

Note Iboval

Bulletin d'information génétique des bovins allaitants

N° 90 – Juillet 2020

Les GEBV produits à partir de l'été 2020 le sont sur la base d'une connaissance affinée du génome bovin

L'évaluation IBOVAL estivale génomique 2020 (réf. 2021_01) utilise une version plus précise de l'assemblage bovin, c'est-à-dire la « carte » du génome, support des évaluations génomiques. Cette mise en œuvre se traduit, toutes choses égales par ailleurs, par de légères variations d'index consécutives à l'amélioration de la fiabilité de cet assemblage. Ces modifications de niveaux s'ajouteront à celles habituellement constatées entre deux évaluations successives pour les valeurs génomiques des animaux génotypés des races concernées (à savoir Charolaise, Limousine et Blonde d'Aquitaine).

Qu'est-ce qu'un assemblage ?

À ce jour, le génome bovin (support des évaluations génomiques) n'est pas connu avec exactitude. La représentation qu'on en a actuellement est appelée « assemblage ». Sa construction résulte directement des méthodes de séquençage qui fournissent un très grand nombre de petites séquences d'environ 150 bases. Ces séquences courtes doivent être positionnées les unes par rapport aux autres, c'est-à-dire assemblées comme un gigantesque puzzle d'environ 2,8 milliards de bases. C'est à partir de cet assemblage qu'est construite la carte génétique utilisée dans les évaluations, qui correspond au positionnement des marqueurs les uns par rapport aux autres, sur chacun des chromosomes.



POURQUOI METTRE À JOUR L'ASSEMBLAGE ?



La qualité de l'assemblage n'est pas parfaite puisque des régions restent non assemblées (et donc non connues) et d'autres, plus ou moins grandes, peuvent être mal placées ou inversées. Or, un assemblage imparfait conduit à des interprétations erronées sur l'organisation et le fonctionnement des gènes. Pour les applications de sélection, un assemblage partiellement inachevé se traduit par des marqueurs à la position fautive et/ou l'absence de marqueurs dans des régions mal connues.

Des efforts récurrents de la communauté scientifique internationale conduisent à des améliorations progressives de l'assemblage. Aucune évolution significative n'avait été publiée depuis 2009. Le nouvel assemblage était donc très attendu. Il comporte 66,2 millions de nouvelles paires de bases positionnées, et corrige un grand nombre d'erreurs de l'assemblage précédent. On a ainsi davantage de marqueurs positionnés, et de façon plus précise, donc une carte génétique plus fiable. Cette bascule vers le nouvel assemblage est progressivement mise en œuvre par tous les centres de calcul d'évaluations génétiques. Son utilisation va donc impacter qualitativement l'imputation des génotypes (la position relative des marqueurs les uns par rapport aux autres ayant pu évoluer, certains allant jusqu'à changer de chromosome), apportant plus de précision et de fiabilité dans la constitution des phases (reconstitution des haplotypes=allèles de différents loci portés par le même brin de chromosome), donc des index. La mise à jour de l'assemblage permet aussi une meilleure définition des haplotypes liés à des gènes d'intérêts ou d'anomalies, et donc une meilleure identification des animaux porteurs de ces haplotypes.

Durant toute l'année 2019 ([Réf n°1](#)) les équipes de l'UMT eBIS (INRA, Aliche, Idele), le centre INRAE-CTIG et GenEval ont collaboré sur le projet de mise à jour de l'assemblage utilisé dans les outils d'évaluation. Valogène a également été associé sur les étapes traitant des tests sur haplotypes. Il en a résulté un ensemble de travaux d'implémentation, de validation, d'amélioration de paramétrages et de contrôles qualité profitables à l'ensemble du dispositif.

Différents tests de comparaison de résultats ont été menés afin de mesurer les impacts de ce nouvel assemblage sur l'ensemble des applications susceptibles d'utiliser les nouvelles phases, notamment l'imputation, les évaluations génomiques, les Indicateurs Précoces de Valeurs génomiques (IPVgénos) et les tests sur haplotypes.

