matrices ont ensuite été validées avec AIREMLF90 (avec quasiment aucun changement entre les deux programmes). Les erreurs-standards des estimées des paramètres génétiques ont été également estimées avec AIREMLF90. EBVs et GEBVs ont ensuite été estimées en utilisant BLUPF90.

### 2.4.1 Modèle statistique

Le modèle statistique utilisé pour l'évaluation était le même pour tous les caractères, et prenait en compte les effets fixés et covariables décrits dans la section 2.3.2, en plus de l'effet génétique aléatoire de l'animal.

Pour les évaluations des données de CC, chaque œuf mesuré était associé à une cage, sans moyen de retrouver la poule à l'origine de l'œuf. Il y a deux mesures en CC, donc chaque poule a en espérance 2 œufs mesurés, mais il n'a pas été possible de prendre en compte cette répétition des mesures par un effet poule environnemental. Un modèle œuf a été utilisé, avec chaque œuf considéré comme un individu, avec père et mère identifiés.

Les évaluations sur les données croisées (50sem et 70sem), ont également été faites avec un modèle œuf, mais avec un père connu et une mère inconnue. Étant donné le faible nombre (3) de générations étudiées, nous n'avons pas inclus d'effet groupe de mères inconnues pour les analyses. Les performances croisées étaient également obtenues dans un élevage en cages collectives, mais avec un seul œuf mesuré par poule dans chaque jeu de données.

Pour toutes ces évaluations, l'estimation de l'héritabilité des caractères a été estimée par le ratio entre la variance de l'effet animal et la somme de la variance animale et résiduelle.

Les évaluations pour les données en CI utilisent des données répétées par poules, et celles-ci peuvent être reliées directement à la poule pondeuse. Un effet poule environnemental a donc été pris en compte dans le modèle animal utilisé pour l'évaluation.

L'héritabilité de ces évaluations a été estimée par le ratio entre la variance animale et la somme des variances animale, environnementale, et résiduelle. La répétabilité du caractère a été définie comme le ratio entre la somme de la variance animale et environnementale et la somme des variances animale, environnementale, et résiduelle.

## 2.4.2 Populations candidates

Pour comparer les précisions des différentes évaluations, les Estimated Breeding Values (EBVs) obtenus par le BLUP et Genomic Estimated Breeding Values (GEBVs) obtenus par le ssGBLUP de candidats à la sélection ont été comparées. Trois populations candidates issues de la lignée pure ont été constituées (cf figure 2.6).

La première population candidate (C1) comprenait des individus mâles de la génération P1, qui avaient des filles (P2) et des petites-filles (P3) avec des performances. Ce groupe était constitué de 174 coqs en CC, et 78 en CI. La seconde population candidate mâle (C2) a été faite dans le but d'augmenter la taille de la population de référence, qui passe de 3 lots pour C1 à 5 pour C2. Cette population contient 175 coqs en CC et 93 en CI.

Pour C1 et C2, le nombre de coqs disponibles en CI est plus réduit qu'en CC. Il y a une première étape de sélection lors du passage de CC à CI, et seuls les pères avec au moins 8 descendantes avec performances ont été conservés pour l'analyse.

En plus des 2 populations candidates mâles, une population candidate femelle (F) a été constituée en utilisant les poules génotypées de la génération P2. Ce groupe comprend 442 femelles en CC et CI, génotypées et phénotypées. Il y a deux différences majeures entre F et les 2 populations mâles : les femelles ont des performances propres en CI, et ont peu de descendantes (1.6 filles/poule en moyenne). Les femelles n'étant génotypées qu'à partir de P2, il n'était pas possible d'avoir 2 populations candidates femelles.

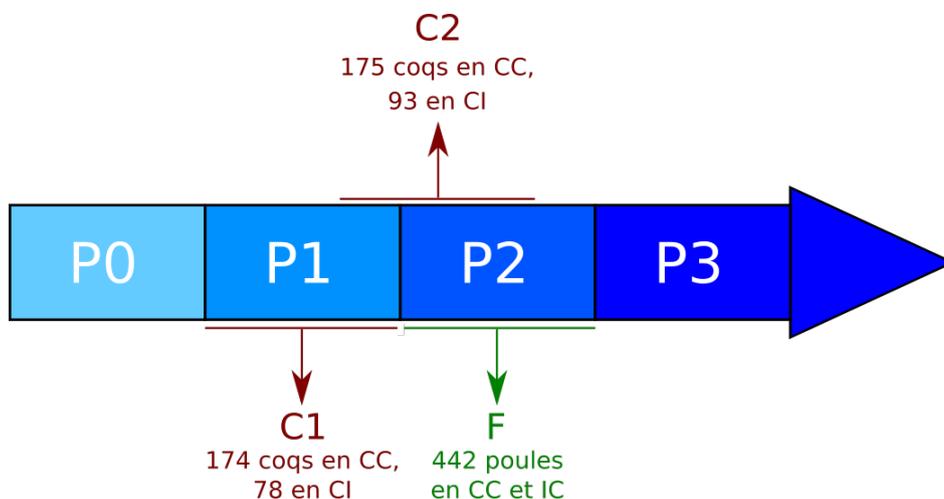


FIGURE 2.6 – Répartition des candidats sur la population de lignée pure

## 2.4.3 Précision de l'évaluation

### Problématique de la référence

Pour estimer la précision de l'évaluation, les EBVs et GEBVs doivent être comparées à la valeur génétique vraie (True Breeding Value, TBV) des candidats. Cependant,

sur données réelles, cette TBV n'est jamais connue. De plus, dans notre cas, elle ne pouvait pas être correctement estimée par une Daughter Yield Deviation (DYD) (VanRaden, 2008), étant donné que nos candidats n'avaient qu'un nombre restreint de descendantes.

Pour pallier à ces limites, la méthode LR (Legarra and Reverter, 2018) a été utilisée pour estimer la précision et les biais de prédiction des évaluations. Cette méthode compare les résultats d'évaluations basées sur des jeux de données partiels avec ceux obtenus sur des jeux de données complets, étant donné que le changement espéré entre évaluations génétiques consécutives a été décrit comme étant fonction de leurs précisions respectives (Reverter et al., 1994).

### Données utilisées

Pour les évaluations sur lignée pure, l'ensemble des données phénotypiques disponibles, de P0 à P3, formait le jeu de données complet. Pour les évaluations utilisant à la fois les performances des lignées pures et celles des poules croisées, ce jeu de données complet comprenait les données disponibles de P0 à P3, et de Cb1 à Cb3. À partir de ces jeux de données complets ont été étudiés 2 cas de jeux de données partiels, construits pour simuler différents types d'évaluation :

- Cas 1 : Évaluation sur ascendance, cas de figure d'une évaluation à la naissance. Le jeu de données est constitué uniquement des performances des ancêtres des candidats, sans information sur leurs performances propres, celles de leurs contemporaines (sœur, demi-sœur, cousines, etc...) ou descendance.
- Cas 2 : Évaluation sur ascendance et collatéraux, cas de figure d'une évaluation à 60 semaines en CC et 80 semaines en CI. Les informations phénotypiques incluent performances des ancêtres, des contemporaines et performances propres dans le cas des femelles. Ce cas correspond au schéma classique de sélection en poule pondeuse.

De plus, pour C1, un troisième cas de figure (Cas 3) a été testé, dans lequel les phénotypes des petites-filles des candidats avaient été retirés. L'objectif était de voir si ces données avaient un impact significatif sur la précision de l'évaluation.

### Méthode LR

La méthode LR repose sur 3 statistiques pour estimer précision et biais d'une évaluation :

- La corrélation entre les (G)EBVs estimées lors de l'évaluation utilisant toute l'information disponible et celles obtenues lors de l'évaluation partielle, pour estimer la précision relative de l'évaluation.
- La différence de moyenne entre les (G)EBVs estimées lors de l'évaluation utilisant toute l'information disponible et celles obtenues lors de l'évaluation partielle, pour estimer le biais.
- La pente de la régression linéaire entre (G)EBVs estimées lors de l'évaluation utilisant toute l'information disponible et celles obtenues lors de l'évaluation

partielle, pour estimer la sur ou sous-dispersion des estimées.

Pour comparer évaluation génétique et génomique, une 4<sup>eme</sup> statistique est utilisée : le ratio entre la précision relative des EBVs et celle des GEBVs. Cette statistique permet de quantifier le gain espéré en termes de précision en passant de l'évaluation génétique à l'évaluation génomique.

## 2.5 Conclusion du matériel et méthodes

Le travail de thèse a donc consisté à évaluer la précision des évaluations génétiques et génomique, sur plusieurs caractères de qualité d'œufs, plusieurs populations candidates, et en utilisant différents volumes de données, représentatifs d'étapes de sélection possibles lors de l'élevage des poules pondeuse. La précision étant estimée en utilisant une évaluation avec le maximum d'information en tant que référence.

Le chapitre suivant présente les résultats obtenus lors de l'étude de la précision de l'évaluation génomique sur les caractères de qualité d'œuf en lignée pure.

# Chapitre 3

## Précision de l'évaluation génomique sur les caractères de qualité d'œuf en lignées pures

### 3.1 Présentation de l'article 1

L'objectif de la première étude était d'estimer, dans différents cas de figure, la précision comparée des évaluations génétique et génomiques dans trois populations candidates (mâles et femelles), en utilisant les performances de qualité d'œuf mesurées sur les poules en lignée pure, à deux âges différents.

Cinq caractères de qualité d'œuf ont été évalués en BLUP et ssGBLUP, et la précision des différentes évaluations a été estimée, ainsi que les biais, pour chaque population candidate. Les évaluations ont utilisé les génotypages et jeux de données réelles en lignée pure décrits dans le chapitre 2.

L'article "Reliability of genomic evaluation for egg quality traits in layers", expose la méthode utilisée et les résultats de ce travail.

RESEARCH ARTICLE

Open Access



# Reliability of genomic evaluation for egg quality traits in layers

David Picard Druet<sup>1</sup> , Amandine Varenne<sup>2</sup>, Florian Herry<sup>1,2</sup>, Frédéric Hérault<sup>1</sup>, Sophie Allais<sup>1</sup>, Thierry Burlot<sup>2</sup> and Pascale Le Roy<sup>1\*</sup>

## Abstract

**Background:** Genomic evaluation, based on the use of thousands of genetic markers in addition to pedigree and phenotype information, has become the standard evaluation methodology in dairy cattle breeding programmes over the past several years. Despite the many differences between dairy cattle breeding and poultry breeding, genomic selection seems very promising for the avian sector, and studies are currently being conducted to optimize avian selection schemes. In this optimization perspective, one of the key parameters is to properly predict the accuracy of genomic evaluation in pure line layers.

**Results:** It was observed that genomic evaluation, whether performed on males or females, always proved more accurate than genetic evaluation. The gain was higher when phenotypic information was narrowed, and an augmentation of the size of the reference population led to an increase in accuracy prediction with regard to genomic evaluation. By taking into account the increase of selection intensity and the decrease of the generation interval induced by genomic selection, the expected annual genetic gain would be higher with ancestry-based genomic evaluation of male candidates than with genetic evaluation based on collaterals. This advantage of genomic selection over genetic selection requires more detailed further study for female candidates.

**Conclusions:** In conclusion, in the population studied, the genomic evaluation of egg quality traits of breeding birds at birth seems to be a promising strategy, at least for the selection of males.

**Keywords:** Laying hens, Egg quality, Genomic evaluation, Accuracy, Single step

## Background

Genomic evaluation, based on thousands of genetic markers in addition to pedigree and phenotype information [1], has become the standard evaluation methodology in dairy cattle breeding programmes over the past years. It has allowed for the improvement of the accuracy of estimated breeding values (EBV) of young bulls and, consequently, their much earlier utilization. Thereby, the generation interval as well as the phenotyping costs have been reduced because of the cessation of the progeny testing of bulls [2]. More recently, avian breeders have started to implement genomic selection in their selection schemes. Indeed, despite the many differences between dairy cattle breeding and poultry breeding, genomic selection is deemed very promising

for the avian sector, especially for layer selection [3–5]. However, to optimize avian selection schemes, one of the key parameters is to properly predict the accuracy of genomic evaluation.

One of the most important factors directly affecting evaluation accuracy is the makeup of the reference population. From the very beginning, genomic evaluation implied that the size of the reference population should not be too small [1, 2, 6]. However, it has also been shown that increasing the size of the reference population does not directly improve evaluation accuracy [7, 8]. Indeed, the close relationship of the reference population with the candidate population is more critical than the size of the reference population. The evaluation becomes more accurate as candidate haplotypes become increasingly well represented in the reference population [7, 9–12]. Aside from the makeup of the reference population, the number

\*Correspondence: [pascale.le-roy@inrae.fr](mailto:pascale.le-roy@inrae.fr)

<sup>1</sup>PEGASE, INRAE, Agrocampus Ouest, 16 Le Clos, 35590 Saint-Gilles, France  
Full list of author information is available at the end of the article



of training generations to use is another important question. Indeed, it has been shown that evaluation accuracy is impacted by the number of training generations used [13, 14], depending on the heritability of the traits.

The present study assesses the relevance of genomic evaluation in comparison with genetic evaluation in order to predict the breeding values of selection candidates for egg quality traits in a pure line of layers. The main objective was to evaluate the expected genetic gain with respect to those traits in order to move from genetic to genomic evaluation.

## Results

Five egg quality traits related to eggshell quality and internal egg quality were studied: egg weight (EW), eggshell colour (ESC), eggshell strength (ESS), albumen height (AH) and eggshell shape index (ESshape).

Two phenotypic datasets were used. The first one, referred to as CC, contains data recorded from young hens (30 to 50 weeks old) bred in collective cages. The second one, referred to as IC, contains data recorded from older hens (60 to 80 weeks old) bred in individual cages.

Three cases scenarios were studied. Case 1: evaluation at birth, Case 2: evaluation at 60 (CC) or 80 (IC) weeks of age, Case 3: evaluation at 140 (CC) or 160 (IC) weeks of age.

### Genetic parameters

Heritabilities remained steady whether REML was carried out with BLUP or with ssGBLUP and independently of the dataset under analysis, e.g., complete or partial (cases 1, 2 and 3). These results were observed whatever the trait or age (CC or IC). Differences ranged from

0% for ESshape in IC to 5% for EW in IC, and values were always higher with ssGBLUP than with BLUP (data not shown). Similar results were obtained for repeatabilities. Genetic correlations were even more stable than heritabilities or repeatabilities. Consequently, variance-covariance matrices were set for the rest of the study. The variance-covariance matrix obtained with genomic evaluation through the use of the complete dataset, i.e., REML carried out with the maximum amount of information, was used to perform subsequent BLUP and ssGBLUP.

Estimates of genetic parameters, obtained using ssGBLUP through the use of the complete dataset, are given in Table 1. EW, ESC and ESshape were highly heritable, while heritability (resp. repeatability) of ESS and AH were more moderate. For all traits, heritability was higher in CC than in IC and was of the same order of magnitude as repeatability in IC. This result may be due to an over-estimation of additive genetic variance, which was partly confounded by common environmental effect in CC, as discussed before (see “Statistical model” section). Genetic correlation between EW and AH was positive and moderate, and showed no significant difference between CC and IC (resp. 0.37 and 0.35) (Table 1). Genetic correlations between ESshape and ESS, or between ESshape and AH, were also positive and moderate, but significantly higher in IC (resp. 0.28 and 0.20 in CC, and 0.45 and 0.41 in IC). Genetic correlation between AH and ESS was slightly lower (0.12 in CC and 0.21 in IC). EW and ESshape were weakly correlated (0.12 in each dataset), while ESC was not correlated with other traits in CC but ESC was weakly correlated with ESS (0.16) and ESshape (-0.16) in IC.

**Table 1** Genetic parameters for studied traits, estimated with ssGBLUP

Traits	EW	ESC	ESS	AH	ESshape
Collective Cages dataset					
EW	0.64 (0.02)	0.02 (0.02)	0.00 (0.02)	0.37 (0.02)	0.12 (0.02)
ESC	0.01 (0.01)	0.59 (0.02)	-0.02 (0.02)	0.03 (0.02)	-0.03 (0.02)
ESS	-0.02 (0.01)	0.05 (0.01)	0.27 (0.02)	0.12 (0.01)	0.28 (0.02)
AH	0.17 (0.01)	0.02 (0.01)	-0.00 (0.01)	0.34 (0.01)	0.20 (0.02)
ESshape	0.02 (0.01)	-0.06 (0.01)	0.18 (0.01)	0.12 (0.01)	0.48 (0.02)
Individual Cages dataset					
EW	0.45/0.69 (0.00)	0.02 (0.01)	0.14 (0.01)	0.35 (0.01)	0.12 (0.01)
ESC	0.02 (0.05)	0.40/0.57 (0.00)	0.16 (0.01)	0.08 (0.01)	-0.16 (0.01)
ESS	0.02 (0.05)	0.08 (0.03)	0.23/0.37 (0.03)	0.21 (0.00)	0.45 (0.01)
AH	0.10 (0.04)	0.02 (0.03)	0.01 (0.00)	0.26/0.36 (0.01)	0.41 (0.01)
ESshape	-0.03 (0.02)	-0.07 (0.01)	0.16 (0.01)	0.10 (0.02)	0.42/0.56 (0.00)

In diagonal: heritability/repeatability; upside diagonal: genetic correlations; downside diagonal: phenotypic correlations. Standard errors are in parentheses  
 Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index

**(G)EBVs relative accuracy for male candidates**

**CC traits**

As expected, relative accuracy estimates (Additional file 1) significantly increased from case 1 to case 3 with the amount of phenotypic information available. In case 1, relative accuracy estimates were not homogeneous and varied depending on the trait. Accuracies for EW and ESshape tended to be low, with relative accuracy between 0.22 (EW genetic C2) and 0.24 (ESshape genetic C2), even though these traits were more heritable than ESS or AH, for which the relative accuracy was approximately 0.32 (AH genetic C2). These differences were less present in case 2 and no longer existed in case 3.

The results of the comparison between genetic evaluation and genomic evaluation are shown in Table 2, which gives the ratio of the relative accuracy of BLUP and ssGBLUP for each case studied. A value of 1 indicates no difference, values below 1 indicate that genomic evaluation is more accurate, while values above 1 indicate that genetic evaluation is more accurate.

Evaluations carried out at birth (case 1) were the ones showing the greatest extent of difference between genetic evaluation and genomic evaluation. Results proved highly trait-dependent and ranged from 1.39 (ESC C1), e.g. a 39% gain in accuracy with genetic evaluation compared with genomic evaluation, to 0.56 (ESshape C2), e.g., a 44% gain in accuracy with genomic evaluation compared with

genetic evaluation. There were strong disparities between C1 results, with a mean of 0.99, and C2 results, with a mean of 0.79. Overall, for C2, genomic evaluation allowed for greater accuracy, which can be explained by the size of the reference population. This advantage of genomic evaluation over genetic evaluation was also trait-dependent.

In the evaluations carried out with respect to ancestry and contemporary relatives (case 2), the differences between genetic evaluation and genomic evaluation were less significant, with values ranging from 1.05 (EW C1) to 0.82 (ESshape C2), and a global mean of 0.93. Differences between C1, with a mean of 0.95, and C2, with a mean of 0.91, were not as significant as they were in case 1, but still existed. As in case 1, the use of ssGBLUP allowed for a relative increase in accuracy.

Evaluations carried out in case 3 showed little difference between BLUP and ssGBLUP, with a mean close to 1.

**IC traits**

As was the case with CC, accuracy estimations were different for each trait (Additional file 2), depending on the evaluation scenario. The evolution of accuracy was also linked to the amount of phenotypic information available.

Here again, the increase in relative accuracy with genomic evaluation, compared to genetic evaluation, was observed in case 1 and in case 2 (Table 3). This increase was more significant for IC traits than it was for CC traits,

**Table 2** Ratio of relative accuracy from BLUP to ssGBLUP with respect to CC traits for candidates

Trait	Candidates	Case 1	Case 2	Case 3
EW	C1	1.05	1.05	1.05
	C2	0.75	0.96	-
	F	0.61	0.92	-
ESC	C1	1.39	1.00	1.01
	C2	0.78	0.92	-
	F	0.88	0.97	-
ESS	C1	0.87	0.89	1.01
	C2	0.99	0.90	-
	F	0.82	0.92	-
AH	C1	0.90	0.93	1.02
	C2	0.87	0.95	-
	F	0.77	0.88	-
ESshape	C1	0.73	0.87	1.01
	C2	0.56	0.82	-
	F	0.83	0.91	-

Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index  
 Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2  
 Case 1: evaluation at birth; Case 2: evaluation at 60 weeks of age; Case 3: evaluation at 140 weeks of age

**Table 3** Ratio of relative accuracy from BLUP to ssGBLUP with respect to IC traits for candidates

Trait	Candidates	Case 1	Case 2	Case 3
EW	C1	0.85	0.89	1.01
	C2	0.65	0.78	-
	F	0.77	1.01	-
ESC	C1	0.84	0.99	1.01
	C2	0.72	0.65	-
	F	0.85	1.01	-
ESS	C1	0.98	0.90	1.01
	C2	0.95	0.80	-
	F	0.69	1.01	-
AH	C1	0.87	0.88	1.01
	C2	0.73	0.77	-
	F	0.79	1.00	-
ESshape	C1	0.70	0.88	1.00
	C2	0.95	0.88	-
	F	0.83	1.01	-

Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index  
 Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2  
 Case 1: evaluation at birth; Case 2: evaluation at 80 weeks of age; Case 3: evaluation at 160 weeks of age

both in case 1 (mean = 0.82) and case 2 (mean = 0.84). The global gain in accuracy observed in genomic evaluations carried out on C2, in comparison to those carried out on C1, was similar to the gain noticed for CC traits.

Evaluations carried out in case 3 showed no differences between BLUP and ssGBLUP, with a mean very close to 1, as was the case for CC.

#### **(G)EBVs biases and dispersion for male candidates**

The bias statistics exhibit an expected value of 0 if evaluation is unbiased. In both CC and IC (Additional files 3 and 4), biases were low and most often negative, indicating an underestimation of (G)EBVs when using partial datasets. The biases increased as the amount of phenotypic information decreased, from approximately 0 in case 3 to -0.11 in case 1. Biases were slightly higher with genomic values, compared to genetic values, in any given trait situation. The differences between traits or between candidate population C1 and candidate population C2 varied, without any clear tendency observed.

Unbiased estimators are supposed to have a regression slope equal to 1. Such unbiased estimators were observed in case 3, with regression coefficients estimated between 0.94 and 1.00 in CC and between 0.97 and 1.04 in IC (R-square values were approximately 0.98). This was the case for both genetic evaluation and genomic evaluation (Additional files 5 and 6).

In both CC and IC, the slopes decreased below 1 every time the amount of phenotypic information decreased. There was no significant difference between genetic evaluation and genomic evaluation: if slopes were closer to 1 when using genetic evaluation on CC traits, it was quite the opposite in the case of IC traits. Conversely, dispersion appeared to be significantly higher in CC than in IC: in the case of IC, slopes remained above 0.7, with few exceptions, even in case 1, while they often decreased below 0.7 in the case of CC (R-square values were between 0.25 and 0.30 in case 1, and between 0.40 and 0.50 in case 2). The slopes were also strongly linked to the evaluated traits, regardless of the candidate population.

#### **(G)EBVs relative accuracy for female candidates**

##### **CC traits**

As was the case for males, accuracy estimations of (G)EBVs for females were not homogeneous, depending on the trait (Additional file 1): some traits were evaluated with greater accuracy than others, and accuracy evolution was not the same for all traits, depending on the scenario. However, these differences were not the same as those noticed with males. This could be a consequence of different genetic determinism of traits in females, for example, the impact of Quantitative Trait Loci carried by sexual chromosomes (females are ZW, and males are ZZ). Relative accuracy of (G)EBVs was nonetheless generally higher

for females in comparison with males, especially in case 2 where females had their performances taken into account.

Furthermore, genomic values were always more accurate than genetic values and, as with males, the gain increased when the amount of phenotypic information was low (Table 2). Indeed, evaluations carried out at birth (case 1), showed a significant increase in accuracy with ssGBLUP evaluation compared to BLUP evaluation. The mean of ratios was close to 0.78, i.e. a 22% gain in accuracy with genomic evaluation compared with genetic evaluation. Regarding evaluations carried out in case 2, where the performances of the females were taken into account, this value was only 0.92, i.e., an 8% gain in accuracy.

##### **IC traits**

As was the case for CC, correlations were moderate (Additional file 2) and varied depending on the trait in case 1, while they were always very high (with a minimum of 0.93) in case 2, where the performances of the females were taken into consideration.

In case 1, the increase in accuracy noticed with ssGBLUP evaluation in comparison with BLUP evaluation (Table 3) was of the same order of magnitude as for CC, with a mean of ratios close to 0.79. In case 2, this value was between 1.00 and 1.01, depending on the trait.

#### **(G)EBVs biases and dispersion for female candidates**

As was the case for males, and both in CC and IC (Additional files 3 and 4), biases were low and often negative (with an exception in case 1 EW and ESC). This indicated an underestimation of (G)EBVs when evaluating partial datasets. A similar increase in biases was observed for females when the amount of phenotypic information decreased. Here again, no clear relationship could be noticed between traits and biases, nor between the type of evaluation carried out (genetic or genomic) and biases, contrary to what was observed with males.

Regarding the regression coefficients in both CC and IC (Additional file 5 and 6: Tables), the results were similar to those observed using male candidates (R-square values were between 0.25 and 0.30 in case 1 and approximately 0.95 in case 2). Except for ESshape in case 1, the regression coefficients decreased below 1 every time the amount of phenotypic information decreased, i.e., results in case 2 were less biased than those in case 1.

As was the case for males, there was no significant difference between genetic evaluation and genomic evaluation, and dispersion seemed higher for CC than for IC. Once again, slopes were strongly linked to the trait evaluated in any given case.

## Discussion

### Genetic parameters

Estimates of heritability, repeatability and genetic correlations (Table 1) were in accordance with the literature [4, 15]. Moderate to high heritability coefficients enable significant expected genetic gains through selection with respect to the five egg quality traits. Selection carried out in order to increase eggshell strength should lead to an increase in albumen height, which is favourable to a combined selection of these traits. Increasing eggshell strength should also lead to an increase in egg short length at a given weight, which is unfavourable: e.g., the egg would be less ovoid. Finally, genetic correlations of eggshell colour to other traits are negligible, with the exception of eggshell strength in IC.

### Relevance of genomic evaluation for male candidates

In case 1 and in case 2, the results generally highlighted a greater accuracy of the evaluation of male candidates with ssGBLUP than with BLUP, in any given scenario, and particularly regarding IC traits. The difference between CC results and IC results can be explained by the nature of the data: indeed, IC data, referring to the hen itself and not just to the full sisters, allowed for the construction of a more accurate evaluation model. Otherwise, the results obtained in case 3 showed that the information about the grand-daughters had little impact on the evaluation. The fact of not using the performances of the grand-daughters does not seem to have any direct negative impact on evaluation accuracy.

The results observed in scenarios using C2 candidates tend to confirm those obtained using C1 candidates, and sometimes even amplified them. The difference between C1 results and C2 results could be explained by the increase in the size of the reference population, which went from 3 batches for C1 to 5 batches for C2, still with the same number of candidates. This increase in the size of the reference population had an impact on evaluation, with more candidate haplotypes represented in the reference population, as shown by Rabier et al. [9].

### Relevance of genomic evaluation for female candidates

The differences between traits, and the relationship between phenotypic information and accuracy noticed with males were also observed with females. Genomic evaluation provided more accurate evaluations than genetic evaluation, except for case 2, where IC traits were used. GEBV accuracy was capped at approximately 0.70 in CC, while it approached 1 in IC. This difference was because in CC, phenotypes are not related to a single bird, but to a cage of full sisters, which means that instead of being performed on the female, the evaluation was performed on its family.

Evaluations performed on a candidate population of females in IC (case 2) showed the relative accuracy of (G)EBV, which came close to 1. This result was because the performances of the females were taken into account for the evaluation: the addition of genomic information did not increase the gain in accuracy. Furthermore, as opposed to male candidates, females had very few daughters with performances: the lack of information about the performances of the daughters had little impact on the estimation of the value of the females.

