# Des analyses GWAS réalisées à l'échelle de la séquence révèlent des régions du génome associées au microbiote fécal en race Holstein.

BRULIN L. (1,2), SANCHEZ M.P. (2), CAI Z. (3), DUCROCQ S. (1,4), EVEN G. (1,4), MARTEL S. (1,4), MERLIN S. (1,4), AUDEBERT C. (1,4), ESTELLE J. (2), SAHANA G. (3), CROISEAU P. (2)
(1) GD Biotech – Gènes Diffusion, Lille, 59000 France
(2) Université Paris-Saclay, AgroParisTech, INRAE, GABI, Jouy-en-Josas 78350, France
(3) Center for Quantitative Genetics and Genomics, Université d'Aarhus, Denmark
(4) PEGASE-Biosciences, Institut Pasteur de Lille, 1 Rue du Professeur Calmette, 59019, Lille, France

**RESUME**
Le microbiote gastro-intestinal des animaux d'élevage, notamment des ruminants, a suscité un intérêt croissant au cours de la dernière décennie, donnant lieu à diverses études. Une partie d'entre elles ont révélé une composante génétique influençant la composition de certaines communautés bactériennes. L'objectif de cette étude est d'approfondir les connaissances sur l'influence de la génétique de l'hôte sur la composition du microbiote fécal de la vache laitière. Pour cela, des échantillons fécaux de 1 875 vaches Holstein issues de 144 fermes commerciales françaises ont été collectés et un séquençage ciblant le gène ARNr 16S a été réalisé. Les abondances de genres précédemment identifiés comme héritables ont été analysées dans des études d'association génomique (GWAS) en utilisant les génotypes des vaches, imputés au niveau de la séquence (13 millions de variants). Ces analyses ont permis d'identifier cinq QTL (quantitative trait loci) associés respectivement à l'abondance de *Paeniclostridium*, *Akkermansia*, *Sutterella*, *Turicibacter* et d'un genre inconnu de la famille des Paludibacteraceae. Les régions génomiques correspondantes contiennent des gènes candidats susceptibles d'influencer l'abondance de ces genres, tel que le gène *ABO*. En conclusion, ces analyses ont mis en évidence des régions et variants génomiques associés à l'abondance de certains genres héritables du microbiote fécal chez la vache Holstein. Ces résultats ouvrent de nouvelles perspectives de recherche pour mieux comprendre l'influence génétique de l'hôte sur son microbiote et évaluer l'intérêt de réaliser une sélection génomique sur la composition de certains genres bactériens.

# Sequence-based GWAS analyses revealed genome regions associated with the fecal microbiota in Holstein.

BRULIN L. (1,2), SANCHEZ M.P. (2), CAI Z. (3), DUCROCQ S. (1,4), EVEN G. (1,4), MARTEL S. (1,4), MERLIN S. (1,4), AUDEBERT C. (1,4), ESTELLE J. (2), SAHANA G. (3), CROISEAU P. (2)
(1) GD Biotech – Gènes Diffusion, Lille, 59000 France

## INTRODUCTION

Au cours des dernières années, de nombreuses études ont mis en évidence l'importance du microbiote gastro-intestinal dans l'expression des performances des animaux d'élevage. Chez les ruminants, les microorganismes sont à l'origine d'une fermentation permettant la digestion des protéines, lipides et glucides, parfois non-digérables par l'hôte, et les valorisent en différents produits tels que les acides gras volatils, absorbés dans le sang. Des études ont notamment révélé des associations significatives entre les performances des ruminants et la composition de leur microbiote ruminal (Jami *et al.*, 2014 ; Xue *et al.*, 2018), l'une des communautés bactériennes les plus étudiées chez ces animaux. Néanmoins, d'autres microbiotes du bovin suscitent également de l'intérêt, tel que le microbiote fécal. En effet, ce dernier a révélé des associations phénotypiques et génétiques avec de nombreux caractères zootechniques, tels que l'efficacité alimentaire, la production laitière, etc. (Brulin *et al*., 2024a, Monteiro *et al*., 2022).
En parallèle, de nombreuses études ont montré l'existence d'une composante génétique influençant le microbiote. La génétique de l'hôte impacte la diversité et la composition de divers microbiotes, notamment celui des fèces (Brulin *et* al., 2024b), avec des héritabilités faibles à modérées.
Ces différents résultats soulèvent la question des mécanismes génétiques à l'origine de cette influence. Une analyse d'association à l'échelle du génome permettrait d'estimer l'effet des différents marqueurs moléculaires. Ces informations permettraient de mieux appréhender l'influence de la génétique de l'hôte sur l'évolution de la composition du microbiote fécal.
Dans cette étude, nous avons analysé les données de microbiote fécal et de génotypage de 1 875 vaches Holstein issues de fermes commerciales françaises. L'objectif est d'identifier les variants et les gènes candidats associés à la composition du microbiote fécal de la vache Holstein grâce à des analyses d'association (GWAS) à l'échelle de la séquence complète.