LES CONSÉQUENCES FAVORABLES DE SON UTILISATION

Le gradient d'impact, systématiquement favorable, dépend de la structure de la population, du nombre de typages et du type de puces utilisées. Ainsi, globalement, les conséquences du nouvel assemblage sur les index sont plus importantes en bovins viande qu'en bovins laitier de même qu'elles sont plus importantes pour les races pour lesquelles la part des typages réalisés avec des puces de basse densité est plus élevée.

Sur l'imputation

Les phases générées avec le nouvel assemblage peuvent, selon les races et le nombre de typages basse (LD = 10 K), moyenne (MD = 50 K) ou haute densité (HD = 800 K) disponibles, présenter des différences avec les phases actuelles :

- pour les 3 races allaitantes bénéficiant actuellement d'une évaluation génomique, les résultats entre les 2 types de phases sont proches, avec en moyenne des résultats présentant 0.2 à 46 marqueurs imputés différents par animal selon les races (selon le taux de typages sur puce 50k ou HD par rapport à l'ensemble),

- pour les autres races allaitantes, les résultats sont très variables en fonction des effectifs et du taux d'animaux typés sur puce 50k (de 0.1 différence en moyenne à plus de 1000 différences).

Sur les valeurs génomiques

Par ailleurs, GenEval a mesuré les impacts de la mise en place du nouvel assemblage, pour toutes les valorisations génomiques produites ([Réf n°2](#)) et toutes les races.

Sur les index génomiques (GEBV)

En utilisant les évaluations IBOVAL officielles 2020_01 (index produits durant l'été 2019) comme référence, les conséquences des seules variations liées au nouvel assemblage ont été estimées. La corrélation globale entre ces deux évaluations pour les différents caractères est d'un niveau très élevé, mais elle n'est pas égale à 1 (ce qui signifierait absence de variation ou lien identique). Le niveau de variation se situe globalement à mi-chemin entre 1 et la variation habituelle de corrélation que l'on constate entre deux évaluations successives. Les résultats sont les plus proches de 1 en race Blonde d'Aquitaine ([cf. Figure n°1](#)), avec des corrélations supérieures à 0,995 pour les index ABjbf et de plus de 0,99 en moyenne pour les index de la naissance jusqu'au sevrage. Cela s'explique en particulier par la part très élevée de typage HD et MD par rapport à l'ensemble des typages disponibles dans cette race sur les index de la naissance au sevrage.

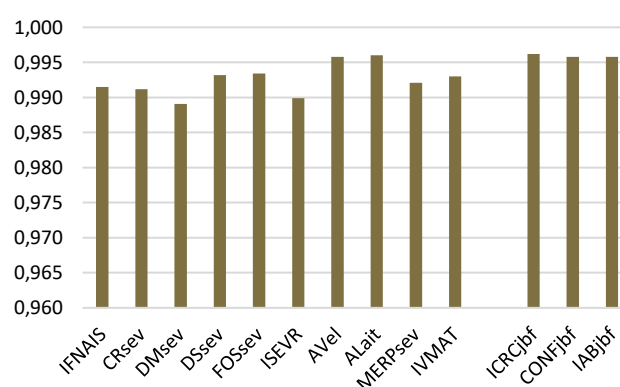


Figure n°1 : Corrélations nouvel/ancien assemblage Race Blonde d'Aquitaine (sur les index 2020_01)

À l'inverse, les corrélations observées en race Charolaise (cf. Figure n°2), sont d'un niveau plus faible, en particulier pour les index de la naissance jusqu'au sevrage (environ 0,97). L'explication réside à la fois au niveau de la structure de la population, du nombre de génotypages et du type de puces utilisées. La situation de la race Limousine (cf. Figure n°3) est intermédiaire à celle des deux autres races, avec des corrélations très élevés sur les index ABjbf et un peu moindres sur les index de la naissance au sevrage.