### Towards the use of genomic evaluation of egg quality traits

In both CC and IC, the reliability of genetic evaluation or of genomic evaluation varied greatly depending on the trait. This heterogeneity in the impact on the evaluation can be explained by the differences found in the genetic architecture of the traits [16]. Indeed, when a trait is influenced by few QTLs with large effects, it would be poorly predicted with ssGBLUP, which assumes a common variance for all SNP effects. Moreover, a trait may be more or less influenced by non-additive effects, i.e., dominance and epistasis, which were not taken into account in the present study.

With regard to the comparison between genomic evaluation and genetic evaluation, with the same amount of phenotypic information used, genomic evaluation proved more accurate than genetic evaluation most of the time. As expected, the increase in accuracy of genomic evaluation was greater when phenotypic information was restricted.

Regarding the size of the reference population, it was observed that adding a generation, from C1 to C2, had an effect on the evaluation, as Weng et al. (2016) [13] showed. Augmentation of the reference population from 1 to 2 generations increased evaluation accuracy, especially when there was little phenotypic information available.

These results are very interesting for the poultry industry. Indeed, genomic evaluation and selection of males at birth would allow for their use at their sexual maturity, i.e., 6 months of age. For CC traits, there would be a significant loss in evaluation accuracy, although the loss would remain acceptable for IC. Depending on the weight of each trait in the breeding goal, this strategy would allow for a significant genetic gain for the global objective through an increment in selection pressure, i.e., it would be feasible to increase the number of male candidates for selection from 200 to 2000 and to reduce the generation interval from 18 months to 6 months.

The genomic evaluation and selection of females at birth would also allow for their use at their sexual maturity, i.e., 6 months of age. However, our results highlighted the fact that, for females, switching from selection at 18 months of age to selection at birth would result in a significant loss

in evaluation accuracy. Moreover, in the case of females, selection pressure should not be significantly increased.

## Conclusions

Our results indicate that it seems advantageous to move from genetic selection at 18 months of age to genomic selection at birth (2 weeks in practice: the time to obtain genotypes and to calculate GEBVs), at least with respect to egg quality traits and as far as males are concerned. This strategy must be studied in greater detail for females in order to assess whether the implementation of genomic evaluation at birth would be an interesting option.

## Methods

### Animals

For the purpose of this study, we have used the data of a pure line of Rhode Island layers selected by the breeding company Novogen (Plédran, France). The hens were hatched in 12 batches, born between 2008 and 2015, which corresponds to four generations (G0 to G3, Fig. 1), with three successive hatches per generation spaced 30 weeks apart. The genealogical information of all of the birds was recorded in the pedigree file, which concerned 2273 breeders: 514 sires and 1759 dams.

In each hatch, chicks were bred in a brooding area until the age of 18 weeks, and then transferred to collective cages of five full sisters for females (2997 collective cages) and to individual cages for males (200 males out of 2000 chicks were kept for the selection). Egg quality was recorded twice during this period, at 30 and 50 weeks of age. This step allowed for the makeup of our first phenotypic dataset, referred to as CC for collective cages, with a total of 14985 hens and 27915 eggs measured.

Then, at 60 weeks of age, a genetic evaluation was performed as a first selection, and 150 males and 650 hens were transferred to individual cages until the end of their

careers at the age of 80 weeks. Egg quality was measured on a weekly basis. A total of 7982 hens, with 74976 performances, were concerned. This step allowed for the makeup of our second phenotypic dataset, referred to as IC for individual cages.

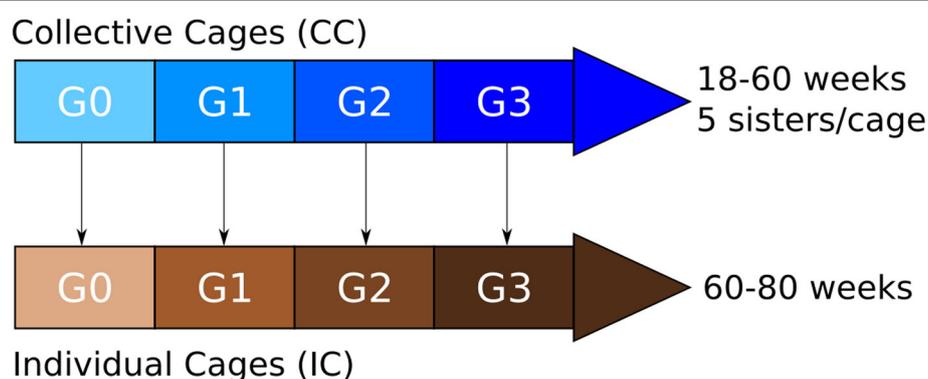
At the end of their careers, animals were slaughtered in poultry slaughterhouse.

### Genotypes

In this population, 2374 birds were genotyped using the 600K Affymetrix<sup>®</sup> Axiom<sup>®</sup> HD genotyping array [17]. Blood samples were collected from the brachial veins of the animals and DNA was extracted. For the first two generations, all male candidates were genotyped by Ark-Genomics (Edinburgh, UK) during the research project UtopIGe [16]. From the G2 generation onward, male and female reproducers were genotyped at the high-throughput genotyping platform Gentyane (Clermont-Ferrand, France).

Each animal was genotyped for 580961 SNP markers. According to the fifth annotation release of the *Gallus gallus* genome [18], these SNPs were distributed over macro-chromosomes (1 to 5), intermediate chromosomes (6 to 10), micro-chromosomes (11 to 28 and 33), one linkage group (LGE64), two sexual chromosomes Z and W, and a group of 3724 SNPs with unknown locations.

Genotypes were filtered through four successive steps: individuals with a call rate <95% were removed (0 animals excluded), SNPs with a MAF <0.05 were removed (258772 SNPs), SNPs with a call rate <95% were removed (7549 SNPs), and SNPs whose genotype frequencies deviated significantly from the Hardy-Weinberg equilibrium ( $P < 10^{-4}$ ) were removed (12538 SNPs). Animals showing pedigree incompatibilities were also removed (12 individuals excluded). Thus, 302102 SNPs and a total of 1214 genotyped males and 1148 genotyped females were retained for the study.



**Fig. 1** Population structure

### Traits

In this paper, traits are named according to Animal Trait Ontology for Livestock [19]. Five egg quality traits related to eggshell quality and internal egg quality were studied: egg weight (EW), eggshell colour (ESC), eggshell strength (ESS), albumen height (AH) and eggshell shape index (ESshape). Summary statistics of the traits are provided in Table 4.

### Trait measurements

At 30 and 50 weeks of age for CC, and once a week for IC, the eggs produced on the farm were collected and quality traits were measured by the company Zootests (Ploufragan, France). The first step consisted in measuring egg short length (SLE in mm) and EW (in g), before calculating ESshape as follows:  $ESshape = \frac{SLE/10}{(EW/10)^{1/3}}$ . Then, eggshell colour was measured using a Minolta chromameter (Nieuwegein, The Netherlands) and three traits were recorded: redness of eggshell  $a^*$ , yellowness of eggshell  $b^*$  and lightness of eggshell  $L^*$ . Eggshell colour was then calculated as follows:  $ESC = 100 - (L^* - a^* - b^*)$ . Third, shell strength was measured using a compression instrument to evaluate the static stiffness of the shell. Eggs were compressed between two flat plates moving at constant speed. ESS is the maximum force recorded before eggshell fracture (in N, multiplied by 100). Finally, the egg was broken and AH (in mm) was measured using a tripod.

Equations for ESC and ESshape were those used by Novogen for egg quality control.

### Choice of environmental effects to introduce into the model and elimination of outliers

For each trait, egg measurements were adjusted for environmental effects using the SAS<sup>®</sup> 9.4 GLM procedure. The

objective was to check which environmental conditions had significant effects on the traits and to remove outliers. The environmental effects tested were the hatch, the cage location in the poultry house during hatching, i.e., the battery and the cage location along the battery, the waiting time between sampling and egg measurement and the age of the hen.

All of these environmental effects were below the significance level ( $P < 0.05$ ), which means that they could be retained in the model of analysis. Raw data were then adjusted using the estimates of all of these effects, and outliers were deleted. The values presenting a deviation from the sire family mean of greater than 5 phenotypic standard deviations (0.4% of the CC performances and 0.01% of the IC performances) were considered as outliers.

Ultimately, a total of more than 25500 records for CC and more than 65800 records for IC were retained (Table 4). All distributions of adjusted phenotypes were symmetrical (data not shown).

### Genetic and genomic evaluations

Performances were centered (by subtracting the general mean) and standardized (by dividing by the standard deviation) before evaluation. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and single-step GBLUP (GEBV) methodologies [20, 21]. BLUP was used as the reference method for genetic evaluations. ssGBLUP was used because a majority of our phenotypes came from non-genotyped birds (13837 hens were phenotyped, but not genotyped, of 14985 total hens), and this method allows those data to be accounted for in our evaluation. Before EBVs and GEBVs estimation, variance-covariance matrices were estimated using REML for both BLUP and ssGBLUP.

**Table 4** Summary statistics on phenotypic data

Traits	EW (g)	ESC (withoutunits)	ESS (N * 100)	AH (mm)	ESshape (mm.g <sup>-1</sup> )
Collective Cages dataset					
Number of records	27915	27932	25578	26447	25538
Mean	59.9	81.2	3785.3	5.5	1.1
Standard deviation	4.56	9.12	691.50	1.24	0.02
Min	41.00	35.36	870.00	1.10	1.04
Max	82.00	109.98	6620.00	11.70	1.17
Individual Cages dataset					
Number of records	74976	73033	65890	72107	67308
Mean	60.0	78.4	3671.7	4.7	1.1
Standard deviation	4.79	9.82	741.73	1.28	0.02
Min	41.70	30.55	682.00	0.90	1.02
Max	82.80	109.29	7187.00	9.00	1.17

Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index

To perform those evaluations, the BLUPF90 family of programmes [22] was used. Variance-covariance matrices were first estimated using REMLF90 to obtain a good starting point for AIREMLF90. AIREMLF90 was then used to validate these variance-covariance components (the components have changed little between the two programmes) and estimate the standard errors of genetic parameter estimates. EBVs and GEBVs were then estimated using BLUPF90 [22].

**Statistical model**

The statistical model was the same for all of the traits and took into account the fixed and covariable environmental effects described before, plus the random genetic effects of the animal. For CC evaluations, each egg measured was associated with one cage of five full sisters, without knowing which hen laid which egg. As the measurements were repeated twice in CC, each hen had the expectation of two measured eggs, but it was not possible to take into account the common environmental effects of the hen. An “egg-animal” model, without common environmental effects, provided slightly more accurate (G)EBV than a sire model (data not shown). The inclusion of a “cage” random effect was also tested (instead of the fixed effects, which described the geolocation in the poultry building). The estimation of this variance-covariance component was very low, but the computation time was significantly higher with this model. The “egg-animal” model was then retained, and the heritability estimation was calculated as the ratio between the animal variance and the sum of animal and residual variances. Conversely, for IC evaluations, several measurements were available for each hen, and random common environmental effects of the hen were taken into account in the model. The heritability coefficient was therefore estimated as the ratio between

the animal variance and the sum of animal, common environmental and residual variances, and the repeatability coefficient was considered as the ratio between the sum of animal and common environmental variances and the sum of animal, common environmental and residual variances.

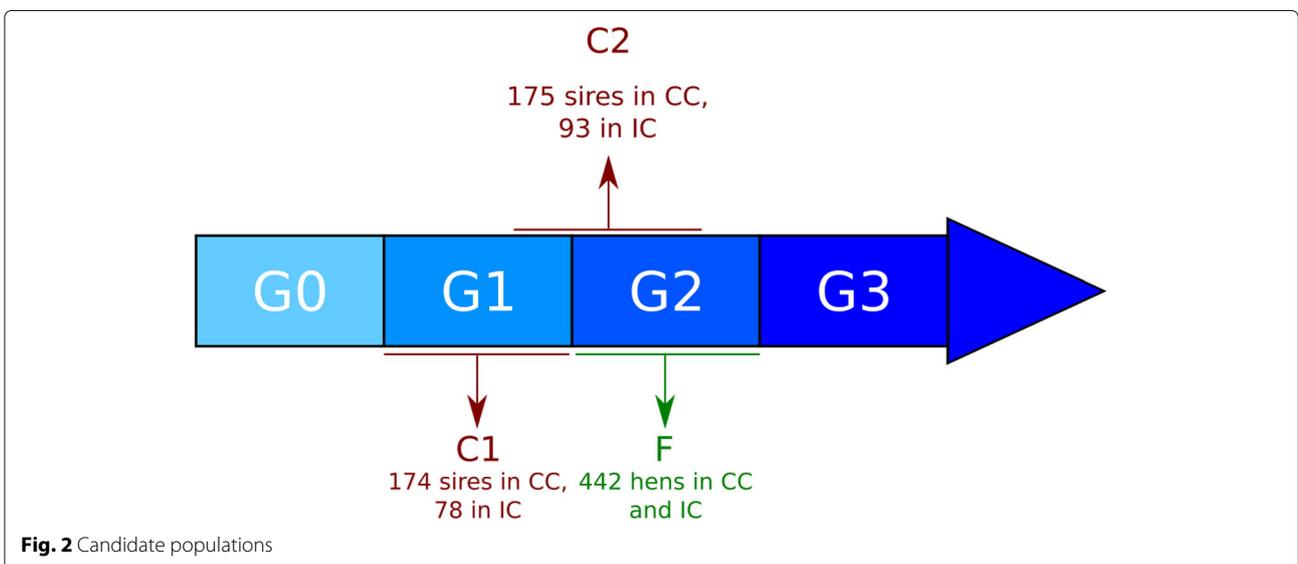
**Candidate populations**

To assess the relevance of genetic evaluation and of genomic evaluation, the statistical properties (described below) of estimated breeding values, EBVs and GEBVs, of selection candidates were used.

A male candidate must have a genotype and have progeny tested. Two male candidate populations were considered: C1 and C2 (Fig. 2). In both, the number of sires in IC was smaller than in CC. This is due to the selection being carried out before moving them from CC to IC and to the fact that in IC, only sires having at least 8 daughters with performances were used.

The first male candidate population (C1) comprised birds from generation G1 that had daughters (G2) and grand-daughters (G3) with performances. This group was made up of 174 sires in CC and 78 sires in IC. A second male candidate population (C2) was considered in order to increase the reference population, therefore changing from three hatches for C1 to five hatches for C2, i.e., it comprised birds from the last hatch of G1 and the two first hatches of G2 (which was the limit to have male progeny with performances), which led to a total of 175 sires in CC and 93 in IC.

In addition to these male populations, a female candidate population (F) was formed using genotyped hens from G2 (Fig. 2). This group comprised 442 females in CC which were then moved to IC. The difference between this population and the male population was that IC females



**Fig. 2** Candidate populations

had performances available for evaluation and had few daughters (between 0 and 19 daughters/hen; 1.6 daughters/hen on average). As opposed to males, females were genotyped starting from G2, which means that it was not possible to have two female candidate populations.

#### **Reliability of prediction**

To assess the reliability of genetic evaluation and genomic evaluation, the estimated breeding values, EBVs and GEBVs, had to be compared to the true breeding values (TBVs) of candidates. However, TBV is never known when working with a real dataset. Moreover, it could not be approximated using Daughter Yield Deviation (DYD) [23], since our candidates had very few offspring. Therefore, to estimate the accuracy and bias of prediction of our evaluations, the LR method [24] was used.

The LR method uses correlation between complete and partial datasets to estimate the accuracy, since the amount of change expected in consecutive genetic evaluations was described as a function of their respective accuracies [25]. This method completes cross-validation approaches with semiparametric elements, based on the mixed model equations, to estimate the “population” accuracy. Population accuracy is relevant to compare the predictive ability of models and to maximize genetic progress.

The LR method relies on three statistics to estimate the accuracy and biases of an evaluation:

- The correlation between (G)EBVs from complete and partial evaluations to estimate relative accuracy
- The difference of means between (G)EBVs from complete and partial evaluations to estimate biases
- The slope of the linear regression of (G)EBVs from complete evaluation on (G)EBVs from partial evaluation to estimate the over- or underdispersion of estimates

To compare genetic evaluation and genomic evaluation using the same amount of data, a fourth statistic was used: the ratio between the relative accuracy of EBVs and the relative accuracy of GEBVs. This statistic allows quantification of the increase in accuracy expected when moving from genetic evaluation to genomic evaluation.

#### **Application to data**

All available pedigrees, phenotypes, and genotypes for GEBV estimation, from G0 to G3, constitute the complete dataset. Two cases of partial datasets were studied based on the amount of phenotypic information available when the evaluation was carried out:

- Case 1: The evaluation was carried out at birth of the candidates without considering the performances of their contemporary relatives nor the performances of the candidates, in the case of females. The evaluated

population was limited to the candidates and their ancestors: all individuals had pedigree information; female ancestors had phenotypes; male ancestors and male (or female) candidates had genotypes for GEBV estimation.

- Case 2: The evaluation was carried out at 60 weeks of age for CC and at 80 weeks of age for IC, without considering the performances of the progeny of candidates. This case corresponds to the scheme classically used in layer selection. The evaluated population included the candidates, their contemporary relatives and their ancestors: all individuals had pedigree information; female ancestors, female contemporary relatives or candidates had phenotypes; male ancestors had genotypes for GEBV estimation; candidates, male or female, had genotypes for GEBV estimation.

Moreover, for C1 male candidates, the potential gain was also assessed taking into account the performances of their grand-daughters. In that case, the pedigrees and phenotypes of G3 hens were removed from the complete dataset to obtain a partial dataset (case 3).

The differences in significance between relative accuracies, e.g., correlations as defined above, were assessed using the Hotelling-Williams test [26]. This test is used to compare two dependent correlations that share a common variable. The null hypothesis means that the two compared correlations are equal. The test statistics under the null hypothesis follow the Student's t-distribution at  $n-3$  degrees of freedom, with  $n$  being the number of observations. Observed correlations were compared two-by-two for EBVs and GEBVs at a significance threshold of 5%.

In layers, candidates are compared within a hatch: comparisons are never made between individuals from different generations for the purposes of selection. However, biases and dispersion statistics were presented here to illustrate the evolution of quality of evaluation according to cases 1 to 3.

#### **Supplementary information**

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12863-020-0820-2>.

**Additional file 1:** Estimates of (G)EBVs relative accuracy in candidate populations with respect to cC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 60 weeks of age; Case 3: evaluation at 140 weeks of age. In parentheses are groups determined by the Hotelling-Williams test at a confidence level of 95%. Groups a, b and c are used when EBVs correlations are compared to other EBVs correlations; d, e, f are used when GEBVs correlations are compared to other GEBVs.

**Additional file 2:** Estimates of (G)EBVs relative accuracy in candidate populations with respect to iC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 80 weeks of age; Case 3: evaluation at 160 weeks of age. In parentheses are groups determined by the Hotelling-Williams test at a confidence level of 95%. Groups a, b and c are used when EBVs correlations are compared to other EBVs correlations; d, e, f are used when GEBVs correlations are compared to other GEBVs.

**Additional file 3:** Bias of (G)EBVs in candidate populations with respect to cC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 60 weeks of age; Case 3: evaluation at 140 weeks of age.

**Additional file 4:** Bias of (G)EBVs in candidate populations with respect to iC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 80 weeks of age; Case 3: evaluation at 160 weeks of age.

**Additional file 5:** Slopes of regression for (G)EBVs in candidate populations with respect to cC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 60 weeks of age; Case 3: evaluation at 140 weeks of age.

**Additional file 6:** Slopes of regression for (G)EBVs in candidate populations with respect to iC traits. Traits: EW: Egg Weight; ESC: Eggshell Colour; ESS: Eggshell Strength; AH: Albumen Height; ESshape: Eggshell shape index. Candidates: C1: male candidates from G1; C2: male candidates from G1 and G2; F: female candidates from G2. Multi-trait evaluations were performed for the five traits using BLUP (EBV) and ssGBLUP (GEBV) methodologies. Case 1: evaluation at birth; Case 2: evaluation at 80 weeks of age; Case 3: evaluation at 160 weeks of age.

#### Acknowledgements

Not applicable.

#### Authors' contributions

DPD performed the statistical analyses and drafted the manuscript. FIH and FrH filtered the genotype data. AVa and TB supervised animal management and production. PLR supervised the analyses and participated in drafting the manuscript. PLR and TB conceived the study. All authors contributed to the ideas and methods. All authors read and approved the final manuscript.

#### Funding

This research project was supported by the French national research agency ANR, within the framework of project ANR-10-GENOM\_BTV-015 UtOplGe. DPD is a PhD fellow supported by the French National Agronomic Research Agency (INRAE), with the SelGen metaprogramme, and the Bretagne region. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

The datasets used and/or analysed throughout the present study are available from the corresponding author upon reasonable request.

#### Funding

This research project was supported by the French national research agency ANR, within the framework of project ANR-10-GENOM\_BTV-015 UtOplGe. DPD is a PhD fellow supported by the French National Agronomic Research Agency (INRAE), with the SelGen metaprogramme, and the Bretagne region. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

The datasets used and/or analysed throughout the present study are available from the corresponding author upon reasonable request.

#### Availability of data and materials

The datasets used and/or analysed throughout the present study are available from the corresponding author upon reasonable request.

#### Ethics approval and consent to participate

All information used in this study was collected from hens raised in the context of layer selection. These animals and the scientific investigations described herein are therefore not to be considered as experimental animals per se, as defined in EU directive 2010/63 and subsequent national application texts. Consequently, we did not seek ethical review or approval of this study regarding the use of experimental animals. All animals were reared on the Novogen nucleus herd in compliance with national regulations pertaining to livestock production and according to procedures approved by the French Veterinary Services.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>PEGASE, INRAE, Agrocampus Ouest, 16 Le Clos, 35590 Saint-Gilles, France. <sup>2</sup>NOVOGEN, 5, rue des Compagnons, 22960 Plédran, France.

Received: 19 July 2019 Accepted: 31 January 2020

Published online: 11 February 2020

#### References

1. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819–29. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461589/>.
2. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92(2):433–443.
3. Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Implementation of genomic selection in the poultry industry. *ResearchGate*. 2016;6(1):23. [https://www.researchgate.net/publication/289494029\\_Implementation\\_of\\_genomic\\_selection\\_in\\_the\\_poultry\\_industry](https://www.researchgate.net/publication/289494029_Implementation_of_genomic_selection_in_the_poultry_industry).
4. Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet Sel Evol*. 2015;47:59. <http://dx.doi.org/10.1186/s12711-015-0133-5>.
5. Liu T, Qu H, Luo C, Shu D, Wang J, Lund MS, et al. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. *BMC Genet*. 2014;15:110. <https://doi.org/10.1186/s12863-014-0110-y>.
6. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*. 2008;3(10):e3395.
7. Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G, Reents R. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genet Sel Evol GSE*. 2011;43(1):19. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3107172/>.
8. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci*. 2012;95(7):4114–29.

9. Rabier CE, Barre P, Asp T, Charmet G, Mangin B. On the Accuracy of Genomic Selection. *PLoS ONE*. 2016;11(6):e0156086. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0156086>.
10. Elsen JM. Approximated prediction of genomic selection accuracy when reference and candidate populations are related. *Genet Sel Evol GSE*. 2016;48:22. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4778372/>. Accessed 3 Mar 2016.
11. Clark SA, Hickey JM, Daetwyler HD, van der Werf JH. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet Sel Evol GSE*. 2012;44(1):4. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3299588/>.
12. Habier D, Fernando RL, Dekkers JCM. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics*. 2007;177(4):2389–97. <http://www.genetics.org/content/177/4/2389>.
13. Weng Z, Wolc A, Shen X, Fernando RL, Dekkers JCM, Arango J, et al. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genet Sel Evol*. 2016;48:22. <http://dx.doi.org/10.1186/s12711-016-0198-9>.
14. Lourenco DaL, Misztal I, Tsuruta S, Aguilar I, Lawlor TJ, Forni S, et al. Are evaluations on young genotyped animals benefiting from the past generations? *J Dairy Sci*. 2014;97(6):3930–42.
15. Beaumont C, Calenge F, Chapuis H, Fablet J, Minvielle F, Tixier-Boichard M. Génétique de la qualité de l'œuf. *Inra Prod Anim*. 2010;23(2):123–32. <http://www6.inra.fr/productions-animales/2010-Volume-23/Numero-2-2010/Genetique-de-la-qualite-de-l-œuf>.
16. Romé H, Varenne A, Hérault F, Chapuis H, Alleno C, Dehais P, et al. GWAS analyses reveal QTL in egg layers that differ in response to diet differences. *Genet Sel Evol*. 2015;47:83. <https://doi.org/10.1186/s12711-015-0160-2>.
17. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*. 2013;14(1):59. <https://doi.org/10.1186/1471-2164-14-59>.
18. Warren WC, Hillier LW, Tomlinson C, Minx P, Kremitzki M, Graves T, et al. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda, Md)*. 2017;7(1):109–17.
19. Livestock Ontologies – ATOL – EOL – AHOL. <http://www.atol-ontology.com/>.
20. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci*. 2010;93(2):743–52.
21. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci*. 2009;92(9):4648–55.
22. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee D. BLUPF90 and related programs. In: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*. vol. 28. World Congress on Genetics Applied to Livestock production; 2002. p. 743.
23. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23.
24. Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol GSE*. 2018;50(1):53.
25. Reverter A, Golden BL, Bourdon RM, Brinks JS. Technical note: detection of bias in genetic predictions. *J Anim Sci*. 1994;72(1):34–7. <https://academic.oup.com/jas/article/72/1/34-37/4632556>.
26. Van Sickle J. Analyzing correlations between stream and watershed attributes. *J Am Water Resour Assoc*. 2003;39(3):717–26.

## Publisher's Note

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

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

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

At BMC, research is always in progress.

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



## 3.2 Conclusions de l'article 1

Cette étude a permis de conclure sur plusieurs aspects concernant l'utilisation de la sélection génomique sur les caractères de qualité d'œuf chez la poule pondeuse.

La précision de l'évaluation, génétique ou génomique, varie fortement en fonction du caractère. Cela peut être une conséquence d'effets de dominance et d'épistasie, qui ne sont pas pris en compte dans le modèle additif du BLUP. Il est également possible que ce soit un effet des différences d'architecture génétique des caractères. Cette hypothèse a été étudiée dans le chapitre 5.

L'évaluation génomique a été estimée comme généralement plus précise que l'évaluation génétique, et d'autant plus dans les cas d'évaluation avec une faible quantité d'informations.

À la lumière des gains de précision observés, il paraît avantageux de passer d'une sélection génétique sur collatéraux à la sélection génomique sur ascendance pour les mâles. La perte de précision est relativement limitée, cela permet d'augmenter fortement le nombre de candidats mâles à la sélection (2 000 au lieu de 200), et de réduire fortement l'intervalle de génération. Cela devrait permettre au final un meilleur progrès génétique des populations sélectionnées.

Pour les femelles, la conclusion est moins claire. Les performances propres apportent énormément d'information, et s'en passer lors de l'évaluation génomique à la naissance produit une baisse de la précision. De plus, les femelles ne bénéficieraient pas de plus de candidates, uniquement de la diminution de l'intervalle de génération. Cette stratégie nécessite d'être étudiée plus en détails pour pouvoir déterminer si l'évaluation génomique à la naissance est une option viable pour la voie femelle.

Le chapitre suivant porte sur l'impact de l'usage des performances des descendantes croisées des individus de lignée pure sur la précision de l'évaluation.

# Chapitre 4

## Apport des informations des descendantes croisées sur la précision de l'évaluation

### 4.1 Présentation de l'article 2

Les objectifs de cette seconde étude étaient de déterminer la corrélation entre caractères mesurés en lignée pure et ceux mesurés dans la descendance croisée, et de l'effet de l'utilisation conjointe de ces données sur la précision de l'évaluation sur une population candidate mâle.

Trois caractères de qualité d'œuf ont été évalués en BLUP et ssGBLUP, en utilisant des jeux de données phénotypiques utilisant soit uniquement les performances mesurées en lignée pure, soit conjointement aux performances en lignée pure celles de leur descendance croisée. La précision et les biais des différentes évaluations ont été évalués, afin de déterminer si l'utilisation conjointe des données de lignée pure et de leur descendance croisée permet un gain en précision de l'évaluation.