## 1. MATERIEL ET METHODES

### 1.1. ECHANTILLONNAGE ET POPULATION

Dans le cadre de cette étude, entre 2020 et 2022, des échantillons fécaux ont été prélevés sur 1 875 vaches Holstein issues de 144 cheptels commerciaux français, grâce à une stimulation rectale réalisée par des techniciens de Gènes Diffusion (Douai, France). Les échantillons ont été immédiatement conservés dans des tubes stériles de 30 mL contenant approximativement 15 mL d'une solution à base d'éthanol, puis transportés et stockés à 4°C au centre de recherche de Gènes Diffusion (Institut Pasteur de Lille, France). Des primipares et multipares (2ème à 5ème lactation) étaient toutes en lactation depuis 5 à 450 jours lors du prélèvement.
L'utilisation d'animaux de fermes commerciales a limité le nombre d'information concernant la conduite d'élevage et leur ration. Cependant, les vaches contemporaines, prélevées lors d'une même campagne, étaient soumises à des pratiques d'élevage similaires. Nous avons alors intégré l'ensemble des effets liées à l'élevage dans une variable cheptel – groupe de séquençage - saison de prélèvement, tel que précédemment fait dans Brulin *et al*. (2024b).

### 1.2. EXTRACTION ADN ET SEQUENCAGE DU GENE 16S ARNr

Les échantillons de cette étude ont été divisés en 18 groupes de séquençage. L'ADN contenu dans les échantillons fécaux a, dans un premier temps, été extrait grâce au Nucleospin® 96 soil kit (Macherey Nagel, Düren, Allemagne) selon les recommandations du fabricant. Ensuite, une stratégie de séquençage dite « dual-indexed paired-end » a été employée afin de construire les librairies de séquençage ciblant les régions variables V3-V4 du gène 16S rRNA. Pour ce faire, deux étapes de PCR ont été réalisées sur les échantillons en utilisant des séquences 5-Tag (amorces sens : 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' et antisens : 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3') ainsi que des amorces spécifiques au gène 16S rRNA (sens : 5'-CCTACGGGNGGCWGCAG-3' ; antisens : 5'-GACTACHVGGGTATCTAATCC-3') (IDT, Iowa, EU). Pour la seconde PCR, deux loci ont été ajoutés : le premier permettant l'indexation des échantillons et le deuxième étant un adaptateur spécifique à la technologie Illumina (San Diego, Californie, EU). Le séquençage a ensuite été réalisé sur la plateforme MiSeq (Illumina) avec le kit MiSeq Reagent sur 600 cycles de séquençage.

### 1.3. TRAITEMENT BIOINFORMATIQUE DES SEQUENCES 16S

Les séquences obtenues ont ensuite été analysées par groupe de séquençage avec le package R dada2 (Callahan *et al.*, 2016). Après un contrôle qualité en suivant les recommandations des auteurs de DADA2, l'algorithme DADA2 a permis de regrouper les séquences en amplicon sequence variants (ASV), avec un seuil d'identité de 100%. Ce processus a abouti à la construction d'une table de comptage. Les tables spécifiques à chaque run de séquençage ont ensuite été assemblées et les chimères ont été éliminées grâce à la fonction removeBimeraDenovo() du package R dada2 (Callahan *et al.*, 2016). Un total de 37 138 ASVs a été identifié. Enfin, la librairie SILVA v.138 (Quast *et al.*, 2013) a été utilisée pour l'assignement taxonomique de chaque séquence à tous les niveaux taxonomiques, *i.e.*, du règne à l'espèce, lorsque la base de données le permettait (sinon le terme « inconnu » était employé).