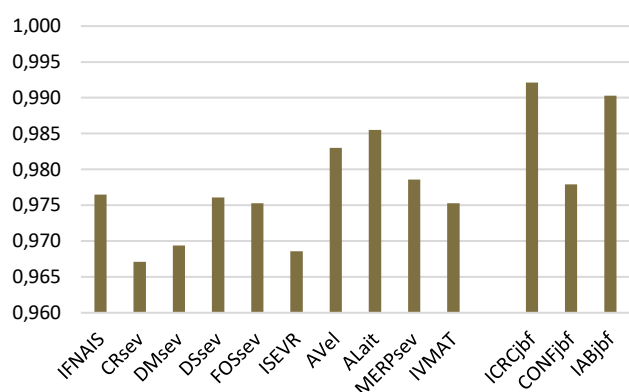


Figure n°2 : Corrélations nouvel/ancien assemblage Race Charolaise (sur les index 2020_01)

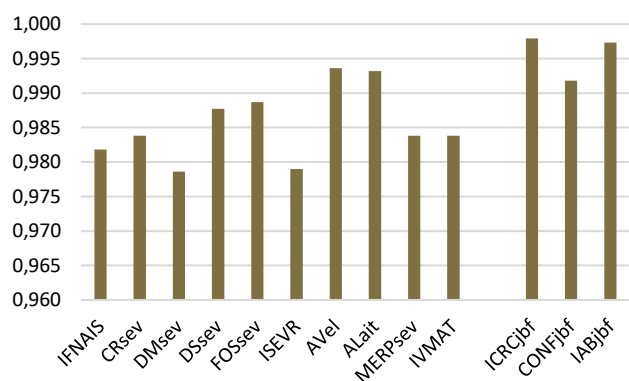


Figure n°3 : Corrélations nouvel/ancien assemblage Race Limousine (sur les index 2020_01)

Un travail complémentaire de comparaisons de résultats a été effectué entre le traitement de référence de l'indexation IBOVAL 2020_01 et le traitement de cet hiver 2020, effectué avec et sans utilisation du nouvel assemblage. Dans les trois races, quelles que soient les évaluations, on observe une très légère baisse des corrélations avec l'utilisation du nouvel assemblage. Les variations en matière de niveau de précision (CD) ou d'effectifs sont marginales.

Sur les IPVgénos

Le nouvel assemblage est utilisé depuis février 2020 pour les évaluations génétiques bovines. Ainsi, les calculs des IPVgénos produits depuis cette époque, bénéficient déjà de cette évolution. L'analyse des différents résultats de comparaison les concernant montre logiquement des tendances similaires sur les IPVgénos à celles observées sur les index génomiques (GEBV). La seule nuance qui peut être notée est le lien encore un peu plus étroit, quels que soit la race et les index concernés, qui est observé entre les résultats établis à partir de l'ancien ou du nouvel assemblage.

De plus, il est utile de signaler que les niveaux élevés à très élevés évoqués en termes de corrélations globales observé au niveau de ces différentes valeurs génomiques peuvent néanmoins masquer des variations individuelles non négligeables mais peu fréquentes.

CONCLUSION

Les différents tests de mesure des conséquences du nouvel assemblage montrent un faible à très faible impact global sur les évaluations génomiques. Les quelques variations générales observées, moitié moindres en moyenne que les changements entre deux évaluations successives officielles classiques, sont une conséquence directe du changement de carte du génome.

La bascule en faveur d'une utilisation du nouvel assemblage est effective depuis le 24/02/2020. Depuis, les nouvelles phases sont utilisées pour le calcul des valeurs privées (IPVgénos). Le traitement IBOVAL complet 2021_01 de l'été 2020 est le premier à utiliser les nouvelles phases en bovins viande. Un surcroît de variation d'index génomiques individuels pourra être expliqué par l'utilisation du nouvel assemblage.

RÉFÉRENCES

1. [GenEval – UMT eBIS – Valogène, 2020, communiqué de presse du projet NOVA, février 2020, 4 pages](#)
2. [Institut de l'Élevage – INRAE, 2017, note IBOVAL n°78, IBOVAL 2017 : Première large diffusion des index génomiques, mars 2017, 7 pages](#)