# Relevance of using crossbred performances in genomic evaluation for egg quality traits in layers

David Picard Druet<sup>1\*</sup>, Amandine Varenne<sup>2</sup>, Frédéric Héroult<sup>1</sup>, Florian Herry<sup>1,2</sup>, Thierry Burlot<sup>2</sup>, Sophie Allais<sup>1</sup> and Pascale Le Roy<sup>1</sup>

Correspondence:

[david.picard-druet@inra.fr](mailto:david.picard-druet@inra.fr)

PEGASE, INRA, Agrocampus

Quest, 16 Le Clos, 35590

Saint-Gilles, France

Full list of author information is

available at the end of the article

## Abstract

### Background:

In poultry selection programs, pyramidal schemes are carried out to produce crossbred animal adapted both to the targeted market and to the consumers needs. Generally, each pure line is selected independently of the others, based on purebred performances only. But performances of purebred (PB) individuals could be, due non additive effects and genotype by environment interactions, an imperfect predictor of performances of crossbred (CB) individuals. Furthermore, performances from crossbred offspring are collected, but not used. A way to valorise these data would be to include them directly in the purebred breeders evaluation.

### Methods:

Both genetic and genomic evaluation were performed on a purebred candidate population of 174 sires, using different sizes of phenotypic records on three egg quality traits, with performances from purebred and their crossbred offspring, and at two different ages. The methodologies used were BLUP & ssGBLUP, and variance-covariance matrices were estimated through REML. To estimate evaluation accuracy, the LR method was implemented. Four statistics were used to assess the relative accuracy of the estimated breeding values of candidates, their biases and dispersion, as well as the difference between purebred and mixed performances evaluation.

### Results:

It was observed that adding crossbred performances in purebred evaluation leads to more accurate evaluation in most of the case scenario. In addition, the genetic correlation found between PB and CB performances allow to confirm that these should be evaluated separately, and considered as 2 different traits.

### Conclusions:

Adding performances from CB offspring to evaluate purebred breeders seems very promising to improve genetic gain on purebred populations, especially combined with genomic selection.

**Keywords:** laying hens; egg quality; genomic evaluation; accuracy; crossbred

1

2

### 3 Introduction

4 In poultry selection programs, crossbreeding is commonly used to produce animals  
5 adapted both to the targeted market and to the consumers needs. As in swine  
6 selection programs, pyramidal schemes are traditionally used to produce crossbred  
7 animals issued from several pure lines. Each pure line is selected independently of  
8 the others and its genetic evaluation is carried out using purebred performances.  
9 One major drawback is that performances of purebred individuals (PB) may prove  
10 to be an imperfect predictor of the performances of crossbred individuals (CB).  
11 Indeed, because of the presence of non additive genetic effects, e.g. dominance  
12 and epistasis, and genotype by environment interactions [1], the genetic correlation  
13 between PB and CB, for a given trait, can be significantly different from 1, as it  
14 was shown in pork [2, 3, 4] and chicken [5, 6].

15  
16 To overcome these limitations, the join use of both PB and CB information  
17 in a Combined Crossbred and Purebred Selection (CCPS) was proposed by Wei  
18 [7]. Recent studies confirmed the interest of this strategy in optimizing genetic  
19 gain [8, 9, 2, 6], but with a potential drawback on inflation rate of inbreeding  
20 [10, 11]. However, the use of genomic prediction seems to solve this problem [11, 12]  
21 and recent studies confirmed the interest of CCPS to optimize long term genomic  
22 selection in crossbreeding schemes [13, 14]. Nevertheless, in practice, this approach  
23 has some limitations in layers because, generally, only purebred selection candidates,  
24 males and/or females, are genotyped. Christensen et al. [15] extended the Wei and  
25 van der Werf [8] model to take advantage of single step methods, which allows  
26 jointly evaluating animals genotyped or not, to solve this problem.

27  
28 Furthermore, a strategy commonly applied by layers breeding companies consists  
29 in checking the overall suitability for the combination of purebred males, by

30 using the performances of different crossbred progeny, eventually reared in various  
31 environments. In this case, genotyping is available only for these purebred males,  
32 that represent only one of the various purebred lines used in the crossbreeding  
33 scheme. The performances are recorded on their crossbred daughters which are half-  
34 daughters with unknown dam pedigree. In Picard Druet *et al.* [16], the relevance  
35 of genomic evaluation of egg quality traits, with the ssGBLUP methodology [17],  
36 was analysed in a purebred line, i.e. using the purebred performances (PB). The  
37 present study focuses on the use of performances from crossbred individuals issued  
38 from the same purebred line. Our objective was to estimate the impact of adding  
39 these crossbred performances on Genomic Estimated Breeding Values (GEBV) of  
40 purebred male candidates to predict PB or CB egg quality traits. We studied  
41 accuracy and biases of GEBVs obtained with bitrait models, i.e. PB and CB for one  
42 trait, and compared these with those obtained with unitrait purebred evaluation, i.e.  
43 only PB. In that context, we also estimated the impact of using genomic evaluation  
44 instead of genetic evaluation, on accuracy and biases.

## 45 **Material and Methods**

46 Purebred population, animals, traits measurement and genotyping were described  
47 in more details in Picard-Druet *et al.* [16].

### 48 **Animals**

#### 49 *Purebreds*

50 Purebred performances concerned a pure line of Rhode Island layers selected  
51 by Novogen. Hens were hatched in 12 batches, born from 2008 to 2015, which  
52 corresponded to 4 generations (P0 to P3, cf figure 1). The genealogy of all the birds  
53 was recorded in the pedigree file, which concerned 2,273 breeders: 514 sires and  
54 1,759 dams.

55

56 Between 18 and 60 weeks of age, hens were bred in collective cages of 5 full-sisters  
57 (14,985 hens). Egg quality traits were measured twice, at 30 and 50 weeks of age  
58 (27,915 eggs measured). At 60 weeks of age a selection was applied (essentially on  
59 pecking) and hens were transferred in individual cages to continue their career from  
60 60 weeks to 80 weeks of age (7,982 hens). Egg quality traits were measured once a  
61 week until the end of their career.

### 62 *Crossbreds*

63 In addition to purebred performances, performances from the crossbred offspring  
64 of purebred males (1,384 sires) were available (cf figure 2). These hens, born from  
65 2010 to 2016, were issued from the 3 generations P0 to P2 (C1 to C3, cf figure  
66 2). Each purebred male had in expectation 45 crossbred daughters with various  
67 genetic types dams (purebred or crossbred). The dam side of the hens pedigree was  
68 unknown.

69

70 Hens (44,460 birds) were hatched in collective cages of 12 half sisters of sire, from  
71 18 weeks to 70 weeks of age. Egg quality traits were measured two times in their life:  
72 at 50 weeks old (41,343 performances), and at 70 weeks old (37,232 performances).

### 73 *Combining performances*

74 Purebred and crossbred performances were analysed with two different data sets,  
75 depending on the age of hens:

76 The first data set was composed with purebred performances from the collective  
77 cages (30 to 50 weeks of age) and crossbred performances at 50 weeks old. These  
78 performances were on young hens, during their laying plateau, so with relatively  
79 homogeneous performances (cf table 1). This group was labeled YH in the following.

80

81 The second data set was formed with the purebred performances from individual  
82 cages (from 60 weeks), and crossbred performances at 70 weeks old. Hens at this

83 age are at career-end, with more variability in performances (cf table 2). This group  
84 was labeled OH in the following.

85

## 86 Genotypes

87 In this population, 2,374 purebred birds were genotyped using the 600K Affymetrix®  
88 Axiom® HD genotyping array [18]. Each animal was genotyped for 580,961 SNP  
89 markers.

90 Genotypes were then filtered through four successive steps: individuals with a call  
91 rate <95% were removed (0 animals excluded); SNPs with a MAF <0.05 were  
92 removed (258 772 SNPs); SNPs with a call rate <95% were removed (7 549 SNPs);  
93 SNPs whose genotype frequencies deviated significantly from the Hardy-Weinberg  
94 equilibrium ( $P < 10^{-4}$ ) were removed (12 538 SNPs). Animals showing pedigree  
95 incompatibilities were also removed (12 individuals excluded).

96 Thus, 302,102 SNPs and a total of 1,214 genotyped males and 1,148 genotyped  
97 females were kept for the study.

## 98 Traits

99 In this paper, traits are named according to Animal Trait Ontology for Livestock  
100 <http://www.atol-ontology.com/>. Three egg quality traits were studied: egg  
101 weight (EW), egg shell color (ESC), and egg shell strength (ESS). Summary  
102 statistics on traits are given in tables 1 & 2. Traits from purebred and crossbred  
103 were considered separately, and labelled respectively PB and CB.

### 104 *Trait measurements*

105 Quality traits were measured by Zootest company (Ploufragan, France). The first  
106 step consisted of measuring EW (in g). Then, egg shell color was measured with  
107 a Minolta chromameter and three traits were recorded: redness of egg shell  $a^*$ ,  
108 yellowness of egg shell  $b^*$  and lightness of egg shell  $L^*$ . Egg shell color was then

109 calculated as:  $ESC = 100 - (L * -a * -b*)$ . This equation is used by the company  
110 Novogen on routine practice for performance quality control. Third, shell strength  
111 was measured using a compression machine to evaluate the static stiffness of the  
112 shell. The egg was compressed between two flat plates moving at constant speed.  
113 ESS is the maximum force recorded before eggshell fracture (in N, multiplied by  
114 100).

#### 115 *Choice of environmental effects to introduce into the model and elimination of* 116 *outliers*

117 For each PB and CB trait independently, egg measurements were adjusted for  
118 environmental effects using the SAS<sup>®</sup> 9.4 GLM procedure. The objective was to  
119 check which environmental conditions had a significant effect on the traits and to  
120 remove outliers.

121 The environmental effects tested were the hatch, the cage location in the poultry  
122 house within hatch, i.e. the battery and the cage location along the battery, the  
123 waiting time between sample and egg measurement and the age of the hen.

124 All these environmental effects were below the significance level ( $P < 0.05$ ), which  
125 means that they could be kept into the model of analysis. Raw data were then  
126 adjusted using the estimates of all these effects and outliers were deleted. Were  
127 considered as outliers the values presenting a deviation from the sire' family mean  
128 higher than 5 phenotypic standard deviations.

129 Finally, 69,313 records in YH and 112,354 records in OH were retained (cf tables 1  
130 and 2). All distributions of adjusted phenotypes were symmetrical (data not shown).

#### 131 Genetic and genomic evaluations

132 Performances were centered (subtracting the general mean) and standardized  
133 (divided by the standard deviation) before evaluation. Each trait (EW, ESC and  
134 ESS) was evaluated separately, applying 3 bi-variate models (EW PB & EW CB,  
135 ESC PB & ESC CB, ESS PB & ESS CB). These evaluations were performed using

136 BLUP (EBV) and single-step GBLUP (GEBV) methodologies [19, 17]. To perform  
137 those evaluations, the BLUPF90 family of programs [20] was used. Variance-  
138 covariance matrices were estimated using REMLF90 to get a good starting point  
139 for AIREMLF90. AIREMLF90 was then used to validate these variance-covariance  
140 components (the components have changed little between the two programs) and  
141 estimate the standard errors of genetic parameters estimates. EBVs and GEBVs  
142 were then estimated using BLUPF90.

#### 143 *Statistical model*

144 The statistical model was the same for all traits and took into account the  
145 environmental effects, fixed and covariables, described before, plus the random  
146 genetic effect of the animal. For CB performances, the dam side of the hens  
147 pedigree was unknown and only the sire was recorded. However, no "unknown-  
148 parent-group" [21] was taken into account because dams were all born in a short  
149 time (3 generations). So they were considered of stable genetic level along the time.

150

151 For YH evaluations, each egg measured in PB was associated to one cage of 5  
152 full sisters, without knowing which hen laid which egg. As the measurements were  
153 repeated twice, i.e. at 30 and 50 weeks of age, each hen had two measured eggs in  
154 expectation, but it was not possible to take into account the common environmental  
155 effect of the hen. A "egg-animal" model, without common environmental effect,  
156 provided slightly more accurate (G)EBV than a sire model (data not shown).  
157 The inclusion of a "cage" random effect was also tested (instead of the fixed  
158 effects, which described the geolocation in the poultry building). The estimate  
159 of this variance-covariance component was very low but the computation time was  
160 significantly higher with this model. The "egg-animal" model was then retained  
161 and the heritability estimate as the ratio between the animal variance and the sum  
162 of animal and residual variances.

163 Conversely, for OH evaluations of PB traits, several measurements were available  
164 by hen, and a random common environmental effect of the hen was taken into  
165 account in the model. We therefore estimated the heritability coefficient as the  
166 ratio between the animal variance and the sum of animal, common environmental  
167 and residual variances, and the repeatability coefficient as the ratio between the sum  
168 of animal and common environmental variances and the sum of animal, common  
169 environmental and residual variances.  
170 Concerning the CB traits, all the measurements were made in collective cages, so  
171 as PB traits in YH, a "agg-animal" model was used for evaluation.

172

### 173 *Candidate population*

174 To assess the relevance of genetic evaluation and of genomic evaluation, the  
175 statistical properties (described below) of estimated breeding values, EBVs and  
176 GEBVs, of selection candidates were used. In this study, we used a male candidate  
177 population which was composed with purebred birds from generation P2, with  
178 purebred daughters in P3 and crossbred daughters in C3 (cf figure 2). This group  
179 contain 174 sires in YH and 93 sires in OH.

### 180 *Reliability of prediction*

181 To assess the ability of genetic or genomic evaluations, the estimated breeding  
182 values EBVs and GEBVs had to be compared to the true breeding values (TBVs)  
183 of candidates. However, TBV is never known when working on real data set and  
184 could not be approximated in our study with Daughter Yield Deviation (DYD) [22]  
185 since our candidates had only few offspring.

186

187 Therefore, to estimate accuracy and bias of prediction of our evaluations, the  
188 LR method [23] was used. This method compares evaluations results based on  
189 complete and partial data sets, since the amount of change expected in consecutive

190 genetic evaluations was described as a function of their respective accuracy [24]. This  
191 method completes cross-validation approaches with semiparametric elements, based  
192 on the mixed model equations, to estimate the “population” accuracy. Population  
193 accuracy is relevant to compare the predictive ability of models and to maximize  
194 genetic progress.

195

196 All available pedigree, phenotypes, and genotypes for GEBV estimation, from G0  
197 to G3, made up the complete data set. Two cases of partial data sets were studied,  
198 based on the amount of phenotypic information available when the evaluation was  
199 carried out:

- 200 • Case 1: The evaluation was carried out at birth of candidates, without  
201 considering the performances of their contemporary relatives. The phenotyped  
202 population was limited to the candidates and their ancestors: all individuals  
203 had a pedigree information; female ancestors had phenotypes; male ancestors  
204 and male candidates had genotypes for GEBV estimation.
- 205 • Case 2: The evaluation was carried out at 60 weeks of age for YH and 80  
206 weeks of age for OH, without considering the performances of the progeny  
207 of candidates. The evaluated population included the candidates, their  
208 contemporary relatives and their ancestors: all individuals had a pedigree  
209 information; female ancestors, female contemporary relatives or candidates  
210 had phenotypes; male ancestors had genotypes for GEBV estimation;  
211 candidates had genotypes for GEBV estimation. This case corresponds to  
212 the classical scheme in layers selection.

213 The LR method relies on three statistics to estimate accuracy and biases of an  
214 evaluation:

- 215 • The correlation between (G)EBVs from complete and partial evaluation to  
216 estimate the relative accuracy

- 217 • The difference of means between (G)EBVs from complete and partial  
218 evaluation to estimate the bias
- 219 • The slope of the linear regression of (G)EBVs from complete evaluation on  
220 (G)EBVs from partial evaluation to estimate the over or under dispersion of  
221 estimates

222 In layers, candidates are compared within hatch and there are never comparisons  
223 made between individuals from different generations for the purposes of selection.  
224 Biases and dispersion statistics were presented here simply to illustrate the evolution  
225 of quality of evaluation according to the case 1 or 2.

226 To compare genetic evaluation to genomic evaluation, in the same case scenario, a  
227 fourth statistics was used: the ratio between the relative accuracy of EBVs and the  
228 relative accuracy of GEBVs. This statistics allows to quantify the expected increase  
229 in accuracy from genetic to genomic evaluation.

230 Finally, for each trait and case scenario, two evaluations for PB were compared  
231 by calculating the ratio of their relative accuracy: (G)EBVs obtained in [16]), i.e.  
232 with PB only, and (G)EBVs obtained in the present study, i.e. jointly with PB and  
233 CB.

234

## 235 **Results**

### 236 Genetic parameters

237 The heritabilities were very steady between REMLs carried out with BLUP or with  
238 GBLUP, whatever the observed trait or age (YH or OH). Same results were obtained  
239 for repeatabilities. Genetic correlations were even more stable than heritabilities or  
240 repeatabilities. For each trait, a strong link was observed between PB and CB with  
241 a genetic correlation estimated around 0.75 - 0.8 (cf table 3 and table 4).

242

243 Consequently, variance-covariance matrices were fixed in the following study. The  
244 variance-covariance matrix obtained with the genomic evaluation using the complete  
245 data set, e.g. the REML carried out with the maximum information, was used to  
246 perform subsequent BLUP and GBLUP. Estimates of genetic parameters were given  
247 table 3 for traits measured in YH and table 4 for traits measured in OH.

#### 248 Accuracy on YH traits

249 As expected, relative accuracy estimates (cf table 5) significantly increased with  
250 the amount of phenotypic information available from case 1 to case 2. In case 1,  
251 estimates were non homogeneous depending on the trait: relative accuracy of EBV  
252 for EW was the lowest, with values of 0.09 (CB) and 0.13 (PB), while this trait  
253 was more heritable than the other for which the relative accuracy was between  
254 0.22 and 0.39. These differences were reduced in case 2. It must be emphasized that  
255 differences between traits were observed by Picard-Druet *et al.* [16] but not the very  
256 low accuracy of EBV observed here (in this case, relative accuracy was estimated  
257 to 0.22-0.23).

258

259 Results of the comparison of genetic to genomic evaluation, were showed table 6,  
260 which gave the relative increase in accuracy from BLUP to GBLUP, in each case.  
261 A value of 1 showed no difference, a value under 1 showed a more accurate genomic  
262 evaluation, a value over 1 a more accurate genetic evaluation.

263

264 Evaluations at birth (case 1) were those with the largest differences between  
265 genetic and genomic evaluations. Results were very trait-dependant, with results  
266 varying from 0.90 (ESC CB), e.g. a gain of accuracy of 10% with genomic evaluation  
267 comparing to genetic evaluation, to 0.29 (EW CB), e.g. a gain of accuracy of 71%  
268 with genomic evaluation comparing to genetic evaluation. Overall, in this case, an  
269 advantage of accuracy in favour of genomic evaluation was estimated, linked to

270 the evaluated trait. For the evaluations at 18 months (case 2), differences between  
271 genetic and genomic evaluations were weaker, with values ranged from 0.97 (ESC  
272 PB) to 0.83 (ESS CB), with a mean of 0.94. As in case 1, globally a relative increase  
273 of accuracy using GBLUP was shown. These results are in adequation with what  
274 we previously found with evaluation with purebred performances.

275

276 Results of comparison between evaluation on purebred and evaluation on mixed  
277 performances were showed table 7, which gave the relative increase in accuracy from  
278 PB evaluation to MP evaluation, in each case. A value of 1 showed no difference,  
279 a value under 1 showed a more accurate MP evaluation, a value over 1 a more  
280 accurate separate evaluation.

281

282 In case 1, results vary depending methodology used, and the trait. In genetic  
283 evaluation, results are on mean more precise using PB evaluation (1.15 on mean,  
284 e.g. a 15% gain using separate evaluation), but their strongly vary from 0.67  
285 (ESC PB) to 1.69 (EW PB). Using genomic evaluation, however, results are more  
286 homogeneous: comprised between 1.16 and 0.70, with a mean of 0.95, they showed  
287 a slightly more accurate results using MP evaluations.

288

289 In case 2, there is no more strong differences between genetic and genomic  
290 evaluation, with values ranged from 0.97 (ESS CB genomic) to 0.66 (EW PB  
291 genomic). In this case, overall, a relative increase of accuracy using MP was shown  
292 on all traits.

### 293 Accuracy on IC traits

294 Results for IC traits are in tables 8, 9. Concerning the effect of genomic evaluation, as  
295 in CC, estimation of accuracies were different for each trait, depending on evaluation  
296 scenario. And accuracy evolution in function of scenarios was also depending on the

297 amount of phenotypic information. A relative increase of accuracy was also observed  
298 for case 1 and 2. However, this increase was estimated stronger in IC traits, both  
299 in case 1 (mean = 0.72) and 2 (mean = 0.78).

300

301 Concerning the comparison between PB evaluation to MP evaluation, results  
302 obtained using IC traits (table 10) were more stable than using CC traits. Globally,  
303 they showed an increase of accuracy using MP evaluation, for all traits (at the  
304 exception of ESS in Case 1 EBV, and EW in case 1 GEBV).

305

#### 306 (G)EBVs biases and dispersion

307 The bias statistics has an expected value of 0 if evaluation is unbiased. Both in CC  
308 and IC (cf tables 13 and 14), biases were low and often negative, signing an under  
309 estimation of (G)EBVs in PB evaluation. Biases increased in the same time as the  
310 amount of phenotypic information decreased, from about 0 in case 2 to between  
311 0 and -0.11 in case 1. Biases were slightly lower for genomic values comparing to  
312 genetic values in case 1, and slightly higher in case 2. The differences between traits  
313 vary without any clear tendency being observed. MP evaluations have stronger  
314 biases than PB in CC, especially in Case 1 (EBV and GEBV). In IC, MP and PB  
315 evaluations present comparable biases.

316

317 Unbiased estimators are supposed to have a regression slope equal to 1. In CC  
318 as in IC (tables 11 and 12), slopes decreased below 1 every time the amount of  
319 phenotypic information decreased. In case 1, regression seems less biased in genomic  
320 evaluation than in genetic evaluation, both in CC and IC traits (at the exception  
321 of ESC in CC). However, an opposite behaviour was observed in case 2, with the  
322 slope of regression being closer to 1 using genetic evaluation compared to genomic  
323 evaluation. The slopes were also strongly link to the evaluated traits in a given

324 case-candidates population. In CC, MP evaluation present stronger biases than PB  
325 evaluation in Case 1, for EW and ESS, but the inverse can be witness

326

## 327 **Discussion**

### 328 Genetic parameters

329 Estimates of heritability, repeatability and genetic correlations on both CC traits  
330 (table 3) and IC traits (table 4) were in accordance to the litterature [25, 26].  
331 Moderate to high heritability coefficients allow to expect significant genetic gains  
332 by selection on these three egg quality traits.

333 Concerning genetic correlation between PB and CB traits, if a strong link was  
334 observed, they can't be considered as the same trait. To evaluate breeders on their  
335 PB traits for CB traits selection objective (or conversly CB for PB objective), or  
336 evaluate them as the same trait, would lead to mediocre evaluation, as they don't  
337 cover the same aspect of breeders performances. As a consequence, PB and CB  
338 should be considered, and evaluate, as 2 different trait.

### 339 MP genetic vs MP genomic

340 In our two case evaluations, our results highlighted a more accurate evaluation of  
341 candidates with GBLUP compared to BLUP, on PB and CB traits, using both CC  
342 and IC data sets. Gains using genomic evaluation instead of genetic evaluation were  
343 also strongly link to the evaluated trait, and the amount of phenotypic information  
344 available. Anyway, the use of genomic selection appeared to be a good way to obtain  
345 more precise MP evaluations.

346

347 In terms of biases, results were less clear, depending on the observed case. With  
348 evaluations at birth (case 1), results tends to show that genomic evaluation leads to  
349 less biased estimations, on both of our studied statistics. In the 18 months evaluation  
350 case however (case 2), results showed the opposite, with genomic evaluations

351 providing more biased estimations. Those biases have an impact on selection in  
352 case of breeder choice between different generations. However, in poultry selection,  
353 breeders are selected in the same generation, as older breeders aren't kept for  
354 reproduction, as it can happened in cattle selection. So, if it's a very interesting  
355 information to have, it should not impact selection schemes.

#### 356 MP evaluation vs PB evaluation

357 At the exception of EW and ESS genetic evaluations in case 1, using CC traits, MP  
358 evaluation was always estimated with results at least as precise, and often more  
359 accurate, compared to PB evaluation, using both CC and IC data sets. Again,  
360 results were strongly linked to the studied trait. It showed that the use of MP  
361 evaluation should provide more accurate evaluations for selection.

362 The fact that EW and (to a lesser extent) ESS show a different behaviour in case  
363 1 is quite interesting. It could be a sign of non-additive genetic effects on crossbred  
364 traits, which in a context of an evaluation with few performances, with purebred  
365 and crossbred performances, destabilize the model. In this case, adding crossbred  
366 performances, strongly impacted by effects as heterosis, could lead to damaging the  
367 model instead of improving it, as performances strongly affected by non additive  
368 effects are used to construct an additive model. This leading to a less accurate  
369 evaluation.

370

## 371 **Conclusion**

372 The results of the present study lead to several conclusions regarding the use of  
373 crossbred offspring performances in purebred evaluation.

374

375 The correlation between PB and CB traits showed that these traits should be  
376 considered, and evaluated, as two different traits.

377 On the comparison of genomic evaluation versus genetic evaluation using MP

378 evaluation, we observed that genomic evaluation is always more accurate than  
379 genetic evaluation, with the same amount of phenotypic information, which confirm  
380 our previous results.

381 Between MP evaluation and PB evaluation, our results showed that in most of the  
382 cases, MP evaluation provided more accurate (G)EBVs for purebred candidates.

383

384 Those results are quite interesting for the poultry industry, on 2 aspects.

385 As PB and CB traits should not be considered as the same trait, breeders evaluation  
386 on CB traits is very important when selection goal is on crossed animals. The use of  
387 CB data in breeders evaluation allow to evaluate candidates on these traits, instead  
388 of PB traits, less adapted for the selection objective. Or, in an other hand, continue  
389 to select on PB traits, but with a better accuracy.

390 Then, on the use of genomic evaluation, we showed this methodology is a way  
391 to improve PB evaluation in layers, and this worked as well for MP evaluation  
392 accuracy. It seems interesting to use genomic evaluation to move from selection  
393 at 18 months to a (near) birth selection for males. The reduction of precision in  
394 evaluation should be compensated by an increase of selection pressure and a shorter  
395 generation interval, and allows an optimization of expected genetic gain.

396

397 Adding performances from CB offspring to evaluate purebred breeders seems very  
398 promising, especially combined with genomic selection. This should allow to further  
399 optimize evaluations schemes, to obtain a better genetic gain each generation.

#### 400 **Ethics statement**

401 All information used in this study were collected from hens raised in the context of layers selection. These animals  
402 and the scientific investigations described herein are therefore not to be considered as experimental animals per se,  
403 as defined in EU directive 2010/63 and subsequent national application texts. Consequently, we did not seek ethical  
404 review and approval of this study as regarding the use of experimental animals. All animals were reared on the  
405 Novogen nucleus herd in compliance with national regulations pertaining to livestock production and according to  
406 procedures approved by the French Veterinary Services.

**407 Availability of data and materials**

408 The datasets used and/or analysed throughout the present study are available from the corresponding author on  
409 reasonable request.

**410 Competing interests**

411 The authors declare that they have no competing interests.

**412 Funding**

413 This research project was supported by the French national research agency ANR, within the framework of project  
414 ANR-10-GENOM.BTV-015 UtOpIge. DPD is a PhD fellow supported by the French national agronomic research  
415 agency (INRA), with SelGen metaprogram, and Bretagne region.

**416 Author's contributions**

417 DPD performed the statistical analyses and drafted the manuscript. FIH filtered the genotype data. AVa and TB  
418 supervised animal management and production. PLR supervised the analyses and participated in drafting the  
419 manuscript. PLR and TB conceived the study. All authors contributed to the ideas and methods. All authors read  
420 and approved the final manuscript.