### 1.4. CREATION DES PHENOTYPES DE COMPOSITION DU MICROBIOTE FECAL

A partir de la table de comptage agglomérée au niveau du genre, les abondances ont été normalisées en utilisant la méthode des ratios log centrés (centered-log ratio – CLR) telle qu'implémentée dans le package R microbiome (Lahti and Shetty, 2017). Cette dernière ajoute un pseudo-compte égal à la moitié de l'abondance relative minimale. Les abondances CLR des 90 genres, avec une prévalence supérieure à 60%, identifiés comme héritables (h² entre 0,05 et 0,21) dans une étude précédente (Brulin *et al.*, 2024b) ont ensuite été ajustées pour l'ensemble des effets non-génétiques avec l'option --blup implémenté dans le logiciel Wombat (Meyer, 2007), en utilisant le modèle linéaire suivant :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{e} \quad [1]$$

avec **y**, le vecteur des abondances CLR des genres, $\boldsymbol{\beta}$ le vecteur des effets non-génétiques et la matrice d'incidence associée $\boldsymbol{X}$ ; $\boldsymbol{\alpha} \sim N(0, \boldsymbol{A}\sigma_\alpha^2)$ le vecteur des effets génétiques aléatoires des animaux avec $\boldsymbol{Z}$, la matrice d'incidence, $\boldsymbol{A}$ la matrice de parenté et $\sigma_\alpha^2$, la variance génétique ; $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2)$ le vecteur des effets aléatoires résiduels, avec $\boldsymbol{I}$, la matrice d'identité et $\sigma_e^2$, la variance résiduelle. Les effets fixes non-génétiques incluaient le cheptel, le groupe de séquençage, la saison de prélèvement, le jour de lactation, la parité et le nombre de séquences dans l'échantillon.

### 1.5. IMPUTATION DES GENOTYPES A LA SEQUENCE

L'ensemble des animaux a été génotypé avec la puce Illumina BovineSNP50 (50K) BeadChip contenant 53 469 SNPs dans sa version actuelle. Les génotypes ont ensuite été imputés jusqu'à la séquence complète en deux étapes. Dans un premier temps, les génotypes ont été imputés de la densité 50 K à la densité 777 K (HD) SNPs avec le logiciel FImpute (Sargolzaei *et al.*, 2014) et une population de référence de 804 animaux de race Holstein génotypés avec la puce Illumina BovineHD BeadChip. La deuxième étape d'imputation, de la densité HD à la séquence, a été réalisée avec le logiciel Minimac (Howie *et al.*, 2012) et la population de référence du neuvième run du projet 1000 Génomes Bovins qui comprend les séquences de 3 414 bovins (dont 1 414 Holstein) (Daetwyler *et al.*, 2014). La précision de l'imputation HD à la séquence a été évaluée avec le R² calculé par Minimac. Parmi tous les variants, 13 420 634 parmi les mieux imputés (R² Minimac ≥ 0,20) et les moins rares (MAF > 0,005) ont été retenus pour les GWAS.

### 1.6. GWAS A LA SEQUENCE

Des analyses d'association ont été réalisées entre les 13 420 634 variants et les abondances CLR ajustées des genres héritables, avec l'option --mlma du logiciel GCTA (Yang *et al.*, 2010) qui applique un modèle linéaire mixte pour chaque variant à tester :

$$\boldsymbol{y} = \boldsymbol{1}\mu + \boldsymbol{x}b + \boldsymbol{u} + \boldsymbol{e} \quad [2]$$

dans lequel **y** est le vecteur des abondances CLR ajustées des genres héritables ; μ est la moyenne ; $b$ est l'effet additif du variant à tester ; **x** est le vecteur des doses des allèles imputés, avec des valeurs entre 0 et 2 ; $\boldsymbol{u}$ ~N(0, $\boldsymbol{G}\sigma_u^2$) est le vecteur des effets polygéniques aléatoires. $\boldsymbol{G}$ est la matrice de parenté génomique calculée à partir des génotypes 50K, et $\sigma_u^2$ est la variance polygénique.