**421 Acknowledgements**

422 This research project was partly supported by the French national research agency ANR within the framework of  
423 project ANR-10-GENOM.BTV-015 UtOpIge. DPD is a PhD fellow supported by the Brittany region (France) and  
424 the INRA metaprogram SelGen.

**425 Author details**

426 <sup>1</sup>PEGASE, INRA, Agrocampus Ouest, 16 Le Clos, 35590 Saint-Gilles, France. <sup>2</sup>NOVOGEN, 5, rue des Compagnons  
427 - Secteur du Vau Ballier, 22960 Plédran, France.

**428 References**

- 429 1. Wei M, Steen HAMvd. Comparison of reciprocal recurrent selection with pure-line selection systems in animal  
430 breeding (A review). *Animal Breeding Abstracts*. 1991; Available from:  
431 <http://agris.fao.org/agris-search/search.do?recordID=NL2012088130>.
- 432 2. Lutaaya E, Misztal I, Mabry JW, Short T, Timm HH, Holzbauer R. Genetic parameter estimates from joint  
433 evaluation of purebreds and crossbreds in swine using the crossbred model. *Journal of Animal Science*. 2001  
434 Dec;79(12):3002–3007. Available from: <https://academic.oup.com/jas/article/79/12/3002/4683799>.
- 435 3. Xiang T, Christensen OF, Vitezica ZG, Legarra A. Genomic evaluation by including dominance effects and  
436 inbreeding depression for purebred and crossbred performance with an application in pigs. *Genetics, selection,*  
437 *evolution: GSE*. 2016;48(1):92.
- 438 4. Xiang T, Nielsen B, Su G, Legarra A, Christensen OF. Application of single-step genomic evaluation for  
439 crossbred performance in pig. *Journal of Animal Science*. 2016 Mar;94(3):936–948.
- 440 5. Wei M, van der Werf JH. Genetic correlation and heritabilities for purebred and crossbred performance in  
441 poultry egg production traits. *Journal of Animal Science*. 1995 Aug;73(8):2220–2226.
- 442 6. Caverio D, Schmutz M, Preisinger R. Genetic evaluation of pure-line and cross-line performance in layers.  
443 *Lohmann Information*. 2010 Oct;45 (2):18.
- 444 7. Ming Wei. Combined crossbred and purebred selection in animal breeding. PhD thesis, Wageningen University,  
445 Netherlands. 1992;.
- 446 8. Wei M, Werf JHJvd. Maximizing genetic response in crossbreds using both purebred and crossbred information.  
447 *Animal Science*. 1994 Dec;59(3):401–413. Available from:

- 448 <https://www.cambridge.org/core/journals/animal-science/article/>  
449 [maximizing-genetic-response-in-crossbreds-using-both-purebred-and-crossbred-information/](https://www.cambridge.org/core/journals/animal-science/article/div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-information/0F465082DDBE9CC925FFF268AF6437A2)  
450 [D8C19321FA46807D4F061A869A7CA990](https://www.cambridge.org/core/journals/animal-science/article/div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-information/0F465082DDBE9CC925FFF268AF6437A2).
- 451 9. Bijma P, Arendonk JAMv. Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both  
452 purebred and crossbred information. *Animal Science*. 1998 Apr;66(2):529–542. Available from:  
453 [https://www.cambridge.org/core/journals/animal-science/article/](https://www.cambridge.org/core/journals/animal-science/article/div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-information/0F465082DDBE9CC925FFF268AF6437A2)  
454 [div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-informa](https://www.cambridge.org/core/journals/animal-science/article/div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-information/0F465082DDBE9CC925FFF268AF6437A2)  
455 [0F465082DDBE9CC925FFF268AF6437A2](https://www.cambridge.org/core/journals/animal-science/article/div-classtitlemaximizing-genetic-gain-for-the-sire-line-of-a-crossbreeding-scheme-utilizing-both-purebred-and-crossbred-information/0F465082DDBE9CC925FFF268AF6437A2).
- 456 10. Bijma P, Woolliams JA, Arendonk JAMv. Genetic gain of pure line selection and combined crossbred purebred  
457 selection with constrained inbreeding. *Animal Science*. 2001 Jan;72(2):225–232. Available from:  
458 [https://www.cambridge.org/core/journals/animal-science/article/](https://www.cambridge.org/core/journals/animal-science/article/genetic-gain-of-pure-line-selection-and-combined-crossbred-purebred-selection-with-constrained-inbreeding/EF8993AEFEB6A2B7DDA1C2DA77366003)  
459 [genetic-gain-of-pure-line-selection-and-combined-crossbred-purebred-selection-with-constrained-inbreeding/](https://www.cambridge.org/core/journals/animal-science/article/genetic-gain-of-pure-line-selection-and-combined-crossbred-purebred-selection-with-constrained-inbreeding/EF8993AEFEB6A2B7DDA1C2DA77366003)  
460 [EF8993AEFEB6A2B7DDA1C2DA77366003](https://www.cambridge.org/core/journals/animal-science/article/genetic-gain-of-pure-line-selection-and-combined-crossbred-purebred-selection-with-constrained-inbreeding/EF8993AEFEB6A2B7DDA1C2DA77366003).
- 461 11. Dekkers JCM. Prediction of response to marker-assisted and genomic selection using selection index theory.  
462 *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie*. 2007  
463 Dec;124(6):331–341.
- 464 12. Lopes MS, Bovenhuis H, Hidalgo AM, van Arendonk JAM, Knol EF, Bastiaansen JWM. Genomic selection for  
465 crossbred performance accounting for breed-specific effects. *Genetics, Selection, Evolution : GSE*. 2017 Jun;49.  
466 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5485705/>.
- 467 13. Esfandyari H, Sørensen AC, Bijma P. A crossbred reference population can improve the response to genomic  
468 selection for crossbred performance. *Genetics Selection Evolution*. 2015 Sep;47(1):76. Available from:  
469 <https://doi.org/10.1186/s12711-015-0155-z>.
- 470 14. Esfandyari H, Sørensen AC, Bijma P. Maximizing crossbred performance through purebred genomic selection.  
471 *Genetics Selection Evolution*. 2015 Mar;47(1):16. Available from:  
472 <https://doi.org/10.1186/s12711-015-0099-3>.
- 473 15. Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances.  
474 *Genetics, selection, evolution: GSE*. 2014 Mar;46:23.
- 475 16. Picard Druet D, Varenne A, Herry F, Hérault F, Allais S, Burlot T, et al. Relevance of genomic evaluation for  
476 egg quality traits in layers. *bioRxiv*. 2019 Jul;p. 704742. Available from:  
477 <https://www.biorxiv.org/content/10.1101/704742v1>.
- 478 17. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize  
479 phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of*  
480 *Dairy Science*. 2010 Feb;93(2):743–752.
- 481 18. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP  
482 genotyping array for chicken. *BMC Genomics*. 2013 Jan;14(1):59. Available from:  
483 <https://doi.org/10.1186/1471-2164-14-59>.
- 484 19. Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree,  
485 and genomic information. *Journal of Dairy Science*. 2009 Sep;92(9):4648–4655.
- 486 20. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee D. BLUPF90 and related programs. In: *Proceedings of*  
487 *the 7th World Congress on Genetics Applied to Livestock Production*. vol. Vol. 28; 2002. p. 743.
- 488 21. Misztal I, Vitezica ZG, Legarra A, Aguilar I, Swan AA. Unknown-parent groups in single-step genomic  
489 evaluation. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie*.  
490 2013 Aug;130(4):252–258.

491 22. VanRaden PM. Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 2008  
 492 Nov;91(11):4414–4423.

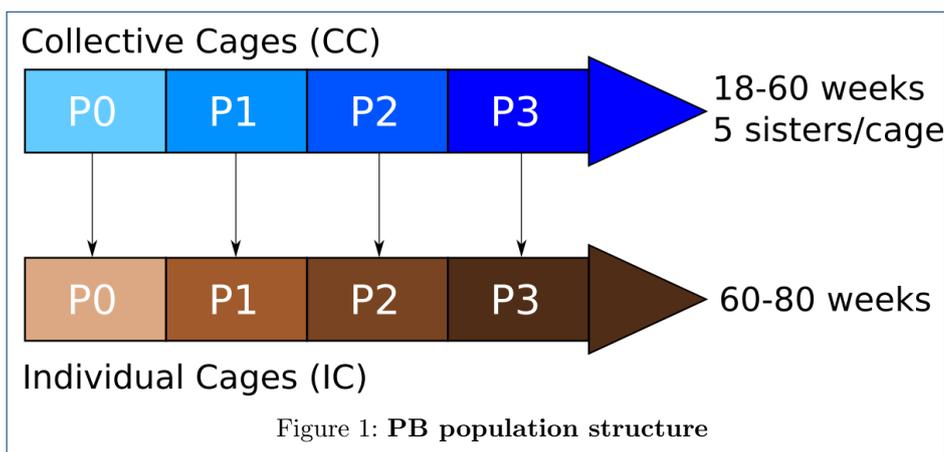
493 23. Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding  
 494 values and future phenotypes using the LR method. *Genetics, selection, evolution: GSE*. 2018 Nov;50(1):53.

495 24. Reverter A, Golden BL, Bourdon RM, Brinks JS. Technical note: detection of bias in genetic predictions2.  
 496 *Journal of Animal Science*. 1994 Jan;72(1):34–37. Available from:  
 497 <https://academic.oup.com/jas/article/72/1/34-37/4632556>.

498 25. Beaumont C, Calenge F, Chapuis H, Fablet J, Minvielle F, Tixier-Boichard M. Génétique de la qualité de l'œuf  
 499 - *Inra Prod.Anim.*, 23 (2), 123-132; 2010. Available from: [http://www6.inra.fr/productions-animales/  
 500 2010-Volume-23/Numero-2-2010/Genetique-de-la-qualite-de-l-aeuf](http://www6.inra.fr/productions-animales/2010-Volume-23/Numero-2-2010/Genetique-de-la-qualite-de-l-aeuf).

501 26. Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Response and inbreeding from a  
 502 genomic selection experiment in layer chickens. *Genetics Selection Evolution*. 2015;47:59. Available from:  
 503 <http://dx.doi.org/10.1186/s12711-015-0133-5>.

504 **Figures**



505 **Tables**

Traits	EW (g)		ESC ( <i>without units</i> )		ESS ( <i>N * 100</i> )	
	EW PB	EW CB	ESC PB	ESC CB	ESS PB	ESS CB
Number of records	27915	41307	27932	41305	25758	38047
Mean	59.9	62.3	81.2	81.5	3785.3	4109.3
Standard deviation	4.56	4.97	9.12	10.55	691.50	749.73
Min	41	38.5	35.36	29.8	870	810
Max	82	89.4	109.981	114.5	6620	7090

Table 1: Summary statistics on phenotypic data in YH  
 EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

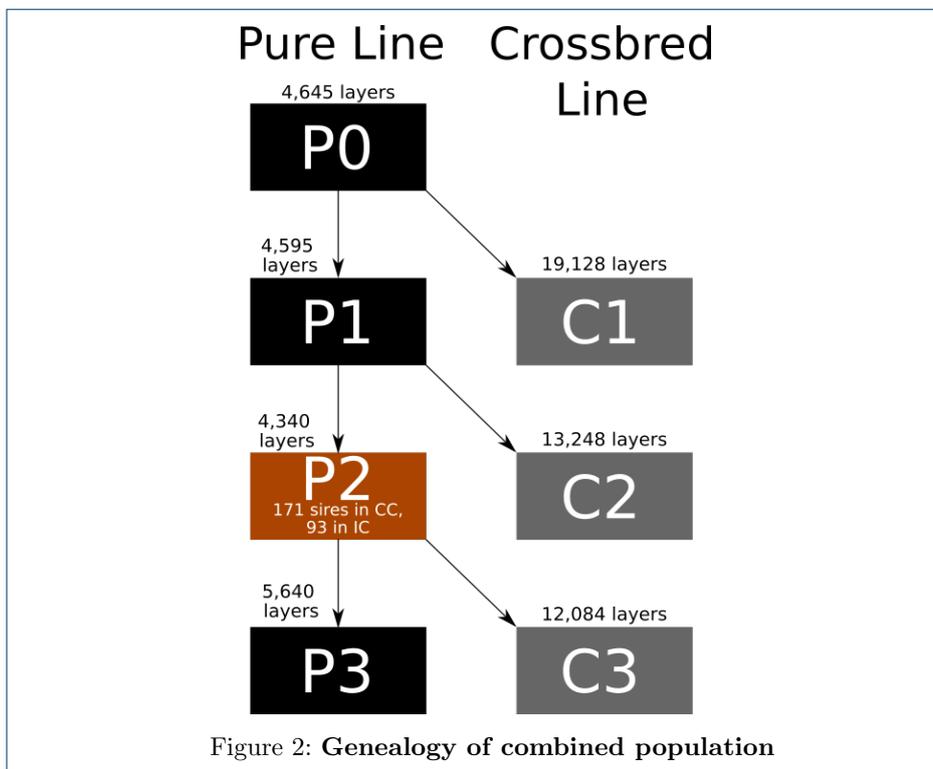


Figure 2: Genealogy of combined population

Traits	EW (g)		ESC (without units)		ESS (N * 100)	
	EW PB	EW CB	ESC PB	ESC CB	ESS PB	ESS CB
Number of records	74976	37212	73033	36702	65890	32904
Mean	60.0	63.1	78.4	80.0	3671.7	3786.1
Standard deviation	4.79	5.49	9.82	11.09	741.73	763.85
Min	41.7	37.7	30.55	28.86	682	640
Max	82.8	91.6	109.295	111.48	7187	6923

Table 2: Summary statistics on phenotypic data in OH

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Traits	heritability PB	heritability CB	Genetic correlation PB-CB (standard error)
EW	0.65	0.8	0.76 (0.03)
ESC	0.62	0.55	0.82 (0.02)
ESS	0.28	0.27	0.77 (0.02)

Table 3: Genetic parameters for YH data

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Traits	heritability/repeatability PB	heritability/repeatability CB	Genetic correlation PB-CB (standard error)
EW	0.33/0.43	0.66/0.76	0.82 (0.03)
ESC	0.34/0.37	0.42/0.48	0.78 (0.02)
ESS	0.19/0.21	0.25/0.28	0.73 (0.02)

Table 4: Genetic parameters for OH data

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.13	0.29	0.65	0.71
EW CB	0.09	0.31	0.54	0.65
ESC PB	0.39	0.47	0.73	0.75
ESC CB	0.4	0.44	0.62	0.71
ESS PB	0.38	0.44	0.58	0.63
ESS CB	0.22	0.38	0.48	0.58
Mean SE	0.07	0.07	0.06	0.06

Table 5: Estimation of accuracy of males evaluation on YH traits, using LR method.  
 EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
 Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1	Case 2
EW PB	0.46	0.92
EW CB	0.29	0.84
ESC PB	0.83	0.97
ESC CB	0.90	0.87
ESS PB	0.87	0.93
ESS CB	0.57	0.83

Table 6: Ratio of the relative increase from BLUP to GBLUP, on YH traits  
 EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
 Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	1.69	1.03	0.69	0.66
ESC PB	0.67	0.7	0.71	0.75
ESS PB	1.11	0.95	0.84	0.86

Table 7: Ratio of the relative increase from PB evaluation to MP evaluation, on YH PB traits  
 EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
 Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.41	0.54	0.65	0.78
EW CB	0.29	0.42	0.59	0.72
ESC PB	0.43	0.57	0.58	0.76
ESC CB	0.37	0.51	0.56	0.76
ESS PB	0.46	0.53	0.7	0.82
ESS CB	0.23	0.44	0.49	0.7
Mean SE	0.1	0.09	0.08	0.07

Table 8: Estimation of accuracy of males evaluation on OH traits, using LR method. EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1	Case 2
EW PB	0.77	0.83
EW CB	0.70	0.82
ESC PB	0.76	0.76
ESC CB	0.72	0.74
ESS PB	0.87	0.85
ESS CB	0.53	0.70

Table 9: Ratio of the relative increase from BLUP to GBLUP, on OH traits. EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.85	1	0.75	0.81
ESC PB	0.7	0.74	0.6	0.71
ESS PB	1.07	0.98	0.81	0.87

Table 10: Ratio of the relative increase from PB evaluation to MP evaluation, on OH PB traits. EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.22	0.46	0.88	0.93
EW CB	0.13	0.45	0.73	0.65
ESC PB	0.77	0.64	0.92	0.82
ESC CB	0.89	0.64	0.88	0.79
ESS PB	0.55	0.57	0.75	0.61
ESS CB	0.4	0.58	0.7	0.59

Table 11: Slope of regression for males (G)EBVs for YH traits. EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength  
Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.64	0.8	0.94	0.78
EW CB	0.53	0.6	0.89	0.72
ESC PB	0.77	0.8	0.78	0.7
ESC CB	0.76	0.78	0.83	0.7
ESS PB	0.59	0.7	0.86	0.7
ESS CB	0.37	0.65	0.66	0.62

Table 12: Slope of regression for males (G)EBVs for OH traits.

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	-0.11	-0.07	0.01	0.07
EW CB	-0.04	-0.03	0.05	0.11
ESC PB	-0.06	-0.07	-0.05	0
ESC CB	-0.05	-0.06	-0.05	-0.01
ESS PB	-0.04	0.02	0	0.07
ESS CB	-0.02	0.03	0	0.08

Table 13: Biases of (G)EBVS for YH traits.

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

Traits	Case 1 EBV	Case 1 GEBV	Case 2 EBV	Case 2 GEBV
EW PB	0.02	0.02	0.05	0.07
EW CB	0.02	0.03	0.08	0.15
ESC PB	-0.04	-0.05	-0.04	-0.03
ESC CB	-0.06	-0.07	-0.06	-0.07
ESS PB	-0.05	-0.01	0.01	0.07
ESS CB	-0.04	0.01	0.02	0.1

Table 14: Biases of (G)EBVS for OH traits.

EW: Egg Weight; ESC: Egg Shell Color; ESS: Egg Shell Strength

Case 1: Evaluation performed using only ancestry performances; Case 2: Evaluation performed using ancestry and contemporary relatives performances

## 4.2 Conclusions de l'article 2

Ces travaux ont permis de conclure sur plusieurs points concernant l'utilisation des performances des descendantes croisées pour l'évaluation des reproducteurs en lignée pure.

Les corrélations obtenues entre caractères mesurés en lignée pure et sur la descendance croisée ont permis de montrer que si ce sont deux caractères fortement corrélés, on ne peut les considérer comme étant identiques, et ils devraient être évalués séparément. La précision comparée entre évaluations génétiques et génomiques en évaluations utilisant les jeux de données mixte confirment les résultats trouvés dans le chapitre 3.

Les évaluations utilisant les jeux de données mixtes ont également obtenues des précisions supérieures ou équivalentes à celles obtenues en évaluations utilisant uniquement les performances en lignée pure.

Ces résultats sont intéressants pour l'industrie sur deux aspects.

Les caractères en lignée pure et sur descendance croisée ne pouvant pas être considérés comme identiques, les reproducteurs devraient être évalués sur les caractères mesurés sur la descendance croisée quand l'objectif de sélection est sur les animaux croisés. L'utilisation des performances croisées en évaluation permet cela ou, si l'objectif de sélection demeure sur les performances en lignée pure, de sélectionner sur les caractères de lignée pure avec une précision améliorée.

Le second aspect concerne l'optimisation du progrès génétique. Les travaux du chapitre précédent avaient montré un intérêt certain à passer à la sélection génomique à la naissance, compensant une baisse légère de la précision par une augmentation de l'intensité de sélection et une diminution de l'intervalle de génération. L'ajout des performances des individus croisés, qui ne sont pas valorisés pour le moment, dans l'évaluation permettrait d'améliorer encore d'avantage le progrès génétique des populations sélectionnées.

## Chapitre 5

# Genome Wide Association Studies : étude des différences d'architecture génétique des caractères de qualité d'œuf, et effets de l'utilisation conjointe des informations de descendants pures et croisées sur la détection des QTLs.

### 5.1 Introduction

Dans les parties précédentes, il a été observé que dans un même contexte, en utilisant la même méthode et la même quantité d'informations, les différents caractères évalués obtenaient des précisions d'évaluations assez différentes, sans que cela soit explicable par l'héritabilité des caractères. Une des raisons pouvant expliquer cela est la différence en termes d'architecture génétique de ces caractères (Romé et al., 2015; Morgante et al., 2018). En effet, Romé et al. (2015) ont montré que les caractères de qualité d'œuf sont influencés par des QTLs différents pour chacun, expliquant un pourcentage particulier de la variance génétique. Certains de ces QTLs pouvant expliquer une part importante de la variance, il est possible que cela rentre en conflit avec le modèle infinitésimal du BLUP, dégradant la précision de l'évaluation. De plus, il a également été montré que du point de vue de l'évaluation génétique ou génomique des animaux, les caractères issus de lignée pure ou de descendance croisée ne pouvaient être considérés comme étant identiques, même si ils sont effectivement fortement corrélés.

Une étude d'association pangénomique (Genome-Wide Association Study, GWAS) sur les différents jeux de données permet d'en apprendre davantage sur l'architecture génétique des caractères dans les différentes populations. Ces informations devraient nous permettre de mieux répondre à la question de l'influence de l'architecture

génétique sur la précision de l'évaluation dans nos populations évaluées. Elles devraient également permettre de confirmer si la différence entre caractère mesuré en lignée pure et sur descendance croisée se confirme dans l'architecture génétique du caractère.

De plus, dans le chapitre 5, nous avons montré que l'utilisation conjointe des performances mesurées en lignées pures et celles des descendantes issues de croisement de lignées permettait une amélioration de la précision de l'évaluation des reproducteurs. À partir de ces observations, et étant donné les interrogations relatives à l'architecture génétique des caractères, s'est posée la question de l'efficacité d'une GWAS utilisant ces deux sources d'informations conjointement pour détecter de nouveaux QTLs.

Pour répondre à ces différentes questions, des GWAS ont été effectuées sur les différents jeux complets de données des performances de qualité d'œuf, utilisés dans les chapitres précédents, composés des performances des individus de lignée pure, de leurs descendantes croisées ou des deux combinés. Les résultats de ces GWAS ont été ensuite utilisés pour déterminer les QTLs significatifs, et ces résultats ont été comparés avec ceux obtenus dans la littérature afin d'en assurer la cohérence avec les connaissances actuelles sur l'architecture génétique des caractères de qualité d'œuf.

## 5.2 Méthodes

Les structures des populations, les animaux et les génotypes utilisés, ainsi que les caractères étudiés dans ces travaux sont décrits de manière exhaustive dans le chapitre 3.

### 5.2.1 Animaux

#### Lignée pure

La lignée pure utilisée dans les GWAS était constituée d'une population de Rhode Island, sélectionnée par la société Novogen. Elle était constituée de 12 lots, élevés de 2008 à 2015, répartis en 4 générations. Le pedigree de tous les individus était connu, et concernait 2 273 reproducteurs : 514 coqs et 1 759 poules.

Les poules âgées de 18 et 60 semaines étaient élevées en cages collectives de 5 pleines sœurs (14 985 poules), et leurs œufs mesurés à l'âge de 30 et 50 semaines (27 915 performances). Elles étaient ensuite sélectionnées (essentiellement sur le piquage) à l'âge de 60 semaines, et transférées en cages individuelles pour terminer leur carrière jusqu'à l'âge de 80 semaines (7 082 poules). Les caractères de qualité d'œufs étaient mesurés une fois par semaine dans ce jeu de données (75 122 performances).

#### Descendantes croisées

Les performances des descendantes croisées des reproducteurs de lignée pure (1 384 coqs) étaient également disponibles. Ces poules étaient réparties en 3 générations, élevées de 2010 à 2016. Chaque coq avait en espérance 45 descendantes croisées avec des femelles d'autres types génétiques (purs ou croisés). La partie femelle du pedigree était inconnue.

Les croisées étaient élevées en cages collectives de 12 demi-sœurs de père, de l'âge de 18 à 70 semaines. Les mesures de qualité d'œufs étaient effectuées 2 fois au cours de leur vie : à 50 semaines (41 343 performances) et 70 semaines (37 232 performances).

### Performances mixtes

Les performances obtenues en lignée pure ou sur descendance croisée ont été divisées en deux groupes, en fonction de l'âge des poules. Le premier groupe était composé de jeunes poules en début de carrière. Pour les performances en lignée pure, les données provenaient des cages collectives, mesurées à 30 et 50 semaines. Pour les performances des descendantes croisées, les performances étaient celles mesurées à 50 semaines. Cela a permis de composer 3 jeux de performances de poules jeunes : uniquement les performances en lignée pure, uniquement celles des descendantes croisées, et les deux précédents combinés.

Le second groupe était composé des performances des poules âgées, qui étaient en fin de carrière. Pour les performances en lignée pure, les données provenaient des cages individuelles, de 60 semaines à 80 semaines d'âge. Pour les performances des descendantes croisées, les performances étaient celles mesurées à 70 semaines. Comme précédemment, 3 jeux de performances ont été constitués : performances en lignée pure, performances des descendantes croisées et les deux combinés.

## 5.2.2 Génotypes

Les génotypes utilisés provenaient de 2 374 reproducteurs et reproductrices de la lignée pure, génotypés avec la puce 600K SNPs Affymetrix<sup>®</sup> Axiom<sup>®</sup>. Les deux plateformes de séquençage ont utilisé la suite d'analyse Axiom pour créer les fichiers de génotype au format PED (Purcell et al., 2007). L'uniformité des génotypes a été testé avec PLINK, qui a identifié 870 SNPs avec une inversion de brin visible, et Génotype Harmonizer (Deelen et al., 2014), qui a identifié 148 SNPs avec une inversion de brin invisible. PLINK a été utilisé pour corriger ces erreurs.

Chaque animal était génotypé sur 580 961 SNPs, et ont été retirés :

- Les individus avec un call rate  $<$  à 95% (aucun individu enlevé)
- Les SNPs avec une fréquence de l'allèle mineur (MAF)  $<$  0.05 (258 772 SNPs)
- Les SNPs avec un call rate  $<$ 95% (7 549 SNPs)
- Les SNPs avec des fréquences qui deviaient significativement de l'équilibre d'Hardy-Weinberg ( $P < 10^{-4}$ ) (12 538 SNPs)
- Les animaux avec des incompatibilités de pédigree (12 individus exclus)

La plupart des SNPs ont été enlevés car ils montraient une MAF de 0. Cela était attendu, étant donné que la puce à SNPs a été conçue à la fois pour les poules pondeuses et les poulets de chair, avec un ratio de 1 :2.

Après ces étapes de filtration, 302 102 SNPs ont été conservés, sur 1 214 coqs et 1 148 poules.

### 5.2.3 Caractères

Les différents caractères sont désignés en accord avec les ontologies des caractères pour l'animal d'élevage (Animal Trait Ontology for Livestock <http://www.atol-ontology.com/>). Cinq caractères ont été étudiés : le poids de l'œuf (Egg Weight, EW), la couleur de la coquille (Egg Shell Color, ESC), la solidité de la coquille (Egg Shell Strength), la hauteur de l'albumen (Albumen Height, AH), et la forme de la coquille (Egg Shell shape index, ESshape).

Les caractères se rapportant aux performances mesurées sur les animaux de lignée pure (caractères PB) sont préfixés par pb, ceux se rapportant aux performances des descendantes croisées (caractères CB) sont préfixés par cb.

#### Mesure des caractères

La première étape dans la récolte des performances de qualité d'œuf est la mesure du diamètre de l'œuf (SLE), et de EW (en grammes). Avec ces deux mesures est calculé l'ESshape, suivant la formule :  $ESshape = \frac{\frac{SLE}{10}}{\frac{EW}{10}^{\frac{1}{3}}}$ .

Ensuite, la couleur de la coquille est mesurée avec un chromamètre Minolta, et 3 paramètres sont enregistrés : la teneur en rouge  $a^*$ , la teneur en jaune  $b^*$  et la clarté  $L^*$  de la coquille. L'ESC est ensuite calculée telle que :  $ESC = 100 - (L^* - a^* - b^*)$ . Puis la solidité de la coquille est mesurée par une machine de compression. L'œuf est placé entre 2 plaques se déplaçant à vitesse constante. L'ESS est la force maximale enregistrée avant rupture de la coquille (en Newton, multiplié par 100). Enfin, l'œuf est cassé, et AH est mesuré (en millimètres) en utilisant un tripode. Cette mesure se fait à mi-chemin entre le bord du jaune et la partie épaisse du blanc d'œuf, au moyen d'un capteur électronique.

#### Choix des effets environnementaux du modèle et suppression des valeurs aberrantes

Pour chaque caractère, les mesures de qualité d'œuf ont été ajustées pour les effets environnementaux en utilisant la procédure GLM de SAS<sup>®</sup>. L'objectif était de vérifier quelle condition environnementale a un effet significatif sur le caractère, et d'enlever les données aberrantes.

Les effets environnementaux testés étaient le lot, la localisation de la cage dans le bâtiment pour le lot, i.e. la batterie à laquelle la cage appartenait et la localisation de la cage dans la batterie, le temps d'attente entre la ponte et la mesure de l'œuf (en jours), et l'âge de la poule (en semaines).

Pour tous les caractères, les effets du modèle étaient en dessous du seuil de significativité (établi à  $p < 0.05$ ), et ont été conservés dans le modèle. Les données brutes ont ensuite été corrigées en utilisant l'estimée de tous les effets, à l'exception de l'effet paternel. Les valeurs aberrantes, considérées comme telles quand présentant un écart à la moyenne supérieure à 5 écarts-types, ont été supprimées, et les données ont été centrées et réduites. Ce seuil assez élevé a été choisi afin de n'éliminer que les données résultant d'une erreur lors de la collecte des performances, et conserver le maximum de performances.

## 5.2.4 Réalisation des GWAS

## 5.2.5 Estimation des pourcentages de variance expliquée par les SNPs

Les GWAS ont été réalisées avec la suite BLUPF90 (Misztal et al., 2002), en évaluant dans un premier temps les GEBVs des individus avec un ssGBLUP (Misztal et al., 2009; Aguilar et al., 2010) utilisant le modèle décrit dans la section 5.2.3 en plus de l'effet génétique aléatoire de l'animal. Les effets des SNPs ont ensuite été estimés à partir des GEBVs (Legarra et al., 2018; Aguilar et al., 2019), en estimant le pourcentage de variance expliquée par groupe de 8 SNPs. Ce chiffre a été choisi car il nous permettait un bon recouvrement de nos blocs haplotypiques, ceux-ci ayant une taille moyenne de 5 SNPs (cf figure 5.1), tout en permettant une vision assez fine de la répartition des effets sur le génotype.

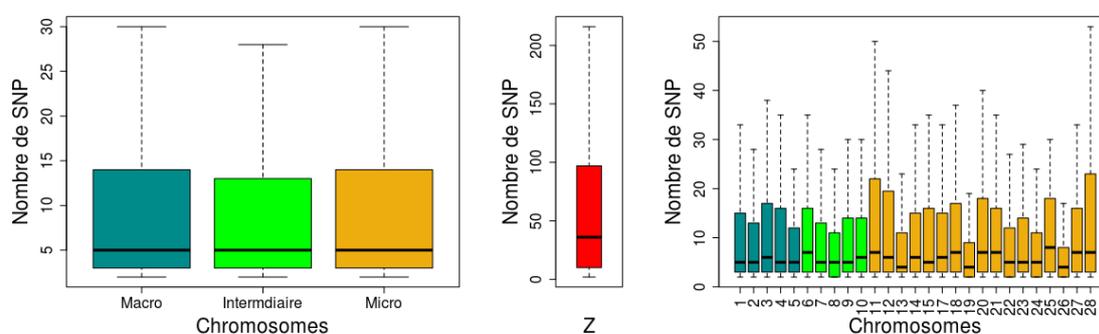


FIGURE 5.1 – Nombre de SNPs dans les blocs haplotypiques, en fonction du chromosome, pour la lignée Rhode Island et la puce à SNPs utilisée dans cette thèse.

Source : Hérault et al. (2018)

Trois GWAS différentes ont donc été réalisées avec les différents jeux de données composés pour chacun des deux âges (50 et 70 semaines). Une GWAS utilisant les performances des poules en lignée pure (GWAS pure), une utilisant les performances des descendantes croisées (GWAS croisée), et une utilisant conjointement les deux jeux précédents (GWAS mixte).

Chacune de ces GWAS a permis d'obtenir une estimation du pourcentage de variance expliquée par groupe de 8 SNPs consécutifs pour chacun des 5 caractères étudiés, PB et CB. Chacun de ces groupes était non chevauchant, et désigné par le numéro et la localisation du premier SNP le constituant.

### Détection des QTLs significatifs

Les QTLs significatifs dans ces GWAS ont été déterminés selon les règles suivantes :

- Au moins un groupe valant 0.02% de la variance totale
- L'intervalle de confiance du QTL était déterminé en prenant en compte tout groupe valant au moins 0.01% de la variance totale à moins de 800 SNPs de distance d'un autre valant au moins 0.01%. Le choix de cette valeur de 800

SNPs de distance a été fait de manière empirique, pour décrire au mieux les différents pics observés dans les GWAS.

- L'effet du QTL était déterminé en faisant la somme de l'ensemble des groupe entre les deux extrémités de l'intervalle de confiance
- Un QTL était significatif si son effet valait au moins 1% de la variance phénotypique totale.

Chaque QTL significatif a reçu un identifiant composé en deux parties séparées par un "-" :

La première partie reprend le caractère de qualité d'œuf sur lequel le QTL a une influence. La seconde partie est liée à la localisation du QTL sur les chromosomes : le chiffre, ou la lettre majuscule dans le cas des chromosomes sexuels. Le chromosome est suivi d'une lettre en minuscule si plusieurs QTLs concernant ce caractère ont été trouvés sur le chromosome. La lettre est attribuée en fonction de la position sur le chromosome, le QTL le plus proche du début du chromosome recevant la lettre a, le suivant la lettre b, etc.

Ainsi, par exemple, le QTL ESC-8a désigne un QTL ayant une influence sur la couleur de la coquille, se trouvant sur le chromosome 8, et il s'agit du plus proche du début du chromosome.

## 5.3 Résultats

### 5.3.1 QTLs significatifs sur les caractères de lignée pure

Les QTLs considérés comme significatifs sur les caractères mesurés en lignée pure sont décrits dans le tableau 5.1 pour les jeunes poules, et le 5.2 pour les poules âgées. Chaque QTL est nommé selon le caractère d'influence et par son chromosome de localisation, et si plusieurs QTLs sont présents sur le même chromosome, une lettre est ajoutée pour les différencier en fonction de leur localisation.

#### Différences entre GWAS pure et GWAS mixte

Des différences importantes sont visibles entre les QTLs trouvés en GWAS pure, et ceux trouvés en GWAS mixte, que ce soit sur les poules jeunes ou âgées.

La première différence se situe dans les QTLs trouvés dans chaque famille de GWAS (pure ou mixte). Seule une partie des QTLs trouvés pour un caractère sont communs entre GWAS pure et GWAS mixte. GWAS pures et mixtes chez les jeunes poules ont des QTLs en commun sur pbEW (1 QTL), pbESC (2 QTLs), pbESS (1 QTL) et pb AH (2 QTLs). Chez les poules âgées, ces QTLs communs sont trouvés sur pbEW (5 QTLs), pbESC (2 QTLs), pbESS (4 QTLs), pbAH (5 QTLs) et pbESshape (4 QTLs).

Des différences supplémentaires se situent sur ces QTLs communs.

La somme de la variance expliquée par les SNPs composant les QTLs est plus élevée en GWAS mixte, comme ESC-12 chez les poules jeunes, qui passe de 1.06% de

Caractère	QTL	GWAS pure			GWAS mixte		
		VE (%)	Intervalle de confiance (Mb)	Top groupe	VE (%)	Intervalle de confiance (Mb)	Top groupe
pbEW	EW-1b	1.34	63.271 - 66.482	65270780	-	-	-
	EW-1c	1.19	89.466 - 92.538	90947321	-	-	-
	EW-2	-	-	-	1.06	29.298 - 32.705	31515850
	<b>EW-3b</b>	2.64	27.181 - 29.755	29046030	1.86	28.952 - 31.495	29196093
	EW-8a	-	-	-	1.26	3.103 - 5.335	4548759
	EW-9	1.01	19.19 - 21.121	20455294	-	-	-
pbESC	ESC-5	-	-	-	1.12	49.615 - 53.636	52095119
	<b>ESC-12</b>	1.06	2.875 - 4.565	2946501	1.59	2.740 - 4.491	2946501
	<b>ESC-18</b>	1.25	3.021 - 4.606	3260324	1.19	3.003 - 4.566	3260324
pbESS	ESS-1a	1.03	48.197 - 51.021	50324720	-	-	-
	ESS-3	-	-	-	1.18	27.137 - 29.689	28363559
	<b>ESS-7</b>	3.67	19.583 - 25.024	24422814	1.09	21.885 - 24.671	22392211
	ESS-20	1.57	9.84 - 12.011	11684352	-	-	-
pbAH	AH-1b	1.09	68.152 - 70.433	69131439	-	-	-
	<b>AH-3</b>	3.33	102.073 - 107.374	105802676	4.45	102.544 - 107.821	102655131
	AH-4	1.13	35.592 - 40.089	38007448	-	-	-
	<b>AH-5b</b>	1.27	49.82 - 53.811	52320304	2.18	49.445 - 53.389	52209883
	AH-6	-	-	-	1.09	30.431 - 32.568	31266912
	AH-7	-	-	-	1.01	22.51 - 24.992	23117257
	AH-9	1.31	19.994 - 22.166	20667316	-	-	-
	AH-12	1.06	15.058 - 17.388	15466942	-	-	-
pbESshape	AH-28a	1.12	0.109 - 1.285	627077	-	-	-
	ESshape-1	1.01	82.512 - 85.14	83633513	-	-	-
	ESshape-3a	1.00	4.328 - 7.384	6540789	-	-	-
	ESshape-3b	1.15	35.47 - 38.306	35879012	-	-	-
	ESshape-5b	1.05	56.804 - 59.554	57227558	-	-	-

TABLE 5.1 – QTLs considérés comme significatifs pour les caractères en lignée pure, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet.

Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL.

Décrits sur les GWAS pure et mixte respectivement, chez les poules jeunes.

variance expliquée en GWAS pure, à 1.59% en GWAS mixte, avec un intervalle de confiance comparable et le même top groupe. À l'inverse, d'autres comme EW-2 chez les poules âgées, expliquent un pourcentage de variance plus fort en GWAS pure. Ces changements d'estimation dans les effets des QTLs ne sont pas systématiques, certains QTLs sont estimés avec un effet proche dans les deux GWAS.

Chez certains QTLs communs aux deux GWAS (pure et mixte), tel ESS-7 chez les jeunes poules, ou AH-6 chez les poules âgées, des différences se situent également dans leur intervalle de confiance, ainsi que dans la localisation de leur top groupe.

Caractère	QTL	GWAS pure			GWAS mixte		
		VE (%)	Intervalle de confiance (Mb)	Top groupe	VE (%)	Intervalle de confiance (Mb)	Top groupe
pbEW	<b>EW-1b</b>	1.45	62.754 - 66.082	62892107	1.4	62.754 - 66.082	65035775
	<b>EW-2</b>	1.44	29.298 - 32.705	30755167	1.1	29.301 - 32.713	31678965
	<b>EW-3b</b>	1.9	27.665 - 30.315	29239088	1.94	27.64 - 30.294	29200043
	EW-8a	-	-	-	1.03	3.095 - 5.314	4414447
	EW-8b	-	-	-	1.33	23.572 - 25.748	24244991
	<b>EW-17</b>	1.16	1.333 - 3.212	3092561	1.09	1.353 - 3.218	2206206
	<b>EW-Z</b>	1.6	0.668 - 73.395	64570157	1.94	2.3 - 73.387	11143131
pbESC	ESC-1a	-	-	-	1.01	65.357 - 68.228	66863660
	ESC-1b	1.56	183.615 - 188.091	184105125	-	-	-
	ESC-2	-	-	-	1.0	83.599 - 87.042	85850556
	ESC-4	-	-	-	1.08	63.445 - 67.706	63568725
	ESC-5	-	-	-	1.26	51.94 - 54.589	54082367
	<b>ESC-8</b>	1.40	25.377 - 27.513	26549534	1.15	25.0 - 27.159	26826597
	<b>ESC-12</b>	1.64	2.783 - 4.516	2946501	3.24	0.451 - 4.642	2956762
	ESC-21	1.07	0.633 - 2.061	897100	-	-	-
pbESS	ESS-1a	-	-	-	1.33	51.681 - 54.103	51970786
	<b>ESS-2a</b>	1.14	0.802 - 3.51	3107257	1.24	0.805 - 3.512	2972969
	ESS-5a	-	-	-	1.03	19.431 - 22.723	22350341
	<b>ESS-5b</b>	2.35	43.02 - 53.383	44804827	2.19	43.006 - 53.389	43979260
	<b>ESS-7</b>	1.01	21.885 - 24.671	22392211	1.26	21.855 - 24.656	22376695
	ESS-13	-	-	-	1.13	11.12 - 13.607	11470303
	ESS-28	1.17	0.278 - 1.642	924747	-	-	-
	<b>ESS-Z</b>	1.41	11.042 - 73.395	45875414	1.88	2.3 - 73.387	45842279
pbAH	<b>AH-1a</b>	1.14	36.53 - 39.236	37967598	1.21	36.53 - 39.236	37947669
	AH-2b	1.13	29.744 - 33.15	30782038	-	-	-
	<b>AH-3</b>	4.13	100.968 - 109.665	106713234	4.21	100.969 - 109.669	102655131
	<b>AH-5a</b>	2.09	21.03 - 25.382	22332876	2.4	20.856 - 25.168	22350341
	AH-5b	1.07	53.454 - 55.947	54870993	-	-	-
	<b>AH-6</b>	2.18	24.15 - 32.249	26038565	1.11	30.431 - 32.568	31283926
	AH-8	1.09	3.744 - 6.236	5517054	-	-	-
	<b>AH-11</b>	1.05	0.098 - 2.19	1366840	1.09	0.088 - 2.175	1351249
	AH-28b	-	-	-	1.07	3.494 - 4.661	4396144
pbESshape	<b>ESshape-5a</b>	1.12	29.543 - 33.103	31154837	1.14	29.547 - 33.112	31264783
	<b>ESshape-5b</b>	1.11	54.66 - 57.055	54870993	1.12	54.859 - 57.187	55193966
	<b>ESshape-6</b>	1.06	29.816 - 32.11	30715578	1.05	30.012 - 32.282	30580954
	<b>ESshape-Z</b>	1.33	17.496 - 73.395	46270822	1.76	11.042 - 73.387	45842279

TABLE 5.2 – QTLs considérés comme significatifs pour les caractères en lignée pure, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet.

Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL.

Décrits sur les GWAS pure et mixte, chez les poules âgées

### Architectures génétiques des caractères

Les deux GWAS pures et mixtes ayant permis de définir les QTLs significatifs sur les caractères, des différences sont apparues en termes d'architecture génétique

entre les mêmes caractères mesurés chez les poules jeunes ou âgées. Ces différences se situent dans les QTLs d'intérêt détecté, et les effets de ces QTLs.

Concernant les différences entre mêmes caractères, observées chez les jeunes poules et chez les poules âgées : les QTLs retrouvés ne sont pas forcément les mêmes. Ainsi, si 3 QTLs sont trouvés pour pbESC chez les jeunes poules, 8 sont trouvés chez les poules âgées, dont 2 sont communs avec ceux trouvés chez les jeunes poules, ESC-5 et ESC-12.

La variance expliquée par les SNPs du QTL est également quelque chose qui peut varier entre QTLs observés à la fois chez les jeunes poules, et chez les poules âgées. ESS-7 est ainsi estimé comme affectant 3.67% de la variance en GWAS pure chez les jeunes poules, chiffre qui descend à 1.01% chez les poules âgées. On observe sur ce même QTL une diminution de l'intervalle de confiance, passant de 5.441 Mb chez les jeunes poules à 2.786 Mb chez les poules âgées, et un décalage du top groupe de 2 Mb.

On a donc une variation de l'architecture génétique entre caractères observés chez les jeunes poules et ceux observés chez les poules âgées.

Par contre, les différentes architectures génétiques des différents caractères sont relativement semblables. Si l'on observe des différences dans le nombre et la localisation des QTLs significatifs trouvés, ce sont en général des QTLs avec une variance phénotypique expliquée par les SNPs assez faible.

Les profils de variance expliquée par groupe de 8 SNPs confirment cela (figures 5.2 et 5.3 pour les profils d'EW, les autres profils sont disponibles en annexe), les différents caractères de qualité d'œuf sont tous très polygéniques. Quelques QTLs, ESC-12, AH-3, AH-5b et ESS-7, sont composés de SNPs expliquant une part de variance bien plus importante que celle de leurs voisins, mais la somme de leurs SNPs n'explique qu'une faible part de la variance globale au final.

### **5.3.2 QTLs significatifs sur les caractères des descendantes croisées**

Les QTLs considérés comme significatifs sur les caractères des descendantes croisées sont décrits dans le tableau 5.3 pour les jeunes poules, et le 5.4 pour les poules âgées. Chaque QTL est nommé selon le caractère d'influence et par son chromosome de localisation, et si plusieurs QTLs sont présents sur le même chromosome, une lettre est ajoutée pour différencier les différentes localisations.

#### **Différences entre GWAS croisée et GWAS mixte**

Comme entre GWAS pure et mixte, des différences importantes sont visibles entre les QTLs trouvés en GWAS croisée, et ceux trouvés en GWAS mixte, que ce soit sur les poules jeunes ou âgées. Comme précédemment, les QTLs trouvés diffèrent en fonction de la famille de GWAS (croisée ou mixte).

Seule une partie des QTLs trouvés pour un caractère sont communs entre GWAS

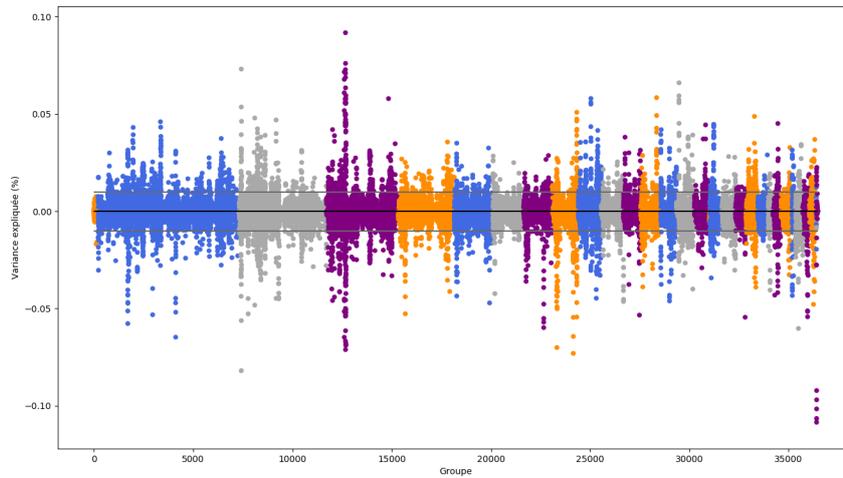


FIGURE 5.2 – Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

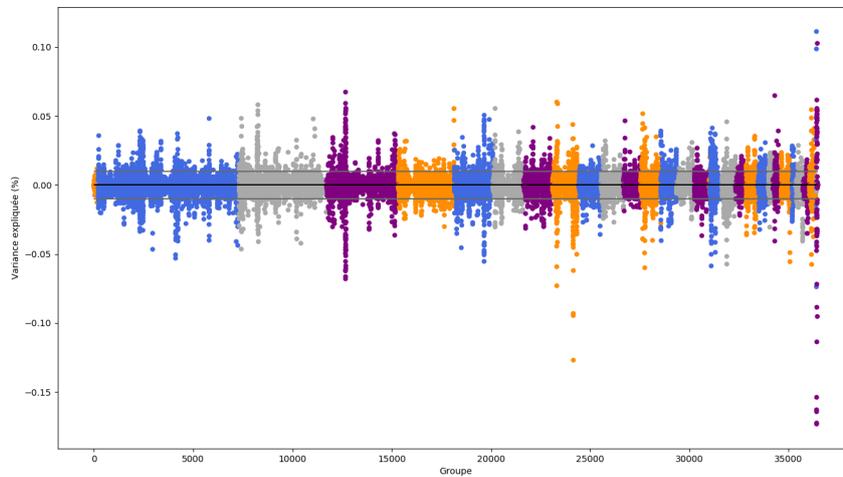


FIGURE 5.3 – Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

pure et GWAS mixte. GWAS pures et mixtes chez les jeunes poules ont des QTLs en commun sur cbEW (1 QTL), cbESC (1 QTL), cbESS (1 QTL), cbAH (4 QTLs) et cbESshape (2 QTLs). Chez les poules âgées, des QTLs communs sont également

Caractère	QTL	GWAS croisée			GWAS mixte		
		VE (%)	Intervalle de confiance (Mb)	Top groupe	VE (%)	Intervalle de confiance (Mb)	Top groupe
cbEW	<b>EW-1a</b>	1.19	43.466 - 47.07	45641104	1.16	43.466 - 47.07	44303499
	EW-2	-	-	-	1.09	29.298 - 32.705	31515850
	EW-3c	1.05	32.938 - 35.199	34394059	-	-	-
	EW-8b	-	-	-	1.2	23.6 - 25.779	24193054
	EW-17	-	-	-	1.03	0.665 - 2.807	2237045
	EW-27	-	-	-	1.17	3.218 - 4.913	3652189
cbESC	ESC-3	1.03	105.694 - 109.723	109543891	-	-	-
	ESC-5	-	-	-	1.01	48.174 - 52.99	52095119
	<b>ESC-12</b>	1.39	0.771 - 3.25	2946501	1.61	2.74 - 4.491	2946501
	ESC-13	1.18	9.654 - 12.143	11627534	-	-	-
cbESS	ESS-1c	1.04	172.476 - 174.888	173874039	-	-	-
	ESS-3	-	-	-	1.11	27.067 - 29.625	27233800
	ESS-4	1.13	30.315 - 34.019	32592335	-	-	-
	<b>ESS-9</b>	1.32	7.82 - 10.332	7882708	1.08	7.82 - 10.332	8900084
	ESS-Z	1.36	0.668 - 73.395	45875414	-	-	-
cbAH	<b>AH-2c</b>	1.11	56.874 - 60.374	59119442	1.04	60.55 - 64.458	61765405
	<b>AH-3</b>	2.87	102.073 - 107.339	102650937	3.76	102.544 - 107.821	102655131
	<b>AH-5b</b>	1.33	48.584 - 53.383	52267402	1.99	49.445 - 53.389	52209883
	AH-6	1.01	30.597 - 32.684	31258417	-	-	-
	<b>AH-7</b>	1.48	22.499 - 24.98	22968050	1.25	22.51 - 24.992	23117257
	AH-9	1.21	21.246 - 23.198	21397372	-	-	-
cbESshape	<b>ESshape-10</b>	1.01	6.672 - 8.994	7710851	1.11	6.665 - 8.983	7687022
	ESshape-12b	1.08	13.86 - 16.31	14041485	-	-	-
	<b>ESshape-Z</b>	2.49	11.042 - 73.395	45875414	1.2	17.496 - 73.387	45842279

TABLE 5.3 – QTLs considérés comme significatifs pour les caractères des descendantes croisées, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet.

Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL.

Décrits sur les GWAS pure et mixte, chez les poules jeunes

trouvés sur pbEW (4 QTLs), pbESC (2 QTLs), pbESS (3 QTLs), pbAH (3 QTLs) et pbESshape (1 QTL).

En plus des QTLs significatifs différents, des changements sont également visibles sur les QTLs communs.

La somme de la variance expliquée par les SNPs composant certains des QTLs communs est plus élevée en GWAS mixte, comme AH-3 chez les poules jeunes, qui passe de 2.87% de variance expliquée en GWAS croisée, à 3.76% en GWAS mixte, avec un intervalle de confiance comparable et le même top groupe. À l'inverse, d'autres comme EW-1c chez les poules âgées, expliquent un pourcentage de variance plus fort en GWAS pure. Ces changements d'estimation dans les effets des QTLs ne sont pas systématiques, certains QTLs sont estimés avec un effet proche dans les deux GWAS, comme ESS-13, dont la somme de la variance expliquée par les SNPs

Caractère	QTL	GWAS croisée			GWAS mixte		
		VE (%)	Intervalle de confiance (Mb)	Top groupe	VE (%)	Intervalle de confiance (Mb)	Top groupe
cbEW	EW-1a	-	-	-	1.0	62.754 - 66.082	65035775
	<b>EW-1c</b>	1.22	108.095 - 110.425	109740104	1.05	108.4 - 110.75	109604843
	EW-2	-	-	-	1.08	29.301 - 32.713	31678965
	EW-3a	1.01	0.886 - 4.717	3191502	-	-	-
	<b>EW-3b</b>	1.2	27.413 - 30.134	29196093	1.48	27.599 - 30.27	29200043
	EW-5	1.33	5.306 - 8.609	7156024	-	-	-
	EW-8a	-	-	-	1.03	3.375 - 5.69	4414447
	<b>EW-8</b>	1.3	23.668 - 25.838	24193054	1.1	24.043 - 26.174	24244991
	<b>EW-Z</b>	1.0	0.668 - 73.395	11041772	1.71	2.3 - 73.387	11143131
cbESC	ESC-1	-	-	-	1.26	65.357 - 68.228	66863660
	ESC-2	-	-	-	1.05	60.209 - 64.034	63568725
	ESC-3	1.01	105.599 - 109.632	105669076	-	-	-
	<b>ESC-5</b>	1.85	51.955 - 54.608	52320304	1.69	51.94 - 54.589	52209883
	<b>ESC-12</b>	1.15	2.636 - 4.433	2968001	2.60	0.451 - 4.642	2956762
	ESC-13	1.03	10.983 - 13.428	11627534	-	-	-
	cbESS	<b>ESS-1b</b>	1.02	64.679 - 67.588	64924393	1.06	63.933 - 66.951
ESS-2b		1.14	10.585 - 14.13	14032105	-	-	-
ESS-2c		1.12	33.77 - 36.865	34944331	-	-	-
<b>ESS-13</b>		1.07	10.664 - 12.99	11462023	1.06	10.738 - 13.132	11470303
ESS-17		-	-	-	1.09	5.707 - 7.235	6126694
<b>ESS-Z</b>		1.63	11.042 - 73.395	38194066	2.51	2.3 - 73.387	45842279
cbAH	AH-1a	-	-	-	1.09	35.635 - 38.716	37947669
	<b>AH-2a</b>	1.04	15.212 - 18.187	17170825	1.09	15.241 - 18.211	17634290
	AH-2d	1.11	90.363 - 93.379	90786427	-	-	-
	AH-3	-	-	-	2.41	102.296 - 107.741	102668880
	AH-5a	-	-	-	1.81	21.947 - 26.039	22350341
	<b>AH-5b</b>	1.23	49.904 - 53.918	52320304	1.33	50.049 - 54.05	52328316
	<b>AH-9</b>	1.05	21.367 - 23.285	23138451	1.37	22.266 - 24.066	23130993
cbESshape	ESshape-2	1.29	87.385 - 90.653	89650130	-	-	-
	ESshape-4	1.07	9.173 - 11.883	10503639	-	-	-
	ESshape-12a	1.1	0.653 - 3.149	2572229	-	-	-
	ESshape-17	1.02	2.313 - 3.92	2560243	-	-	-
	<b>ESshape-Z</b>	2.23	11.042 - 73.395	45875414	2.22	11.042 - 73.387	45842279

TABLE 5.4 – QTLs considérés comme significatifs pour les caractères des descendantes croisées, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet.

Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL.

Décrits sur les GWAS pure et mixte, chez les poules âgées

qui le compose est de 1.07 en GWAS croisée, et 1.06 en GWAS mixte.