Comme proposé par Sahana *et al*. (2023), nous avons considéré l'effet d'un variant significatif si sa -$\log_{10}$(p-*value*) ≥ 7,3. Les intervalles de confiance des QTL (quantitative trait loci) ont ensuite été définis en considérant l'ensemble des variants localisés dans le tiers supérieur du pic et à moins de 2 millions de paires de bases (pb) du variant le plus significatif. Le pourcentage de variance phénotypique expliquée par chaque QTL ($\%\sigma_p^2$) a été estimé avec la formule suivante :

$$\%\sigma_p^2 = 100\left(\frac{2p(1-p)\alpha^2}{\sigma_p^2}\right) \quad [3]$$

avec $p$ et $\alpha$, la fréquence et l'effet du variant le plus significatif du QTL.

## 2. RESULTATS

Les GWAS ont révélé cinq QTL significativement associés à *Paeniclostridium*, *Akkermansia*, un genre indéterminé de la famille des Paludibacteraceae, *Sutterella* et *Turicibacter* (Tableau 1 ; Figure 1). Ces QTL, respectivement localisés sur les chromosomes 9, 11, 14, 16 et 22, ne sont pas communs à plusieurs genres fécaux.

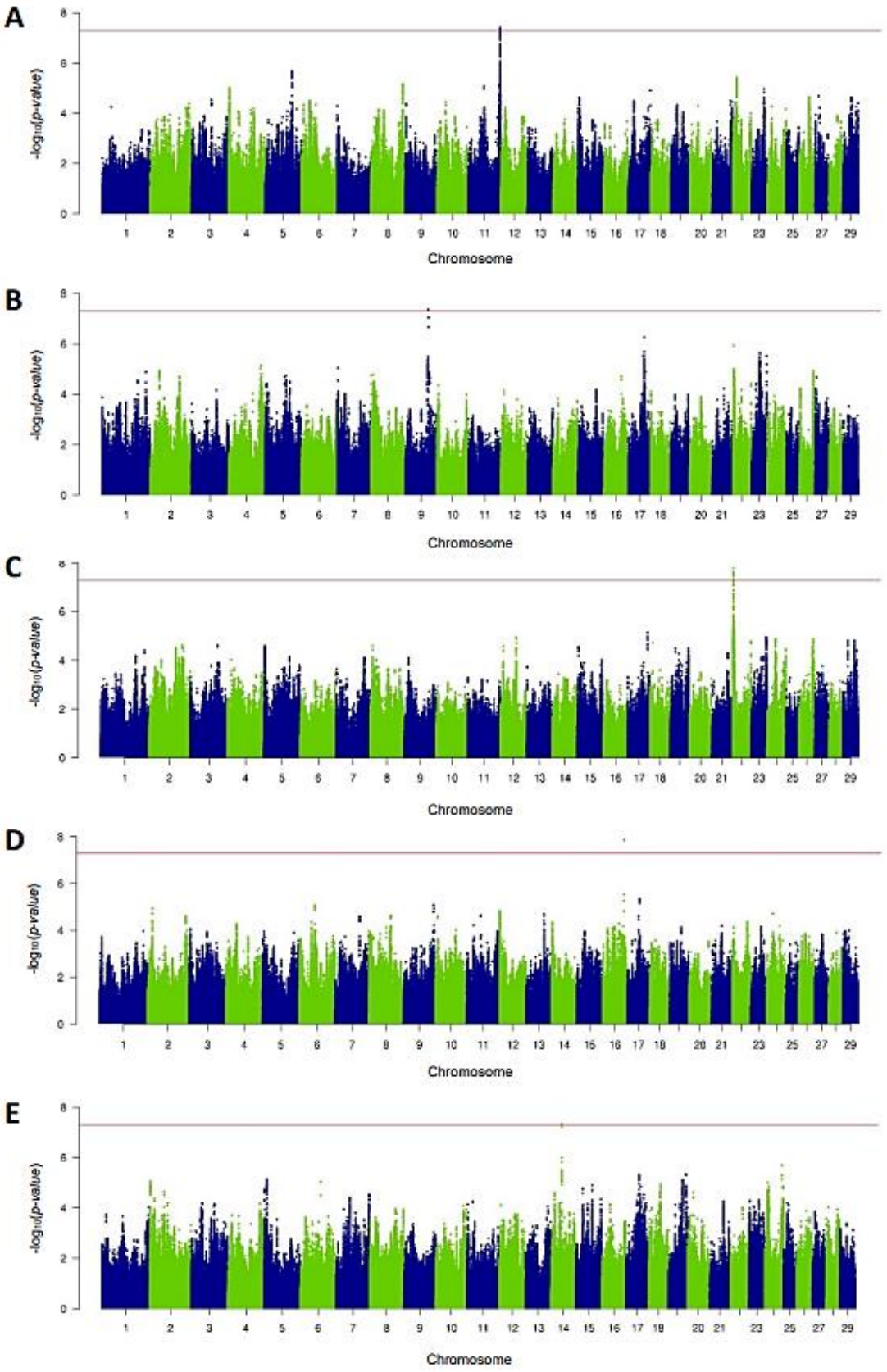


Figure 1 : Manhattan plots représentant les -$\log_{10}$(*p-value*) des effets des variants sur l'abondance *d'Akkermansia* (**A**), de *Paeniclostridium* (**B**), d'un genre inconnu de la famille des Paludibacteraceae (**C**), de *Sutterella* (**D**) et de *Turicibacter* (**E**) en fonction de leur position sur le génome de *bos taurus*. La ligne rouge horizontale représente le seuil de significativité de - $\log_{10}$(*p-value*) = 7,3.

Avec des intervalles de confiance compris entre environ 500 kb et 3 Mb, les QTLs observés regroupent entre 9 et 36 gènes chacun. Parmi les variants avec les effets les plus significatifs de chaque QTL, deux variants, sont particulièrement intéressants puisqu'associés à des fréquences (MAF) proches de 0,50. Ces derniers, pourvus de -$\log_{10}$(p-value) de 7,40 et 8,01, expliquent respectivement 2,26% et 2,02% des abondances d'*Akkermansia* et *Turicibacter*. De façon intéressante, le variant associé à l'abondance d'*Akkermansia* est localisé en aval du gène *ABO.