Chez certains QTLs communs aux deux GWAS (croisée et mixte), tel ESshape-Z chez les jeunes poules, ou ESS-Z chez les poules âgées, des différences se situent également dans leur intervalle de confiance, ainsi que dans la localisation de leur top

groupe.

## **Architectures génétiques des caractères**

Les différences d'architectures génétiques des caractères chez les poules jeunes et les poules âgées, sur les QTLs d'intérêt détecté, et les effets de ces QTLs, sont également observés entre GWAS croisées et GWAS mixtes.

Concernant les différences entre mêmes caractères, observés chez les jeunes poules et chez les poules âgées : les QTLs identifiés dans chaque caractère changent. Ainsi, si 3 QTLs sont trouvés pour cbESshape chez les jeunes poules, 5 sont trouvés chez les poules âgées, dont un seul est commun avec ceux trouvés chez les jeunes poules, ESshape-Z.

Comme précédemment, la variance expliquée par les SNPs du QTL est également quelque chose qui peut évoluer entre QTLs observés à la fois chez les jeunes poules, et chez les poules âgées, de même que l'intervalle de confiance ou le top groupe. On retrouve donc une variation de l'architecture génétique entre caractères observés chez les jeunes poules et ceux observés chez les poules âgées et également sur les caractères des descendantes croisées.

Les différents caractères CB présentent également des architectures génétiques assez semblables. On observe quelques différences dans les QTLs significatifs trouvés, mais de manière générale, comme pour les caractères PB, ce sont des QTLs avec une variance phénotypique expliquée par les SNPs assez faible.

Les profils de variance expliquée par groupe de 8 SNPs confirment cela (figures 5.4 et 5.5 pour les profils d'EW, les autres profils sont disponibles en annexe), les différents caractères de qualité d'œuf sont tous très polygéniques.

Quelques QTLs, EW-Z, ESC-12, AH-2, AH-5b et ESshape-Z, sont composés de SNPs expliquant une part de variance bien plus importante que celle de leurs voisins, mais la somme de leurs SNPs n'explique au final qu'une faible part de la variance globale.

### **5.3.3 Différences entre caractère en lignée pure et sur descendantes croisées**

Comme entre les caractères chez les poules jeunes et chez les poules âgées, on observe des différences entre les architectures génétiques des caractères PB et CB, à âge équivalent.

Les QTLs significatifs identifiés entre les deux types de caractères (PB ou CB) varient, et seule une partie est commune entre les deux types de caractères (voir tableau 5.5).

Il peut y avoir des légers décalages en termes d'intervalle de confiance (comme pour le QTL AH-5a en GWAS mixte chez les poules jeunes, passant de 4,312 Mb en PB à 4,092 Mb en CB), ou de top groupe (chez les poules jeunes, le top groupe de ESC-5 en GWAS mixte est décalé de 2Mb entre le caractère PB et CB), mais

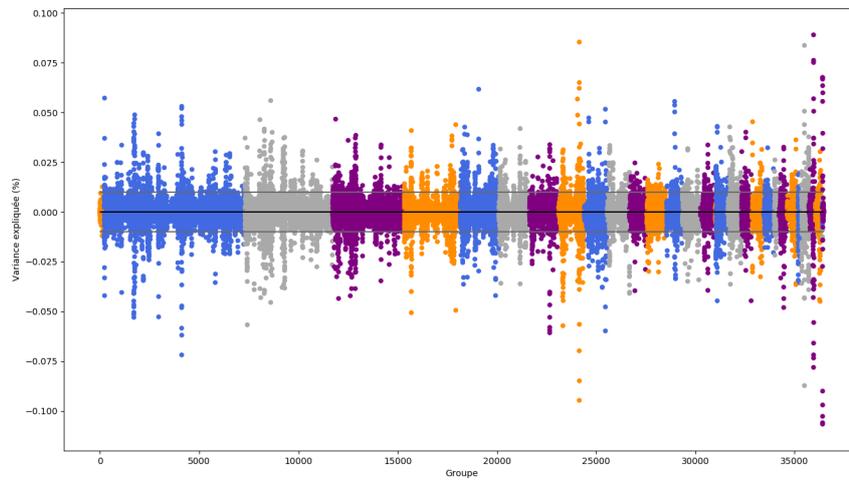


FIGURE 5.4 – Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

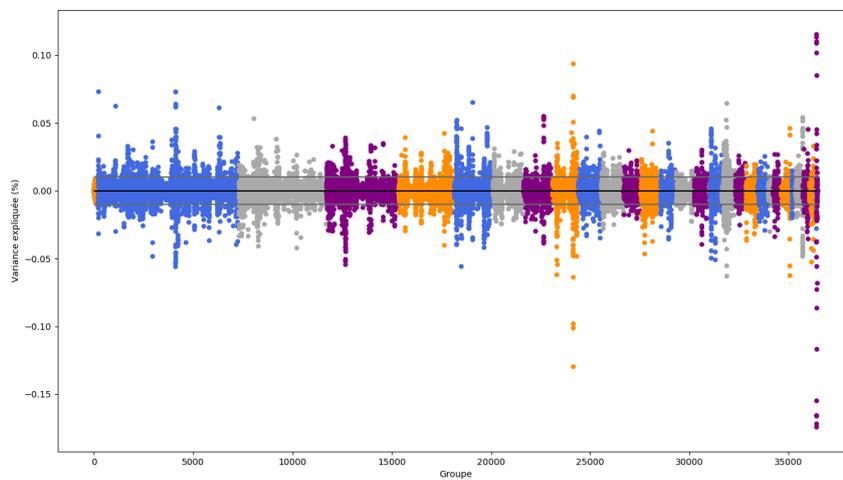


FIGURE 5.5 – Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

dans l'ensemble ces deux paramètres s'avèrent proches entre QTLs communs aux caractères PB et CB.

Caractère	Nombre de QTLs chez poules jeunes			Nombre de QTLs chez poules âgées		
	PB	CB	Communs	PB	CB	Communs
EW	6	6	3	7	9	5
ESC	3	4	2	8	6	4
ESS	4	5	1	8	7	2
AH	9	6	5	9	7	4
ESshape	4	3	0	4	5	1

TABLE 5.5 – QTLs communs entre caractères PB et CB

## 5.4 Discussion

### 5.4.1 Résultats obtenus par rapport à la littérature

Comme il a été détaillé dans le chapitre 1, les caractères de qualité d’œufs ont déjà été caractérisés dans la bibliographie par un certain nombre de QTLs, résumés dans le tableau 5.6.

Sur les caractères de lignée pure, 3 QTLs chez les poules jeunes et 7 chez les poules âgées sont déjà connus dans la bibliographie.

Chez les poules jeunes, en GWAS pure, ces QTLs sont EW-1a, ESC-12 et AH-3, alors qu’en GWAS mixte, ce sont EW-2, ESC-12 et AH-3. Chez les poules âgées, en GWAS pure et mixte ce sont EW-1a, EW-2, ESC-5, ESC-12, ESS-Z, AH-1a et AH-3 qui sont retrouvés.

Sur les caractères des descendantes croisées, des QTLs correspondent également à ceux décrits dans la littérature.

Sur les données des jeunes poules, 4 QTLs sont communs avec la littérature. Les deux GWAS ont en commun EW-1a, ESC-12 et AH-3, la GWAS croisée détectant également ESS-Z et la mixte EW-2.

Chez les poules âgées, 3 QTLs connus sont trouvés en GWAS pure, ESC-12, ESS-Z et EW-Z. La GWAS mixte en détecte 2 supplémentaires, EW-2 et AH-3.

Romé et al (2015) avaient également réalisé des GWAS sur les données de performances des descendantes croisées utilisées ici.

Quatre QTLs significatifs trouvés sur les caractères des descendantes croisées ont été trouvés en commun, EW-Z, ESC-12, ESS-9 et ESshape-10, sachant que 8, 13, 2, 5 et 7 QTLs avaient été décrits comme significatifs pour respectivement EW, ESC, ESS, AH et ESshape. Les GWAS sur les performances des poules jeunes retrouvent EW-Z et ESC-12, et celles sur poules âgées ESC-12, ESS-9 et ESshape-10.

Sur les caractères de lignée pure, 3 QTLs sont communs, EW-Z, ESC-12 et ESS-28. Chez les poules jeunes, seul ESC-12 est trouvé, mais tous sont trouvés chez les poules âgées.

L’étude de Romé et al. (2015) a été réalisé en utilisant la méthode Müller (Müller et al., 2011) pour estimer les effets des SNPs et la significativité des QTLs. Cette méthode estime directement les effets des SNPs à partir du génotype et du phénotype des individus, et non pas à partir des GEBVs, ce qui explique les différences en termes de résultats.

Une partie des QTLs significatifs identifiés dans les différentes GWAS est donc

Caractère	Chromosome	Position	Publications
EW	1	25Mb	Tuiskula-Haavisto et al. (2002); Kerje et al. (2003), Sasaki et al. (2004); Honkatukia et al. (2005), Schreiweis et al. (2006); Wright et al. (2008), Wolc et al. (2012); Goraga et al. (2012), Wolc et al. (2014); Romé et al. (2015)
		50Mb	
	2	30Mb	
	4	50Mb	
	Z	75Mb	
ESC	3	80Mb	Wardecka et al. (2002); Schreiweis et al. (2006), Wolc et al. (2014); Romé et al. (2015), Tuiskula-Haavisto et al. (2018); Liu et al. (2018)
		74Mb	
	12	3Mb	
ESS	1	80Mb	Tuiskula-Haavisto et al. (2002); Sasaki et al. (2004), Dunn et al. (2009); Tuiskula-Haavisto et al. (2011), Wolc et al. (2014); Sun et al. (2015), Tuiskula-Haavisto et al. (2018)
	2	102Mb	
	4	50Mb	
	6	20Mb	
	Z	10Mb	
AH	1	45Mb	Wardecka et al. (2002); Hansen et al. (2005), Liu et al. (2011a); Goto et al. (2014), Wolc et al. (2014)
		125Mb	
	2	85Mb	
	3	100Mb	
	4	73Mb	
ESshape	2	108Mb	Hansen et al. (2005); Dunn et al. (2009), Yao et al. (2010); Tuiskula-Haavisto et al. (2018)
		4	
	8	20Mb	

TABLE 5.6 – QTLs affectant les caractères de qualité d’œuf, d’après Romé and Le Roy (2016). La position correspond à la localisation moyenne du pic des QTL.

déjà connue dans la littérature. Cela tend à confirmer à la fois la cohérence des résultats trouvés avec ce qui est déjà connu, et la fiabilité de ces résultats.

### 5.4.2 Différences d’architecture entre caractères

Les différents caractères de qualité d’œufs présentent des architectures relativement proches, très polygéniques. Certains QTLs, comme ESC-12 ou AH-3 se distinguent fortement, du fait de SNPs les composant avec un effet sur la variance estimée bien plus élevé que les autres.

Ces différences pourraient correspondre à celles observées en termes de précision lors des évaluations, ESC étant notamment un caractère estimé avec une précision assez faible dans le chapitre 3. En effet, le modèle infinitésimal du GBLUP suppose que chacun des SNPs explique une part de la variance additive, et une architecture déviant fortement par rapport à ce modèle pourrait impacter la précision de l’évaluation. Cependant, ce type de QTL est également observé dans le caractère AH, qui lui-même est estimé avec une très bonne précision dans le chapitre 3, dans les mêmes conditions. Les différences de précision observées entre les caractères sont donc probablement liées à un autre aspect du caractère, peut être à des effets de dominance, ou d’épistasie. Ces résultats confirment ce qu’avait conclu Daetwyler et al. (2010a,

2010b), sur le fait que le GBLUP n'est que peu sensible à l'architecture du caractère évalué.

Entre caractères mesurés sur les poules jeunes et les poules âgées, que ce soit sur la lignée pure ou sur leurs descendants croisés, on observe d'importantes variations dans les QTLs, en termes de variabilité expliquée ou de localisation. Le caractère mesuré à 50 semaines a une architecture génétique assez différente de celle du caractère à 70 semaines, ce sont deux caractères bien distincts.

De même, les caractères PB et CB présentent, malgré une base commune, des différences dans leur architecture. Cela confirme les résultats obtenus dans le chapitre 4 : les caractères mesurés en lignée pure et en descendance croisée ne peuvent être considérés comme identiques. C'est une conséquence du croisement de différentes lignées pures entre elles. Ces lignées étant chacune optimisée pour la production d'un type d'œuf particulier, elles présentent des fréquences alléliques qui leur sont propres, pouvant aller jusqu'à la fixation d'allèles favorables. Les descendants croisés de deux lignées auront des fréquences alléliques intermédiaires, et donc une architecture génétique différente de celles de leurs géniteurs de lignée pure.

### 5.4.3 Utilisation de la GWAS mixte

L'idée derrière la GWAS mixte était de voir si l'utilisation combinée de ces deux sources d'informations pouvait permettre d'améliorer la détection des QTLs.

Concernant la fiabilité de ce type de GWAS, dans tous les cas, on retrouve des QTLs qui sont connus dans la littérature. On peut cependant observer quelques différences sur la localisation du top groupe de ces QTLs, notamment sur les QTLs AH-3 (pour le caractère PB) et ESS-Z (pour le caractère CB), où le top-SNP est plus proche de celui indiqué en moyenne dans la littérature.

À priori, la GWAS mixte, par l'utilisation conjointe des deux jeux de données, permettrait d'augmenter la puissance de l'estimation, et d'ainsi détecter de nouveaux QTLs.

Les QTLs significatifs et inconnus de la bibliographie constituent cependant la plus grande part des QTLs trouvés dans les GWAS. Parmi ces QTLs, certains ne sont détectés qu'en GWAS mixte, d'autres uniquement en GWAS pure ou croisée. Pour les QTLs trouvés uniquement en GWAS mixte, et pas en GWAS pure ou croisée, cela peut être causé par l'ajout des données supplémentaires. L'ajout de performances, dont la corrélation génétique est comprise entre 0.75 et 0.80 (cf chapitre 4), permet une augmentation de la puissance de l'estimation, et une détection de nouveaux QTLs.

Concernant les QTLs trouvés en GWAS pure et croisée, et pas en GWAS mixte, il est possible que ce soit des zones dont les effets ne sont plus observés dans le jeu de données rajouté, et la forte corrélation génétique induit une réduction des effets lors de la GWAS mixte, ce qui fait que l'on ne détecte plus ces QTLs.

En tout cas, la prise en compte des performances en lignée pure conjointement à celles des descendantes croisées permet aux GWAS mixtes de détecter de nouveaux QTLs, jusqu'ici inconnus sur les caractères de qualité d'œuf, et qui n'étaient pas

détectés dans les GWAS pure et croisée.

## 5.5 Conclusion

Les différentes GWAS réalisées ont permis de conclure sur plusieurs points. Tout d'abord, les différences observées dans les structures des caractères n'ont pas permis d'expliquer les différences de précision lors des évaluations constatées dans les chapitres 3 et 4.

Ensuite, l'architecture génétique est assez différente entre caractères PB et CB, ce qui confirme notre conclusion de l'article 4 les concernant, préconisant de les évaluer séparément.

Enfin, la GWAS mixte paraît être une méthode intéressante pour détecter de nouveaux QTLs et mieux appréhender la réalité biologique derrière les caractères de qualité d'œuf.

Les étapes suivantes pour approfondir ces travaux seront de localiser sur le génome les différents QTLs, afin d'identifier de potentiels gènes candidats, positionnels ou fonctionnels, et d'estimer les parts de variance expliquées par les effets de dominance pour les différents caractères de qualité d'œuf.

# Chapitre 6

## Discussion générale et perspectives

Les travaux effectués dans cette thèse ont permis d'apporter un nombre conséquent d'informations sur différents aspects liés à la précision de l'évaluation génomique chez la poule pondeuse.

Les travaux sur la précision de l'évaluation génomique sur lignée pure ont permis de conclure de l'intérêt de celle-ci pour la filière.

Sur la voie mâle, la sélection génomique à la naissance des candidats présente un intérêt indéniable. La précision est suffisamment améliorée pour se retrouver équivalente à celle obtenue en évaluation génétique sur collatéraux, et cela permet de fortement augmenter l'intensité de sélection et de diviser par 3 l'intervalle de génération. L'utilisation de l'évaluation génomique à la naissance permettrait donc de gagner sur tous les composants du progrès génétique dépendants de la méthode de sélection. Ces résultats ont confirmé le potentiel de cette approche auprès de nos partenaires de Novogen, qui ont commencé à l'appliquer sur leurs mâles reproducteurs, avec des premiers résultats très encourageants.

Sur la voie femelle, l'intérêt d'utiliser la sélection génomique à la naissance est plus discutable. Les performances propres apportent énormément d'informations, ce qui donne un très grand écart de précision entre évaluation génomique à la naissance et évaluation génétique sur collatéraux. De plus la voie femelle ne gagnerait pas en intensité de sélection, seulement en intervalle de génération.

La valorisation des performances des descendantes croisées en évaluation mixte a permis d'obtenir des informations supplémentaires sur comment améliorer la précision de l'évaluation sur la voie mâle.

Concernant la corrélation entre caractères en lignée pure et ceux observés sur la descendance croisée, nous avons pu constater que si ces caractères sont fortement liés, on ne peut pas les considérer comme étant identiques. Sélectionner des individus sur leurs caractères en lignée pure ne sera donc pas optimal si l'objectif de sélection se porte sur les performances des descendantes croisées. Il faudra donc sélectionner les reproducteurs sur le caractère correspondant le mieux à l'objectif de sélection. En tout cas, l'utilisation conjointe en évaluation des performances en lignée pure avec celles des descendantes croisées, s'est avérée avoir un impact positif sur la précision de l'évaluation. Valoriser les performances des descendantes croisées en évaluation

est une piste viable pour optimiser encore davantage le progrès génétique sur la voie mâle. Ceci, combiné avec les gains possibles en sélectionnant les reproducteurs à la naissance, peut contribuer à faire fortement augmenter le progrès génétique obtenu par la sélection.

Les GWAS effectuées sur nos données nous ont permis de répondre aux questionnements liés à l'architecture génétique de nos caractères. Elles nous ont permis de confirmer que les caractères mesurés en lignée pure présentent une architecture génétique différente de leurs équivalents mesurés sur les descendantes croisées, comme les corrélations génétiques l'avaient indiqué dans le chapitre 4. Les résultats ont également montré que les différences de précision observées entre les différents caractères lors des évaluations n'étaient pas liées à l'architecture génétique des caractères, ceux-ci présentant tous une architecture génétique très multigénique.

À cette occasion, nous avons également testé le potentiel d'une GWAS utilisant conjointement les performances des poules de lignées pures et celles de leurs descendantes croisées. Les résultats de ces GWAS se sont révélés très intéressants, mettant en lumière de nouveaux QTLs qui n'étaient pas détectés jusqu'alors.

Ces trois travaux principaux ont été détaillés dans leurs chapitres respectifs, mais des questions supplémentaires se sont posées au vu des résultats obtenus.

Concernant la sélection génomique à la naissance, se posait la question de la diversité génétique des individus sélectionnés par rapport à celle des individus sélectionnés par évaluation génétique sur collatéraux. C'est un aspect assez important, car la diminution de l'intervalle de génération par l'utilisation de l'évaluation à la naissance accélère l'évolution temporelle de la consanguinité. En théorie, la sélection génomique étant capable de mieux différencier les individus intra-famille, sélectionnerait plus d'individus non apparentés à quantité d'informations phénotypiques équivalentes. Mais dans le cas de figure qui nous intéresse, la quantité d'informations phénotypiques n'est pas la même.

Afin d'observer cela, ont été comparés le nombre de familles différentes sélectionnées en évaluation génétique sur collatéraux et celui en évaluation génomique sur ascendance, sur les caractères de lignée pure. Les (G)EBVs de tous les mâles des lots à l'origine de C1 (672 coqs, population notée M1), ceux des lots à l'origine de C2 (650 coqs, population notée M2), ont été considérés, et les 150 meilleurs pour chaque caractère ont été sélectionnés. Le nombre de familles de père différentes dans chaque cas de figure a ensuite été déterminé. Un coq pouvant être apparié à plusieurs femelles (mais pas le contraire), cela permettait de grouper tous les individus partageant au moins un parent en commun.

Les résultats varient en fonction des caractères, mais on peut y observer que la sélection génomique sur ascendance a un nombre de familles différentes sélectionnées supérieur à celui en sélection génétique sur collatéraux chez les mâles. Avec les individus M1, l'évaluation génétique sur les données en CC (CI) sélectionne en moyenne 59 (63) familles différentes pour 150 individus, là où l'évaluation génomique en sélectionne 66 (76). Avec les individus M2, l'évaluation génétique sur les données en CC (CI) sélectionne en moyenne 61 (67) familles différentes pour 150 individus, là

où l'évaluation génomique en sélectionne 66 (74). L'usage de la sélection génomique à la naissance chez les mâles ne devrait donc pas provoquer une augmentation de la consanguinité par génération, par rapport à ce qui se produisait en évaluation génétique sur collatéraux.

Un autre aspect qui posait question était la variabilité des résultats entre caractères. Une des hypothèses pouvant expliquer les disparités de précision observées entre caractères était l'impact d'effets de dominance sur les différents caractères. Les évaluations étant faites à partir de modèles additifs, ces effets de dominance ne sont pas pris en compte, ce qui pourrait en partie expliquer les différences de précision observées entre les caractères.

Une étude est en cours sur les effets de dominance sur les caractères de qualité d'œuf. Pour cela, étant donné la nécessité d'utiliser des animaux à la fois génotypés et phénotypés, nous avons utilisé une population de 1 148 femelles génotypées, avec 10 509 performances dans le jeu de données CI. Les génotypages ont ensuite été utilisés pour générer les matrices de relation additive (G) et de dominance (D) en utilisant la méthode décrite par (Vitezica et al., 2013). Les matrices G et D obtenues ont ensuite été inversées pour pouvoir être utilisées directement dans le programme REMLF90. Ce programme a ensuite été utilisé afin d'estimer les parts de variance expliquées par les effets de dominance et ceux d'additivité, en utilisant en plus des effets déjà utilisés dans les évaluations précédentes une covariable de consanguinité. Les premiers résultats obtenus sur les caractères de qualité d'œuf ont montré que la dominance n'expliquerait qu'une faible part de la variance phénotypique totale. Sur les 5 caractères observés, les effets de dominance expliquent environ 3% de la variance phénotypique totale. Ces effets de dominance peuvent cependant représenter une part importante de la variance génétique des individus, entre 3% et 35% en fonction du caractère étudié.

L'étape suivante sera d'estimer les corrélations génétiques des effets d'additivité et de dominance entre les différents caractères de qualité d'œuf. Il serait également très instructif d'estimer les effets d'épistasie sur ces caractères de qualité d'œuf, ce type d'effet pouvant également impacter la précision de l'évaluation. Il est aussi envisagé d'étudier l'effet de l'intégration des effets de dominance en évaluation ssGBLUP mixte des performances de lignées pures et de descendance croisée (Vitezica et al., 2016; Xiang et al., 2016). L'objectif étant d'évaluer le potentiel d'une telle évaluation pour obtenir des individus croisés avec de meilleures performances.

D'autres pistes sont envisageables dans la poursuite de ce travail sur la précision de l'évaluation génomique chez la poule pondeuse.

Tout d'abord, les caractères de qualité d'œuf étudiés ici sont des caractères avec une héritabilité élevée, allant de 0.40 à 0.60. Les caractères de qualité d'œuf ne représentaient qu'une partie des caractères mesurés dans les jeux de données utilisés ici, et certains (e.g. le taux de ponte) sont connus pour avoir une héritabilité assez faible. Étant donné que l'héritabilité à un effet sur la précision de l'évaluation, il serait très instructif de refaire ces expérimentations sur ces caractères moins héréditaires, mais qui revêtent une importance certaine dans l'objectif de sélection. Cela permettrait d'estimer plus précisément le progrès génétique que l'on peut

attendre de l'utilisation de l'évaluation génomique à la naissance.

Ensuite, à cause de limitations dans le jeu de données à notre disposition, il n'a pas été possible de réaliser l'étude sur l'utilisation conjointe des données de lignée pure et des descendantes croisées sur la voie femelle. Les résultats devraient être les mêmes, avec une augmentation de la précision, mais chiffrer l'étendue de l'amélioration possible serait pertinent pour mieux estimer les conséquences (positives ou négatives) que pourraient produire le passage de l'évaluation génétique sur collatéraux à la sélection génomique à la naissance pour la voie femelle.

Il serait également assez intéressant de tester l'évaluation conjointe des lignées pures et descendantes croisées en ayant accès au pedigree des mères, pour déterminer la précision de l'évaluation dans le cas où l'on a accès aux deux lignées pures à l'origine des descendantes croisées.

Enfin, il serait également très instructif, maintenant que l'architecture génétique des caractères est mieux connue dans la population étudiée, d'approfondir le travail sur les zones du génome ayant un effet sur les caractères de qualité d'œuf. Il serait intéressant de séquencer des individus provenant de cette population, pour aller rechercher des polymorphismes causaux. Cela permettrait de progresser dans la connaissance des mécanismes génétiques impliqués dans la qualité de l'œuf.

Sur les aspects plus liés à la filière, plusieurs questions se posent également sur l'impact que va avoir la mise en place de l'évaluation génomique à la naissance sur la voie mâle. Celle-ci semble très avantageuse, permettant une diminution importante de l'intervalle de génération, une forte augmentation de l'intensité de sélection, pour une précision comparable. Elle permet également de prendre plus facilement en compte des caractères qu'il n'était pas envisageable de tester sur des candidats à la sélection, comme des caractères de robustesse ou de santé, qui sont très intéressants dans une filière où les animaux doivent s'adapter à des conditions d'élevage variées. Cependant, si son utilisation permet de réduire l'intervalle de génération, il faut prendre en compte la nécessité de garder à jour la population de référence, et trouver un compromis entre la réduction au minimum de l'intervalle de génération et le temps nécessaire à la récolte de nouvelles performances.

De plus, si le génotypage permet d'augmenter la précision de l'évaluation et de différencier les individus d'une même famille, cela a un coût qui peut être important. Une solution pour le diminuer est de recourir à des puces à basse densité, moins chères, et de les imputer vers des puces HD. Mais le DL de la population évoluant avec le temps, il faudra également déterminer quand tenir à jour le modèle en génotypant de nouveau des individus à haute densité, afin de garder des imputations de qualité.