* Les variants les plus significatifs des autres QTL, avec des MAF plus faible (1 à 2%) sont annotés dans des régions intergéniques et expliquent 4,38%, 5,90% et 25,49% de la variance phénotypique des abondances de *Paeniclostridium*, d'un genre inconnu de la famille des Paludibacteraceae et de *Sutterella*.

**Tableau 1** Caractéristiques des QTL (quantitative trait loci) significativement associés aux genres du microbiote fécal des vaches Holstein.

| Genres | QTL | | | Variants avec les effets les plus significatifs | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | BTA | Bornes de l'IC (pb) | Nombre de gènes | Position (pb) | -log(p-*value*) | b ± SE | $\%\sigma_p^2$ | MAF | Annotation |
| *Akkermansia* | 11 | 103 501 870 - 104 240 306 | 33 | 104 174 750 | 7,40 | 1,67 ± 0,30 | 2,26 | 0,34 | Aval du gène *ABO* (<3 K pb) |
| *Paeniclostridium* | 9 | 73 845 265 - 76 787 603 | 36 | 75 786 168 | 7,37 | -6,26 ± 1,14 | 4,38 | 0,02 | Région inter-génique |
| Inconnu Paludibacteraceae | 14 | 35 347 692 - 35 842 081 | 10 | 35 842 081 | 7,34 | -7,78 ± 1,42 | 5,90 | 0,01 | Région inter-génique |
| *Sutterella* | 16 | 69 370 397 - 71 115 512 | 24 | 69 370 397 | 8,65 | -17,86 ± 2,99 | 25,49 | 0,02 | Région inter-génique |
| *Turicibacter* | 22 | 1 871 656 - 3 025 335 | 9 | 2 552 694 | 8,01 | 1,19 ± 0,21 | 2,02 | 0,44 | Région inter-génique |

Abréviations : BTA = *Bos taurus* autosome ; IC = intervalle de confiance ; b = effet du variant ; MAF = fréquence de l'allèle mineur ; pb = paires de bases ; $\%\sigma_p^2$ = pourcentage de variance phénotypique expliquée ; SE = erreur standard.

## 3. DISCUSSION

La réalisation de GWAS à la séquence a permis d'identifier précisément des régions du génome bovin associées à l'abondance de certains genres du microbiote fécal dans la race Holstein. Cette étude, qui est à notre connaissance la première réalisée chez les bovins, a révélé cinq QTL associés à *Paeniclostridium*, *Akkermansia*, *Sutterella*, *Turicibacter* et un genre inconnu de la famille des Paludibacteraceae. En dépit du fait que nous n'ayons identifié des QTL que pour un nombre restreint de genres, ces résultats sont en accord avec d'autres études menées sur le microbiote ruminal (Li *et al.*, 2019 ; Martinez Boggio *et al.*, 2022). Il faut cependant noter que les QTL identifiés dans la présente étude diffèrent de ceux observés dans le microbiote ruminal, impliquant des différences au niveau des mécanismes de déterminisme génétique influençant la composition de ces deux microbiotes d'intérêt. Par la suite, il pourrait être intéressant de confronter les résultats de GWAS obtenus avec d'autres rangs taxonomiques. Dans le cadre de cette étude, le rang taxonomique des genres nous a permis de bénéficier à la fois d'un certain détail sur les spécificités taxonomiques des bactéries, tout en gardant, contrairement aux ASVs, des prévalences élevées et en limitant les biais induits par le séquençage 16S.

Les cinq genres identifiés dans cette étude sont tous des genres communément observés dans le tractus digestif des bovins (Cendron *et al.*, 2020 ; Fan *et al.*, 2021 ; Jiang *et al*., 2023). Notre étude a notamment mis en avant une association intéressante entre le gène *ABO* et l'abondance d'*Akkermansia*. Ces résultats font écho à l'action du gène *ABO* sur l'abondance de certains genres du microbiote fécal chez l'Homme (Qin *et al.*, 2022) et le porc (Yang *et al.*, 2022). A notre connaissance, notre étude est toutefois la première à rapporter une association entre *ABO* et *Akkermansia*. Le genre *Akkermansia*, avec une héritabilité de 0,21 (Brulin *et al.*, 2024b) et une abondance relative (AR) moyenne de 1% ($max_{AR}$ = 19%) dans nos études, peut s'avérer être bénéfique pour son hôte, en le protégeant des pathogènes si la ration apportée est suffisamment riche en fibre (Wolter *et al.*, 2024). Une autre association pertinente impliquait *Turicibacter*. En effet, malgré une héritabilité ($h^2$ = 0,08) et une abondance relative moyenne (AR = 0,6%, $max_{AR}$ = 11%) plus faibles, 2,01% de la variance phénotypique de ce genre ont pu être expliqué par un variant avec une MAF assez élevée. Ce dernier est notamment impliqué dans la production et composition de la bile (Lynch *et al.*, 2023). Cependant, ce variant est localisée dans une région intergénique, rendant l'élucidation du mécanisme génétique plus complexe. Des variants significatifs en région intergénique ont également été mis en avant pour *Paeniclostridium*, un genre inconnu de la famille des Paludibacteraceae et *Sutterella*. *Paeniclostridium* est le genre identifié comme le plus héritable ($h^2$ = 0,21) avec une abondance de 2% ($max_{AR}$ = 13%). Cependant, ce genre est considéré comme un pathogène, impliqué dans des infections du tractus digestif des bovins (Wang et al., 2022, Jiang et al., 2023). Au contraire, la famille des Paludibacteraceae associée ici à un genre inconnu héritable ($h^2$ = 0,16, AR = 1,3%, $max_{AR}$ = 7%) est plutôt enrichie dans les animaux en bonne santé (Gu *et al.*, 2023) et dégrade la lignocellulose permettant la valorisation des aliments en acides gras volatils. *Sutterella* ($h^2$ = 0,14 ; AR = 0,07% ; $max_{AR}$ = 24%) est également un genre impliqué dans la santé de l'hôte, grâce à son action pro-inflammatoire permettant d'entrainer la réponse immunitaire de l'hôte (Jiang et al., 2023). La grande proportion de variance expliquée pour *Sutterella* (25,50%) est à considérer avec précaution en raison de la MAF faible du variant associé (MAF = 0,015). Cependant, l'obtention de variants avec des effets significatifs sur la composition du microbiote dans une population commerciale, impliquant un plus grand nombre d'animaux mais des informations limitées sur les conduites d'élevage, suggèrent la possibilité d'utiliser les données génomiques pour prédire et sélectionner la composition du microbiote fécal afin de favoriser un microbiote bénéfique pour la vache et la filière laitière.