# Annexes

## 6.1 GWAS pure vs GWAS mixte

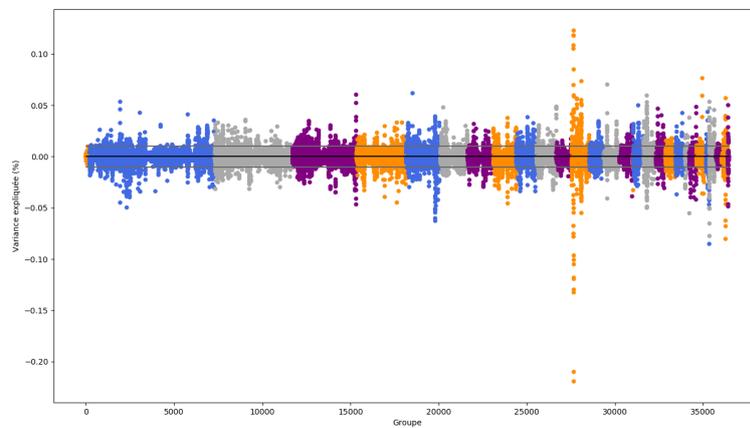


FIGURE 6.1 – Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

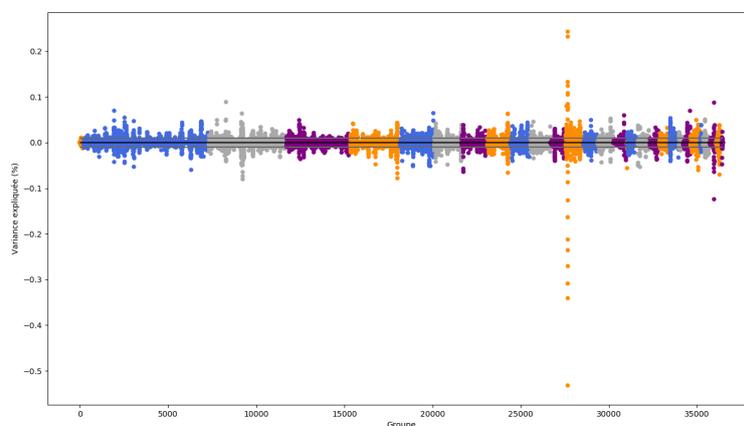


FIGURE 6.2 – Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

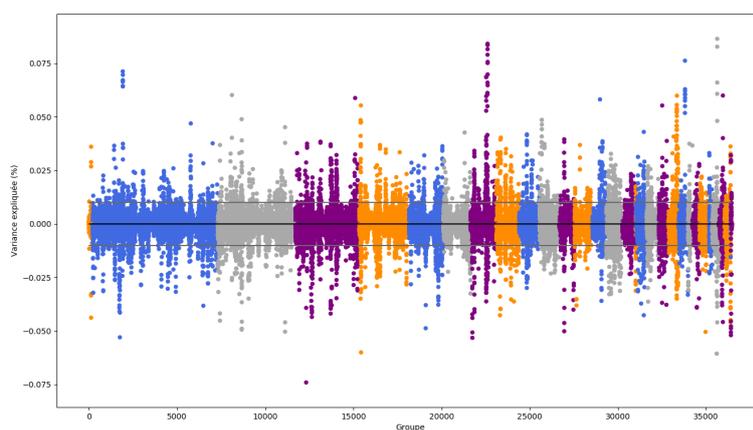


FIGURE 6.3 – Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

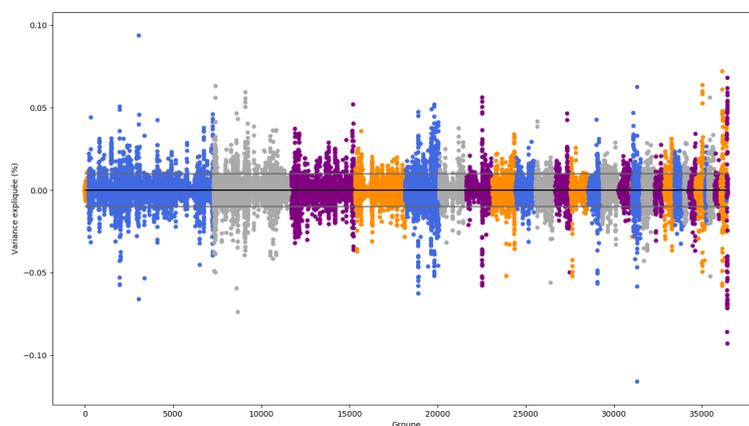


FIGURE 6.4 – Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

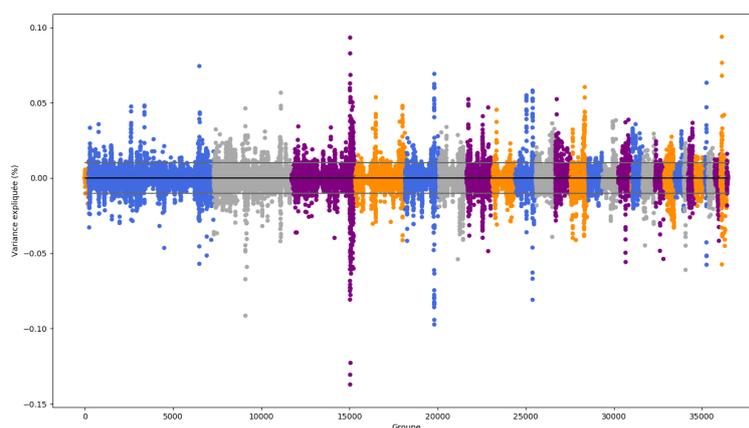


FIGURE 6.5 – Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

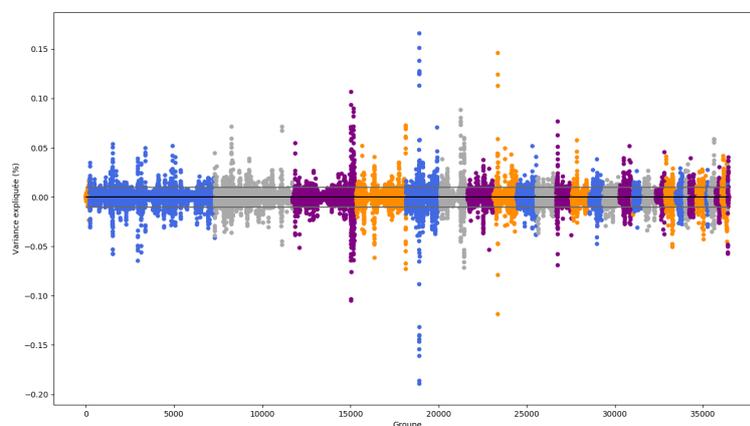


FIGURE 6.6 – Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

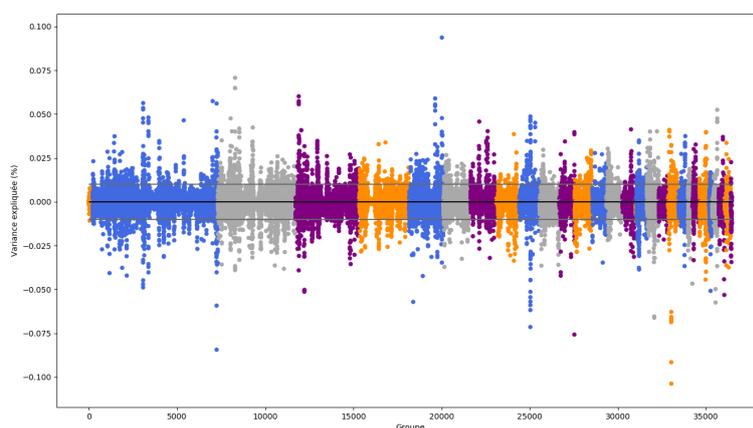


FIGURE 6.7 – Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

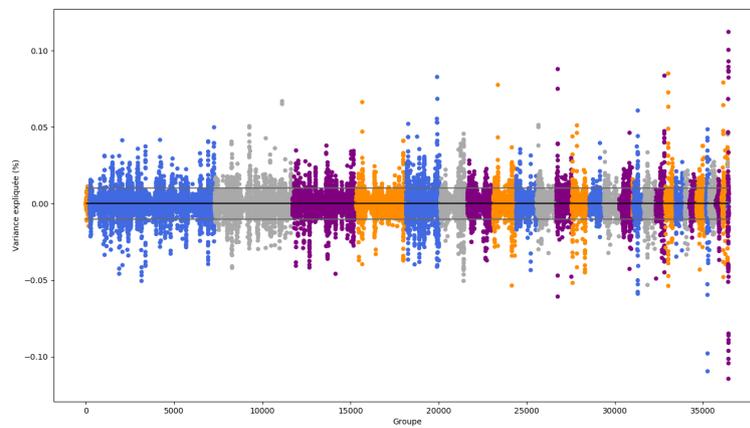


FIGURE 6.8 – Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte

## 6.2 GWAS croisée vs GWAS mixte

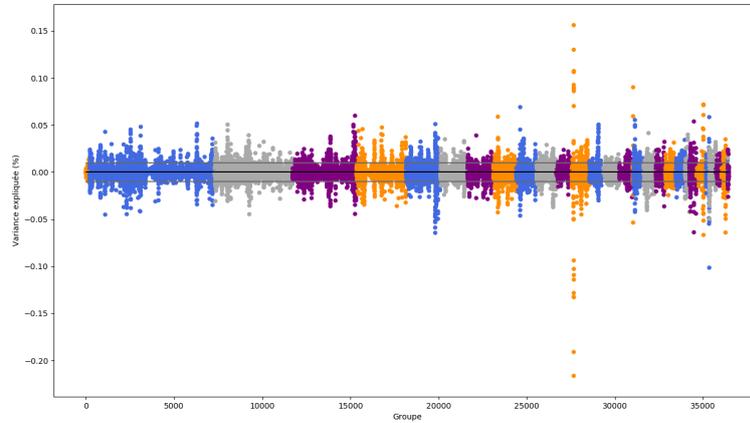


FIGURE 6.9 – Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

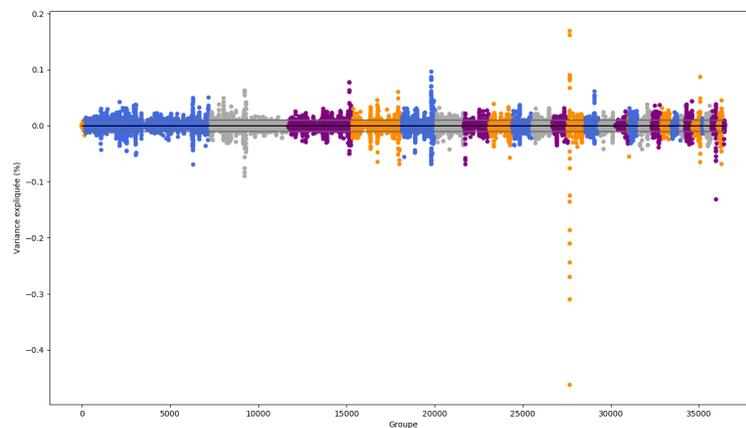


FIGURE 6.10 – Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

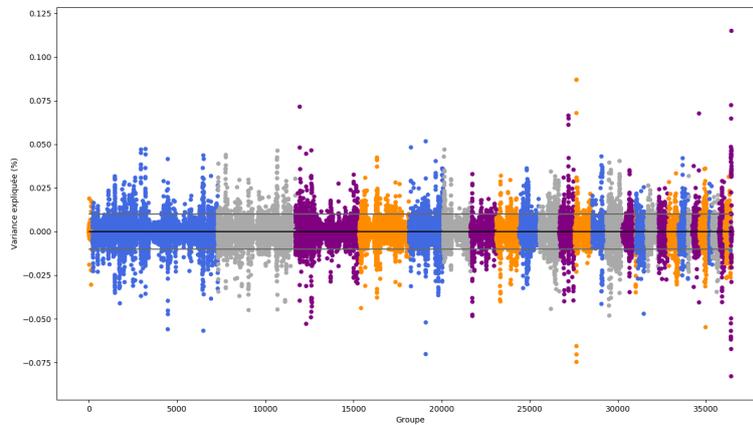


FIGURE 6.11 – Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

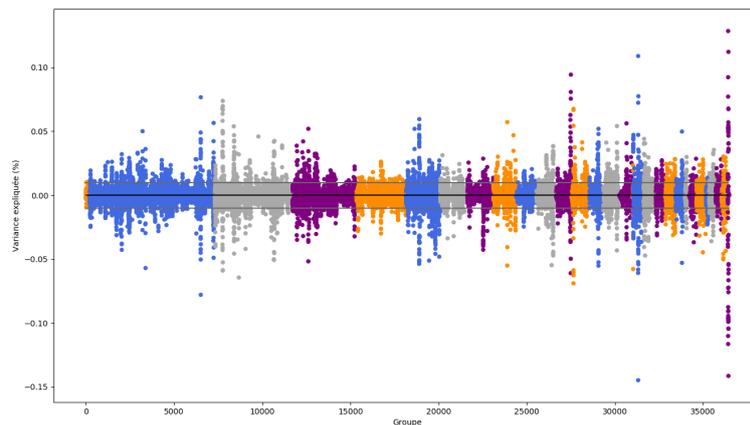


FIGURE 6.12 – Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

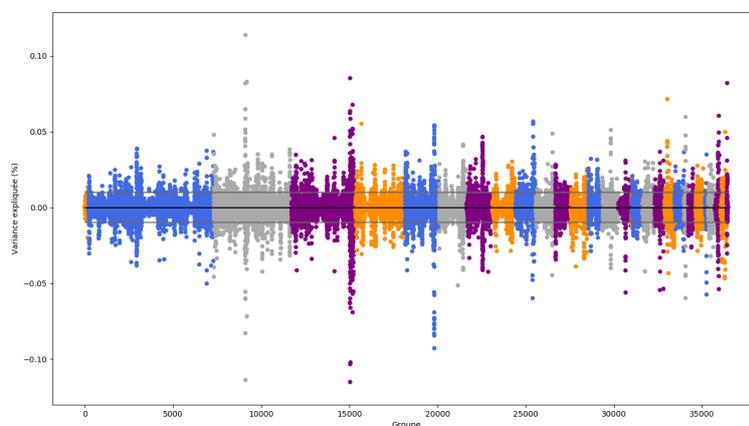


FIGURE 6.13 – Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS croisée

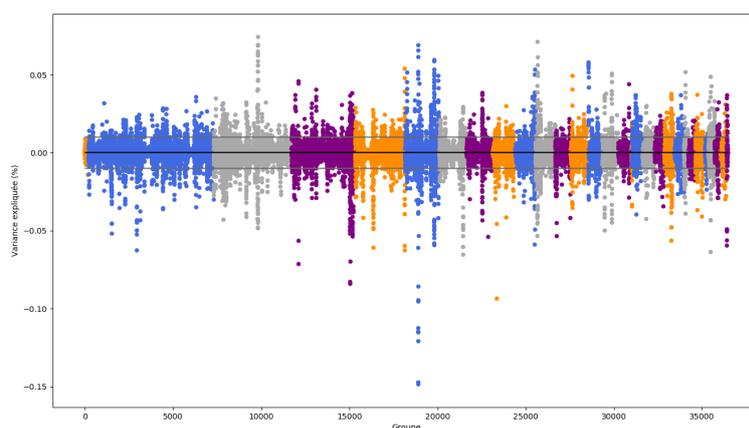


FIGURE 6.14 – Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

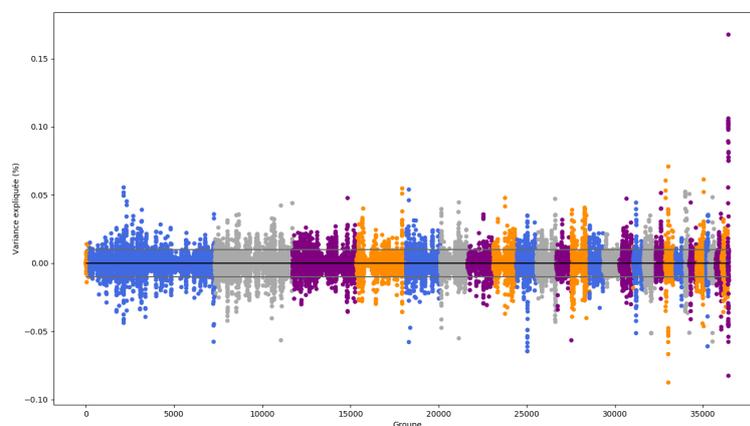


FIGURE 6.15 – Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

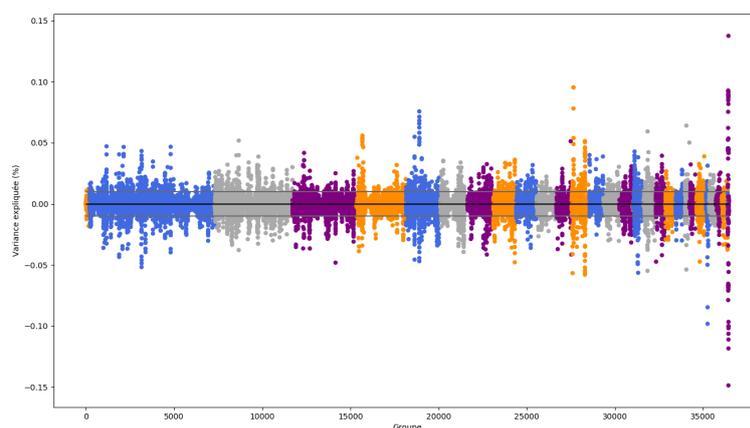


FIGURE 6.16 – Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte

# Bibliographie

- I. Aguilar, I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. Hot topic : a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, 93(2) :743–752, Feb. 2010. ISSN 1525-3198. doi : 10.3168/jds.2009-2730.
- I. Aguilar, A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genetics, Selection, Evolution : GSE*, 51, June 2019. ISSN 0999-193X. doi : 10.1186/s12711-019-0469-3. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6584984/>.
- F. G. Asoro, M. A. Newell, W. D. Beavis, M. P. Scott, and J.-L. Jannink. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *The Plant Genome*, 4(2) :132–144, July 2011. ISSN 1940-3372. doi : 10.3835/plantgenome2011.02.0007. URL <https://dl.sciencesocieties.org/publications/tpg/abstracts/4/2/132?q=publications/tpg/abstracts/4/2/132>.
- J. W. M. Bastiaansen, A. Coster, M. P. L. Calus, J. A. M. van Arendonk, and H. Bovenhuis. Long-term response to genomic selection : effects of estimation method and reference population structure for different genetic architectures. *Genetics, selection, evolution : GSE*, 44 :3, Jan. 2012. ISSN 1297-9686. doi : 10.1186/1297-9686-44-3.
- C. Beaumont, F. Calenge, H. Chapuis, J. Fablet, F. Minvielle, and M. Tixier-Boichard. Génétique de la qualité de l’œuf - Inra Prod.Anim., 23 (2), 123-132, 2010. URL <http://www6.inra.fr/productions-animales/2010-Volume-23/Numero-2-2010/Genetique-de-la-qualite-de-l-aeuf>.
- S. Brard and A. Ricard. Is the use of formulae a reliable way to predict the accuracy of genomic selection? *Journal of Animal Breeding and Genetics*, 132(3) :207–217, 2015. ISSN 1439-0388. doi : 10.1111/jbg.12123. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jbg.12123>.
- F. V. Brito, J. B. Neto, M. Sargolzaei, J. A. Cobuci, and F. S. Schenkel. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics*, 12 :80, Sept. 2011. ISSN 1471-2156. doi : 10.1186/1471-2156-12-80. URL <https://doi.org/10.1186/1471-2156-12-80>.

- D. W. Burt. Origin and evolution of avian microchromosomes. *Cytogenetic and Genome Research*, 96(1-4) :97–112, 2002. ISSN 1424-8581. doi : 63018.
- M. P. L. Calus. Genomic breeding value prediction : methods and procedures. *Animal : An International Journal of Animal Bioscience*, 4(2) :157–164, Feb. 2010. ISSN 1751-7311. doi : 10.1017/S1751731109991352.
- H. D. Daetwyler, B. Villanueva, and J. A. Woolliams. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS One*, 3(10) :e3395, 2008. ISSN 1932-6203. doi : 10.1371/journal.pone.0003395.
- H. D. Daetwyler, B. J. Hayes, J. M. Hickey, J. M. Henshall, S. Dominik, J. H. J. V. D. Werf, and B. Gredler. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science*, 50(11-12) :1004–1010, Jan. 2010a. ISSN 1836-0939. doi : 10.1071/AN10096.
- H. D. Daetwyler, R. Pong-Wong, B. Villanueva, and J. A. Woolliams. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185(3) : 1021–1031, July 2010b. ISSN 1943-2631. doi : 10.1534/genetics.110.116855.
- H. D. Daetwyler, M. P. L. Calus, R. Pong-Wong, G. de Los Campos, and J. M. Hickey. Genomic prediction in animals and plants : simulation of data, validation, reporting, and benchmarking. *Genetics*, 193(2) :347–365, Feb. 2013. ISSN 1943-2631. doi : 10.1534/genetics.112.147983.
- R. Dassonneville, R. F. Brøndum, T. Druet, S. Fritz, F. Guillaume, B. Guldbrandtsen, M. S. Lund, V. Ducrocq, and G. Su. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *Journal of Dairy Science*, 94(7) :3679–3686, 2011. URL <http://www.sciencedirect.com/science/article/pii/S0022030211003651>.
- P. Deelen, M. J. Bonder, K. J. van der Velde, H.-J. Westra, E. Winder, D. Hendriksen, L. Franke, and M. A. Swertz. Genotype harmonizer : automatic strand alignment and format conversion for genotype data integration. *BMC Research Notes*, 7, Dec. 2014. ISSN 1756-0500. doi : 10.1186/1756-0500-7-901. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4307387/>.
- I. C. Dunn, M. Bain, A. Edmond, P. W. Wilson, N. Joseph, S. Solomon, B. De Ketelaere, J. De Baerdemaeker, M. Schmutz, R. Preisinger, and D. Waddington. Heritability and genetic correlation of measurements derived from acoustic resonance frequency analysis; a novel method of determining eggshell quality in domestic hens. *British Poultry Science*, 46(3) :280–286, June 2005. ISSN 0007-1668. doi : 10.1080/00071660500098574.
- I. C. Dunn, N. T. Joseph, M. Bain, A. Edmond, P. W. Wilson, P. Milona, Y. Nys, J. Gautron, M. Schmutz, R. Preisinger, and D. Waddington. Polymorphisms in eggshell organic matrix genes are associated with eggshell quality measurements in pedigree Rhode Island Red hens. *Animal Genetics*, 40(1) :110–114, Feb. 2009. ISSN 1365-2052. doi : 10.1111/j.1365-2052.2008.01794.x.

- J.-M. Elsen. Approximated prediction of genomic selection accuracy when reference and candidate populations are related. *Genetics, Selection, Evolution : GSE*, 48, Mar. 2016. ISSN 0999-193X. doi : 10.1186/s12711-016-0183-3. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4778372/>.
- M. Erbe, B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95(7) :4114–4129, July 2012. ISSN 1525-3198. doi : 10.3168/jds.2011-5019.
- M. Erbe, B. Gredler, F. R. Seefried, B. Bapst, and H. Simianer. A Function Accounting for Training Set Size and Marker Density to Model the Average Accuracy of Genomic Prediction. *PLoS ONE*, 8(12), Dec. 2013. ISSN 1932-6203. doi : 10.1371/journal.pone.0081046. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3855218/>.
- M. T. Falconer D.S. *Introduction to quantitative genetics*. 4th edition, 1996. ISBN 0-582-24302-5.
- M. Goddard. Genomic selection : prediction of accuracy and maximisation of long term response. *Genetica*, 136(2) :245–257, June 2009. ISSN 1573-6857. doi : 10.1007/s10709-008-9308-0.
- M. E. Goddard and B. J. Hayes. Genomic selection. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzucht Und Zuchtungsbiologie*, 124(6) :323–330, Dec. 2007. ISSN 0931-2668. doi : 10.1111/j.1439-0388.2007.00702.x.
- M. E. Goddard, B. J. Hayes, and T. H. E. Meuwissen. Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzucht Und Zuchtungsbiologie*, 128(6) :409–421, Dec. 2011. ISSN 1439-0388. doi : 10.1111/j.1439-0388.2011.00964.x.
- Z. S. Goraga, M. K. Nassar, and G. A. Brockmann. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines : I. egg production traits. *Animal Genetics*, 43(2) :183–189, Apr. 2012. ISSN 1365-2052. doi : 10.1111/j.1365-2052.2011.02233.x.
- T. Goto, A. Ishikawa, N. Goto, M. Nishibori, T. Umino, and M. Tsudzuki. Mapping of main-effect and epistatic quantitative trait loci for internal egg traits in an Fsub(2) resource population of chickens. *Journal of Poultry Science*, 2014, 51(4) :4, 2014. ISSN 1346-7395. URL <http://agris.fao.org/agris-search/search.do?recordID=JP2019009552>.
- D. Habier, R. L. Fernando, and J. C. M. Dekkers. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics*, 177(4) :2389–2397, Dec. 2007. ISSN 0016-6731, 1943-2631. doi : 10.1534/genetics.107.081190. URL <http://www.genetics.org/content/177/4/2389>.

- D. Habier, R. L. Fernando, and J. C. M. Dekkers. Genomic Selection Using Low-Density Marker Panels. *Genetics*, 182(1) :343–353, May 2009. ISSN 0016-6731, 1943-2631. doi : 10.1534/genetics.108.100289. URL <http://www.genetics.org/content/182/1/343>.
- D. Habier, R. L. Fernando, K. Kizilkaya, and D. J. Garrick. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(1) :186, May 2011. ISSN 1471-2105. doi : 10.1186/1471-2105-12-186. URL <https://doi.org/10.1186/1471-2105-12-186>.
- D. Habier, R. L. Fernando, and D. J. Garrick. Genomic BLUP decoded : a look into the black box of genomic prediction. *Genetics*, 194(3) :597–607, July 2013. ISSN 1943-2631. doi : 10.1534/genetics.113.152207.
- C. Hansen, N. Yi, Y. M. Zhang, S. Xu, J. Gavora, and H. H. Cheng. Identification of QTL for production traits in chickens. *Animal Biotechnology*, 16(1) :67–79, 2005. ISSN 1049-5398.
- B. J. Hayes, P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. Invited review : Genomic selection in dairy cattle : progress and challenges. *Journal of Dairy Science*, 92(2) :433–443, Feb. 2009. ISSN 1525-3198. doi : 10.3168/jds.2008-1646.
- C. R. Henderson. SIRE EVALUATION AND GENETIC TRENDS. *Journal of Animal Science*, 1973(Symposium) :10–41, Jan. 1973. ISSN 0021-8812. doi : 10.1093/ansci/1973.Symposium.10. URL <https://academic.oup.com/jas/article/1973/Symposium/10/4697208>.
- C. R. Henderson, O. Kempthorne, S. R. Searle, and C. M. von Krosigk. The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics*, 15(2) :192–218, 1959. ISSN 0006-341X. doi : 10.2307/2527669. URL <https://www.jstor.org/stable/2527669>.
- F. Herry, F. Hérault, D. Picard Druet, A. Varenne, T. Burlot, P. Le Roy, and S. Allais. Design of low density SNP chips for genotype imputation in layer chicken. *BMC Genetics*, 19, Dec. 2018. ISSN 1471-2156. doi : 10.1186/s12863-018-0695-7. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6278067/>.
- M. Honkatukia, M. Tuiskula-Haavisto, D.-J. de Koning, A. Virta, A. Mäki-Tanila, and J. Vilkki. A region on chicken chromosome 2 affects both egg white thinning and egg weight. *Genetics, selection, evolution : GSE*, 37(5) :563–577, Oct. 2005. ISSN 0999-193X. doi : 10.1051/gse:2005016.
- Z.-L. Hu, C. A. Park, and J. M. Reecy. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*, 47(D1) :D701–D710, Jan. 2019. ISSN 0305-1048. doi : 10.1093/nar/gky1084. URL <https://academic.oup.com/nar/article/47/D1/D701/5165267>.

- F. Hérault, F. Herry, A. Varinne, T. Burlot, D. Picard-Druet, F. Recoquillay, C. Macé, F. Fagnoul, S. Allais, and P. Le Roy. A linkage disequilibrium study in layers and broiler commercial chicken populations. *11. World Congress on Genetics Applied to Livestock Production (WCGALP), Feb 2018, Auckland, New Zealand*, Feb. 2018. URL <https://hal.archives-ouvertes.fr/hal-01791806/document>.
- J. Jensen, G. Su, and P. Madsen. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. *BMC Genetics*, 13(1) :44, June 2012. ISSN 1471-2156. doi : 10.1186/1471-2156-13-44. URL <https://doi.org/10.1186/1471-2156-13-44>.
- S. Kerje, O. Carlborg, L. Jacobsson, K. Schütz, C. Hartmann, P. Jensen, and L. Andersson. The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. *Animal Genetics*, 34(4) :264–274, Aug. 2003. ISSN 0268-9146.
- L. Kibala, I. Rozempolska-Rucinska, K. Kasperek, G. Zieba, M. Lukaszewicz, L. Kibala, I. Rozempolska-Rucinska, K. Kasperek, G. Zieba, and M. Lukaszewicz. Eggshell Qualities as Indicative of Eggshell Strength for Layer Selection. *Brazilian Journal of Poultry Science*, 20(1) :99–102, Mar. 2018. ISSN 1516-635X. doi : 10.1590/1806-9061-2017-0590. URL [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S1516-635X2018000100099&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1516-635X2018000100099&lng=en&nrm=iso&tlng=en).
- A. Kranis, A. A. Gheyas, C. Boschiero, F. Turner, L. Yu, S. Smith, R. Talbot, A. Pirani, F. Brew, P. Kaiser, P. M. Hocking, M. Fife, N. Salmon, J. Fulton, T. M. Strom, G. Haberer, S. Weigend, R. Preisinger, M. Gholami, S. Qanbari, H. Simianer, K. A. Watson, J. A. Woolliams, and D. W. Burt. Development of a high density 600k SNP genotyping array for chicken. *BMC Genomics*, 14(1) : 59, Jan. 2013. ISSN 1471-2164. doi : 10.1186/1471-2164-14-59. URL <https://doi.org/10.1186/1471-2164-14-59>.
- R. Lande and R. Thompson. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124(3) :743–756, Mar. 1990. ISSN 0016-6731, 1943-2631. URL <https://www.genetics.org/content/124/3/743>.
- Z. Ledvinka, L. Zita, and L. Klesalová. Egg quality and some factors influencing it : A review. *Scientia Agriculturae Bohemica*, 2012 :46–52, Jan. 2012.
- A. Legarra and A. Reverter. Can we frame and understand cross-validation results in animal breeding? July 2017.
- A. Legarra and A. Reverter. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics, selection, evolution : GSE*, 50(1) :53, Nov. 2018. ISSN 1297-9686. doi : 10.1186/s12711-018-0426-6.

- A. Legarra, C. Robert-Granié, E. Manfredi, and J.-M. Elsen. Performance of Genomic Selection in Mice. *Genetics*, 180(1) :611–618, Sept. 2008. ISSN 0016-6731. doi : 10.1534/genetics.108.088575. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2535710/>.
- A. Legarra, I. Aguilar, and I. Misztal. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9) :4656–4663, Sept. 2009. ISSN 1525-3198. doi : 10.3168/jds.2009-2061.
- A. Legarra, A. Ricard, and L. Varona. GWAS by GBLUP : Single and Multimarker EMMAX and Bayes Factors, with an Example in Detection of a Major Gene for Horse Gait. *G3 : Genes/Genomes/Genetics*, 8(7) :2301–2308, May 2018. ISSN 2160-1836. doi : 10.1534/g3.118.200336. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6027892/>.
- T. Liu, H. Qu, C. Luo, D. Shu, J. Wang, M. S. Lund, and G. Su. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. *BMC Genetics*, 15 :110, Oct. 2014. ISSN 1471-2156. doi : 10.1186/s12863-014-0110-y. URL <https://doi.org/10.1186/s12863-014-0110-y>.
- W. Liu, D. Li, J. Liu, S. Chen, L. Qu, J. Zheng, G. Xu, and N. Yang. A genome-wide SNP scan reveals novel loci for egg production and quality traits in white leghorn and brown-egg dwarf layers. *PloS One*, 6(12) :e28600, 2011a. ISSN 1932-6203. doi : 10.1371/journal.pone.0028600.
- Z. Liu, F. R. Seefried, F. Reinhardt, S. Rensing, G. Thaller, and R. Reents. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genetics, Selection, Evolution : GSE*, 43 (1) :19, May 2011b. ISSN 0999-193X. doi : 10.1186/1297-9686-43-19. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3107172/>.
- Z. Liu, C. Sun, Y. Yan, G. Li, F. Shi, G. Wu, A. Liu, and N. Yang. Genetic variations for egg quality of chickens at late laying period revealed by genome-wide association study. *Scientific Reports*, 8(1) :10832, July 2018. ISSN 2045-2322. doi : 10.1038/s41598-018-29162-7.
- D. a. L. Lourenco, I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. Are evaluations on young genotyped animals benefiting from the past generations? *Journal of Dairy Science*, 97(6) :3930–3942, 2014. ISSN 1525-3198. doi : 10.3168/jds.2013-7769.
- T. Luan, J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics*, 183(3) :1119–1126, Nov. 2009. ISSN 0016-6731, 1943-2631. doi : 10.1534/genetics.109.107391. URL <http://www.genetics.org/content/183/3/1119>.
- J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, 39(7) :906–913, July 2007. ISSN 1061-4036. doi : 10.1038/ng2088.

- H. A. McQueen, G. Siriaco, and A. P. Bird. Chicken Microchromosomes Are Hyperacetylated, Early Replicating, and Gene Rich. *Genome Research*, 8(6) :621–630, June 1998. ISSN 1088-9051. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC310741/>.
- H.-J. Megens, R. P. Crooijmans, J. W. Bastiaansen, H. H. Kerstens, A. Coster, R. Jalving, A. Vereijken, P. Silva, W. M. Muir, H. H. Cheng, O. Hanotte, and M. A. Groenen. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. *BMC Genetics*, 10(1) :86, Dec. 2009. ISSN 1471-2156. doi : 10.1186/1471-2156-10-86. URL <https://doi.org/10.1186/1471-2156-10-86>.
- T. Meuwissen, B. Hayes, and M. Goddard. Accelerating improvement of livestock with genomic selection. *Annual Review of Animal Biosciences*, 1 :221–237, Jan. 2013. ISSN 2165-8110. doi : 10.1146/annurev-animal-031412-103705.
- T. H. Meuwissen. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics, Selection, Evolution : GSE*, 41(1) :35, June 2009. ISSN 0999-193X. doi : 10.1186/1297-9686-41-35. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2708128/>.
- T. H. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4) :1819–1829, Apr. 2001. ISSN 0016-6731. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461589/>.
- I. Misztal, S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. BLUPF90 and related programs. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, volume Vol. 28, page 743, Jan. 2002.
- I. Misztal, A. Legarra, and I. Aguilar. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92(9) :4648–4655, Sept. 2009. ISSN 1525-3198. doi : 10.3168/jds.2009-2064.
- F. Morgante, W. Huang, C. Maltecca, and T. F. C. Mackay. Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity*, 120(6) :500–514, 2018. ISSN 1365-2540. doi : 10.1038/s41437-017-0043-0.
- G. Moser, M. S. Khatkar, B. J. Hayes, and H. W. Raadsma. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genetics Selection Evolution*, 42(1) :37, Oct. 2010. ISSN 1297-9686. doi : 10.1186/1297-9686-42-37. URL <https://doi.org/10.1186/1297-9686-42-37>.
- W. M. Muir. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzucht Und Zuchtungsbiologie*, 124(6) :342–355, Dec. 2007. ISSN 0931-2668. doi : 10.1111/j.1439-0388.2007.00700.x.

- B. U. Müller, B. Stich, and H.-P. Piepho. A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. *Heredity*, 106(5) :825–831, May 2011. ISSN 0018-067X. doi : 10.1038/hdy.2010.125. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186238/>.
- S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. PLINK a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3) :559–575, Sept. 2007. ISSN 0002-9297. doi : 10.1086/519795.
- A. Reverter, B. L. Golden, R. M. Bourdon, and J. S. Brinks. Technical note : detection of bias in genetic predictions2. *Journal of Animal Science*, 72(1) :34–37, Jan. 1994. ISSN 0021-8812, 1525-3163. doi : 10.2527/1994.72134x. URL <https://academic.oup.com/jas/article/72/1/34-37/4632556>.
- C. Robert-Granié, A. Legarra, and V. Ducrocq. Principes de base de la sélection génomique. *INRA Prod. Anim.*, (24 (4)) :331–340, 2011.
- H. Romé and P. Le Roy. Régions chromosomiques influençant les caractères de production et de qualité des oeufs de poule. *INRA Prod. Anim.*, 29(2) :117–128, 2016.
- H. Romé, A. Varenne, F. Héroult, H. Chapuis, C. Alleno, P. Dehais, A. Vignal, T. Burlot, and P. Le Roy. GWAS analyses reveal QTL in egg layers that differ in response to diet differences. *Genetics Selection Evolution*, 47 :83, Oct. 2015. ISSN 1297-9686. doi : 10.1186/s12711-015-0160-2. URL <https://doi.org/10.1186/s12711-015-0160-2>.
- O. Sasaki, S. Odawara, H. Takahashi, K. Nirasawa, Y. Oyamada, R. Yamamoto, K. Ishii, Y. Nagamine, H. Takeda, E. Kobayashi, and T. Furukawa. Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens. *Animal Genetics*, 35(3) :188–194, June 2004. ISSN 0268-9146. doi : 10.1111/j.1365-2052.2004.01133.x.
- M. A. Schreiweis, P. Y. Hester, P. Settar, and D. E. Moody. Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens. *Animal Genetics*, 37(2) :106–112, Apr. 2006. ISSN 0268-9146. doi : 10.1111/j.1365-2052.2005.01394.x.
- F. Sirri, M. Zampiga, A. Berardinelli, and A. Meluzzi. Variability and interaction of some egg physical and eggshell quality attributes during the entire laying hen cycle. *Poultry Science*, 97(5) :1818–1823, May 2018. ISSN 1525-3171. doi : 10.3382/ps/pex456.
- C. Sun, L. Qu, G. Yi, J. Yuan, Z. Duan, M. Shen, L. Qu, G. Xu, K. Wang, and N. Yang. Genome-wide association study revealed a promising region and candidate genes for eggshell quality in an F2 resource population. *BMC genomics*, 16 :565, July 2015. ISSN 1471-2164. doi : 10.1186/s12864-015-1795-7.

- M. Tuiskula-Haavisto, M. Honkatukia, J. Vilkki, D. J. de Koning, N. F. Schulman, and A. Mäki-Tanila. Mapping of quantitative trait loci affecting quality and production traits in egg layers. *Poultry Science*, 81(7) :919–927, July 2002. ISSN 0032-5791. doi : 10.1093/ps/81.7.919.
- M. Tuiskula-Haavisto, M. Honkatukia, R. Preisinger, M. Schmutz, D. J. de Koning, W. H. Wei, and J. Vilkki. Quantitative trait loci affecting eggshell traits in an F(2) population. *Animal Genetics*, 42(3) :293–299, June 2011. ISSN 1365-2052. doi : 10.1111/j.1365-2052.2010.02131.x.
- M. Tuiskula-Haavisto, M. Honkatukia, I. C. Dunn, M. M. Bain, D. J. De Koning, R. Preisinger, M. Schmutz, J. Arango, D. Fischer, and J. Vilkki. Validated quantitative trait loci for eggshell quality in experimental and commercial laying hens. *Animal Genetics*, 49(4) :329–333, Aug. 2018. ISSN 1365-2052. doi : 10.1111/age.12671.
- P. M. VanRaden. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11) :4414–4423, Nov. 2008. ISSN 1525-3198. doi : 10.3168/jds.2007-0980.
- P. M. VanRaden, C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. Invited review : reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1) :16–24, Jan. 2009. ISSN 1525-3198. doi : 10.3168/jds.2008-1514.
- K. L. Verbyla, B. J. Hayes, P. J. Bowman, and M. E. Goddard. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research*, 91(5) :307–311, Oct. 2009. ISSN 1469-5073. doi : 10.1017/S0016672309990243.
- K. L. Verbyla, P. J. Bowman, B. J. Hayes, and M. E. Goddard. Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings*, 4(Suppl 1) :S5, Mar. 2010. ISSN 1753-6561. doi : 10.1186/1753-6561-4-S1-S5. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2857847/>.
- Z. G. Vitezica, L. Varona, and A. Legarra. On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. *Genetics*, 195(4) :1223–1230, Dec. 2013. ISSN 0016-6731, 1943-2631. doi : 10.1534/genetics.113.155176. URL <https://www.genetics.org/content/195/4/1223>.
- Z. G. Vitezica, L. Varona, J.-M. Elsen, I. Misztal, W. Herring, and A. Legarra. Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genetics Selection Evolution*, 48(1) : 6, Jan. 2016. ISSN 1297-9686. doi : 10.1186/s12711-016-0185-1. URL <https://doi.org/10.1186/s12711-016-0185-1>.
- B. Wardecka, R. Olszewski, K. Jaszczak, G. Zieba, M. Pierzchała, and K. Wicińska. Relationship between microsatellite marker alleles on chromosomes 1-5 originating from the Rhode Island Red and Green-legged Partridge breeds and egg

- production and quality traits in F(2) mapping population. *Journal of Applied Genetics*, 43(3) :319–329, 2002. ISSN 1234-1983.
- W. C. Warren, L. W. Hillier, C. Tomlinson, P. Minx, M. Kremitzki, T. Graves, C. Markovic, N. Bouk, K. D. Pruitt, F. Thibaud-Nissen, V. Schneider, T. A. Mansour, C. T. Brown, A. Zimin, R. Hawken, M. Abrahamsen, A. B. Pyrkosz, M. Morisson, V. Fillon, A. Vignal, W. Chow, K. Howe, J. E. Fulton, M. M. Miller, P. Lovell, C. V. Mello, M. Wirthlin, A. S. Mason, R. Kuo, D. W. Burt, J. B. Dodgson, and H. H. Cheng. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda, Md.)*, 7(1) :109–117, 2017. ISSN 2160-1836. doi : 10.1534/g3.116.035923.
- K. L. Weber, R. M. Thallman, J. W. Keele, W. M. Snelling, G. L. Bennett, T. P. L. Smith, T. G. McDanel, M. F. Allan, A. L. Van Eenennaam, and L. A. Kuehn. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes,. *Journal of Animal Science*, 90 (12) :4177–4190, Dec. 2012. ISSN 0021-8812. doi : 10.2527/jas.2011-4586. URL <https://academic.oup.com/jas/article/90/12/4177/4717862>.
- Z. Weng, A. Wolc, X. Shen, R. L. Fernando, J. C. M. Dekkers, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, and D. J. Garrick. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genetics Selection Evolution*, 48 :22, 2016. ISSN 1297-9686. doi : 10.1186/s12711-016-0198-9. URL <http://dx.doi.org/10.1186/s12711-016-0198-9>.
- Y. C. J. Wientjes, R. F. Veerkamp, and M. P. L. Calus. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics*, 193(2) :621–631, Feb. 2013. ISSN 1943-2631. doi : 10.1534/genetics.112.146290.
- A. Wolc, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *ResearchGate*, 43(1) :23, June 2011. ISSN 1297-9686. doi : 10.1186/1297-9686-43-23. URL [https://www.researchgate.net/publication/51238224\\_Persistence\\_of\\_accuracy\\_of\\_genomic\\_estimated\\_breeding\\_values\\_over\\_generations\\_in\\_layer\\_chickens](https://www.researchgate.net/publication/51238224_Persistence_of_accuracy_of_genomic_estimated_breeding_values_over_generations_in_layer_chickens).
- A. Wolc, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, W. G. Hill, and J. C. M. Dekkers. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Animal Genetics*, 43 Suppl 1 :87–96, July 2012. ISSN 1365-2052. doi : 10.1111/j.1365-2052.2012.02381.x.
- A. Wolc, J. Arango, T. Jankowski, I. Dunn, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, R. L. Fernando, D. J. Garrick, and J. C. M. Dekkers. Genome-wide association study for egg production and quality in layer chickens. *Journal of Animal Breeding and Genetics = Zeitschrift Fur Tierzucht Und*

*Zuchtungsbiologie*, 131(3) :173–182, June 2014. ISSN 1439-0388. doi : 10.1111/jbgi.12086.

- A. Wolc, H. H. Zhao, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, C. Stricker, D. Habier, R. L. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. Response and inbreeding from a genomic selection experiment in layer chickens. *Genetics Selection Evolution*, 47 :59, 2015. ISSN 1297-9686. doi : 10.1186/s12711-015-0133-5. URL <http://dx.doi.org/10.1186/s12711-015-0133-5>.
- A. Wolc, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, J. C. M. Dekkers, R. Fernando, and D. J. Garrick. Mixture models detect large effect QTL better than GBLUP and result in more accurate and persistent predictions. *ResearchGate*, 7(1) :7, Feb. 2016a. ISSN 1674-9782. doi : 10.1186/s40104-016-0066-z. URL [https://www.researchgate.net/publication/294258215\\_Mixture\\_models\\_detect\\_large\\_effect\\_QTL\\_better\\_than\\_GBLUP\\_and\\_result\\_in\\_more\\_accurate\\_and\\_persistent\\_predictions](https://www.researchgate.net/publication/294258215_Mixture_models_detect_large_effect_QTL_better_than_GBLUP_and_result_in_more_accurate_and_persistent_predictions).
- A. Wolc, A. Kranis, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, A. Avendano, K. A. Watson, J. M. Hickey, G. d. l. Campos, R. L. Fernando, D. J. Garrick, and J. C. M. Dekkers. Implementation of genomic selection in the poultry industry. *ResearchGate*, 6(1) :23, Jan. 2016b. ISSN 2160-6064. doi : 10.2527/af.2016-0004. URL [https://www.researchgate.net/publication/289494029\\_Implementation\\_of\\_genomic\\_selection\\_in\\_the\\_poultry\\_industry](https://www.researchgate.net/publication/289494029_Implementation_of_genomic_selection_in_the_poultry_industry).
- D. Wright, S. Kerje, H. Brändström, K. Schütz, A. Kindmark, L. Andersson, P. Jensen, and T. Pizzari. The genetic architecture of a female sexual ornament. *Evolution; International Journal of Organic Evolution*, 62(1) :86–98, Jan. 2008. ISSN 0014-3820. doi : 10.1111/j.1558-5646.2007.00281.x.
- T. Xiang, O. F. Christensen, Z. G. Vitezica, and A. Legarra. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genetics Selection Evolution*, 48(1) : 92, Nov. 2016. ISSN 1297-9686. doi : 10.1186/s12711-016-0271-4. URL <https://doi.org/10.1186/s12711-016-0271-4>.
- J. F. Yao, Z. X. Chen, G. Y. Xu, X. L. Wang, Z. H. Ning, J. X. Zheng, L. J. Qu, and N. Yang. Low-density lipoprotein receptor-related protein 8 gene association with egg traits in dwarf chickens. *Poultry Science*, 89(5) :883–886, May 2010. ISSN 0032-5791. doi : 10.3382/ps.2009-00613.
- F. Ytournal. Déséquilibre de liaison et cartographie de QTL en population sélectionnée. Jan. 2008. URL <https://pastel.archives-ouvertes.fr/pastel-00003789>.
- L. C. Zhang, Z. H. Ning, G. Y. Xu, Z. C. Hou, and N. Yang. Heritabilities and genetic and phenotypic correlations of egg quality traits in brown-egg dwarf layers. *Poultry Science*, 84(8) :1209–1213, Aug. 2005. ISSN 0032-5791. doi : 10.1093/ps/84.8.1209.

S. Zhong, J. C. M. Dekkers, R. L. Fernando, and J.-L. Jannink. Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines : A Barley Case Study. *Genetics*, 182(1) :355–364, May 2009. ISSN 0016-6731. doi : 10.1534/genetics.108.098277. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2674832/>.

# Table des figures

1.1	Production d'œufs entre 1961 et 2017 dans le monde (millions de tonnes)	8
1.2	Répartition des différents systèmes d'élevage dans l'UE en 2017 (ITAVI, 2018)	9
1.3	Schéma de sélection pyramidal : les lignées pures sont conservées en noyau de sélection, et Grands-parentaux (GP), Parentaux (P) et produits finaux sont diffusés à travers le monde.	10
1.4	Courbes de référence concernant la viabilité, le pourcentage de ponte, le poids d'œuf, le poids corporel et la consommation journalière en fonction de l'âge en semaine pour la souche ISA Brown (Source : Hendrix-ISA)	12
1.5	Illustration du progrès génétique $\Delta G$ par la sélection des reproducteurs de la génération 0 à la génération 1.	17
1.6	Décroissance du DL en fonction du temps, taux de recombinaison ( $r$ ) et DL initial ( $D_0$ ). Source : Ytournal (2008)	20
1.7	Déséquilibre de liaison entre marqueurs en fonction de la distance les séparant, dans 2 lignées de poules pondeuses (L et RI) et une lignée de poulets de chair. Source : Hérault et al. (2018)	21
1.8	Taille des blocs haplotypiques et nombre de SNPs dans ces blocs, en fonction du chromosome, pour la lignée Rhode Island et la puce à SNPs utilisée dans cette thèse. Source : Hérault et al. (2018)	21
1.9	Nombre d'individus nécessaires (en ordonnée), pour atteindre une précision donnée en fonction de l'héritabilité du caractère (en abscisse), d'après les formules de Goddard and Hayes (2007). Source : Hayes et al. (2009)	26
1.10	Précision de l'évaluation, estimée et observée, en fonction du nombre d'animaux dans la population de référence, et du nombre de marqueurs utilisés pour le génotypage. Le plan est le résultat théorique obtenue à partir d'une équation modifiée de Daetwyler et al. (2010b), les flèches représentent les résultats obtenus sur les données réelles des taureaux. Source : Erbe et al. (2013)	27

1.11	Précision des résultats du GBLUP en fonction de la taille de la population de référence et du type de relation entre population de référence et candidate. Résultats obtenus par simulation. La ligne continue bleu foncé présente les précisions obtenues avec des candidats apparentés aux individus dans la population de référence (individus réels). La ligne hachurée en bleu clair présente les précisions obtenues par des candidats simulés en utilisant le DL et les fréquences alléliques de la population de référence. En rouge sont les résultats quand la population candidate est simulée à partir des fréquences alléliques de la référence. En jaune, quand les candidats sont simulés en combinant les haplotypes de la population de référence. En vert, quand les candidats sont simulés par combinaison de chromosomes de la population de référence. Source : Wientjes et al. (2013) . . . . .	28
1.12	Évolution de la précision de l'évaluation sur plusieurs caractères de production d'œuf chez la poule pondeuse, en fonction du nombre de générations prises en compte. Caractères : eCO : couleur de la coquille, eEW : poids de l'œuf, eC3 : couleur des 3 premiers œufs, eSM : age de maturité sexuelle, eAH : hauteur d'albumen, eYW : poids du jaune, ePD : taux ponte, ePS : perforation de la coquille. Performances récoltées entre 26 et 28 semaines. Source : Weng et al. (2016) . . . . .	29
2.1	Structure des populations de la lignée pure. . . . .	35
2.2	Généalogie détaillée des différents lots composant la lignée pure. Pour chaque lot est indiqué le nombre de pondeuses le composant, ainsi que le nombre de parents, mâles et femelles. . . . .	36
2.3	Relations de parenté entre lignée pure et cheptel de poules croisées. . . . .	37
2.4	Différentes couleurs de coquille d'œuf . . . . .	39
2.5	Mesure de la hauteur de l'albumen . . . . .	40
2.6	Répartition des candidats sur la population de lignée pure . . . . .	43
5.1	Nombre de SNPs dans les blocs haplotypiques, en fonction du chromosome, pour la lignée Rhode Island et la puce à SNPs utilisée dans cette thèse. Source : Hérault et al. (2018) . . . . .	89
5.2	Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	94
5.3	Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	94

5.4	Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	98
5.5	Variance expliquée pour EW par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	98
6.1	Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	107
6.2	Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	108
6.3	Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	108
6.4	Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	109
6.5	Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	109
6.6	Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	110
6.7	Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	110

6.8	Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS mixte . . . . .	111
6.9	Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	112
6.10	Variance expliquée pour ESC par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	112
6.11	Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	113
6.12	Variance expliquée pour ESS par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	113
6.13	Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS pure, négatif, GWAS croisée . . . . .	114
6.14	Variance expliquée pour AH par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	114
6.15	Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules jeunes. Les différentes couleurs sont là différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	115
6.16	Variance expliquée pour ESshape par les groupes de 8 SNPs, en fonction de leur emplacement, sur les poules âgées. Les différentes couleurs sont là pour différencier les chromosomes, répartis de gauche à droite du 0 au 28, suivis du 33 (en bleu), du W (en gris) et du Z (en violet). En positif, GWAS croisée, négatif, GWAS mixte . . . . .	115

# Liste des tableaux

1.1	Héritabilité des différents caractères de qualité d'œuf étudiés, dans différentes populations . . . . .	13
1.2	Régions génomiques affectant les caractères de qualité d'œuf, d'après Romé and Le Roy (2016). La position correspond à la localisation moyenne du pic des QTL. . . . .	14
1.3	Fréquence des haplotypes pour 2 loci bi-alléliques en situation d'équilibre d'association gamétique . . . . .	18
1.4	Fréquence des haplotypes pour 2 loci bi-alléliques en situation de déséquilibre de liaison . . . . .	19
2.1	Résumé des différentes étapes de filtration du contrôle qualité des SNPs	39
2.2	Statistiques élémentaires des performances de CC . . . . .	41
2.3	Statistiques élémentaires des performances de 50sem . . . . .	41
2.4	Statistiques élémentaires des performances en CI . . . . .	41
2.5	Statistiques élémentaires des performances en 70sem . . . . .	41
5.1	QTLs considérés comme significatifs pour les caractères en lignée pure, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet. Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL. Décrits sur les GWAS pure et mixte respectivement, chez les poules jeunes. . . . .	91
5.2	QTLs considérés comme significatifs pour les caractères en lignée pure, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet. Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL. Décrits sur les GWAS pure et mixte, chez les poules âgées . . .	92
5.3	QTLs considérés comme significatifs pour les caractères des descendantes croisées, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet. Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL. Décrits sur les GWAS pure et mixte, chez les poules jeunes . . . . .	95

5.4	QTLs considérés comme significatifs pour les caractères des descendantes croisées, leur effet en % de variance expliquée, intervalle de confiance (en Mb), et la localisation du groupe de 8 SNPs avec le plus fort effet. Les QTLs communs sont indiqués en gras, l'acronyme "VE" indique la variance expliquée par l'ensemble des SNPs composant le QTL. Décrits sur les GWAS pure et mixte, chez les poules âgées . . . . .	96
5.5	QTLs communs entre caractères PB et CB . . . . .	99
5.6	QTLs affectant les caractères de qualité d'œuf, d'après Romé and Le Roy (2016). La position correspond à la localisation moyenne du pic des QTL. . . . .	100

---

**Titre :** Précision de l'évaluation génomique chez la poule pondeuse

**Mots clés :** poules pondeuses, évaluation génomique, précision de l'évaluation

**Résumé :** La sélection est particulièrement importante pour le domaine de la production d'œufs, qui est un marché en expansion au niveau mondial. Celui-ci présente une forte segmentation du marché, c'est à dire que chaque pays aura un cahier des charges particulier, dépendant des attentes des consommateurs. Cela signifie que les entreprises de sélection ont à produire des individus adaptés, ou adaptable, aux conditions d'élevage diverses, ainsi que capables de répondre aux différents besoin des consommateurs. C'est également un secteur fortement concurrentiel, détenu au niveau mondial par 3 grands groupes. Cela fait de la sélection des différentes lignées, qui répondent aux différents critères de production, une étape particulièrement critique. Le progrès génétique des populations sélectionnées est fortement diffusé partout dans le monde, et l'optimiser est un moyen de se distinguer des concurrents.

Depuis plusieurs années, la méthodologie de la sélection génomique est mise en place dans différentes filières. Elle a notamment permis d'accélérer le progrès génétique dans la filière bovine, qui a été pionnière dans son utilisation. Cette méthode d'évaluation a été jugée prometteuse pour la filière avicole. Pour estimer précisément l'effet de l'usage de cette méthode d'évaluation sur le progrès génétique des populations sélectionnées, il est nécessaire de connaître la précision de l'évaluation génomique.

L'objectif premier de cette thèse est d'estimer la précision de l'évaluation génomique sur les caractères de qualité d'œuf chez la poule pondeuse dans différents cas de figures, en comparaison de celle obtenue par l'évaluation génétique utilisée jusqu'alors.

---

**Title :** Accuracy of genomic selection in layers

**Keywords :** layers, genomic evaluation, evaluation accuracy

**Abstract :** Selection is very important for the egg production domain, which is a growing market worldwide. This market has a very segmented nature, with every country having special specifications, depending on consumers expectations. This means that breeding companies have to produce birds which are adapted, or adaptable, to various breeding conditions, while still being able to produce eggs suitable for consumers expectations. Egg production market is also a very competitive, detained worldwide by only 3 enterprise groups. This means that the selection of the different lines, which are dedicated to fulfill differents production criteria, is a critical step. The genetic progress of selected populations is distributed worldwide, and optimize it is a way to stand out from the competitors.

For the last few years, genomic selection is put in place in different sectors. Among other things, this methodology allowed to improve genetic progress in cattle industry, which was a precursor in the utilisation of it. This evaluation methode seems highly promising for the avian sector. To precisely estimate the effect of using it, it is necessary to know the accuracy of genomic evaluation.

The main objective of this work is to estimate the precision of genomic evaluation on layers egg quality traits under different scenarios, in comparison with what is observed when using genetic evaluation.