## CONCLUSION

Cette étude a permis l'identification de zones du génome de la Holstein qui sont associées à l'abondance de certains genres du microbiote fécal. Ces résultats sont prometteurs à plusieurs égards : ils ouvrent de nouvelles perspectives de recherche afin de mieux appréhender les relations entre l'hôte et son microbiote, tout en soulignant le potentiel d'utilisation de marqueurs génomiques pour une sélection ciblée de la composition du microbiote chez la vache laitière.

*Les auteurs souhaiteraient remercier les éleveurs et les techniciens qui ont permis la collection des échantillons fécaux, ainsi que Gérard Heumez (Gènes Diffusion) pour son aide dans la collecte des données.*
*Les auteurs voudraient également remercier la base nationale française (CTIG) et Valogène pour avoir fourni les données ; et la contribution du 1 000 Bull Genome Project. Par ailleurs, les auteurs remercient également le centre de bioinformatique MIGALE (INRAE) (doi : 10.15454/1.5572390655343293E12) qui a apporté les ressources computationnelles nécessaires aux analyses. GD Biotech souhaiterait également exprimer sa gratitude envers ANR-10-EQPX-07-01 LIGAN-PM-UMR8199 pour son support technique avec la plateforme de séquençage MiSeq.*

**Brulin L., Ducrocq S., Estellé J., Even G., Martel S., Merlin S., Audebert C., Croiseau P., Sanchez M.P., 2024a,** animal**.** 18(8):101243.
**Brulin L., Ducrocq, S., Estellé, J., Even, G., Martel, S., Merlin, S., Audebert, C., Croiseau, P., Sanchez, M.P., 2024b**. J. Dairy Sci., accepté le 8 août 2024.
**Callahan B.J., McMurdie P.J. Rosen M.J., Han A., Holmes S.P., 2016**. Nat Methods. 13 (7) 581-583.
**Cendron , F., Niero G., Carlino G., Penasa M., Cassandro M., 2020**. J Appl Genet. 61 593-605.
**Daetwyler H.D., Capitan A., Pausch H., Stothard P., van Binsbergen R., Brøndum R.F., Liao X., Djari A., Rodriguez S.C., Grohs C., Esquerré D., Bouchez O., Rossignol M.N., Klopp C., Rocha D., Fritz S., Eggen A., Bowman P.J., Coote D., Chamberlain A.J., Anderson C., VanTassell C.P., Hulsegge I., Goddard M.E., Guldbrandtsen B., Lund M.S., Veerkamp R.F., Boichard D.A., Fries R., Hayes B.J., 2014.** Nat Genet. 46, 858-865
**Fan P., Nelson C.D. Driver J.D., Elzo M.A., Peñagaricano F., Jeong K.C., 2021.** ISME J. 15, 2306-2321.
**Gu F., Zhu S., Tang Y., Liu X., Jia M., Malmuthuge N., Valencak T.G., McFadden J.W., Liu J.X., Sun H.Z., 2023.** Microbiome. 11:40.
**Howie B., Fuchsberger C., Stephens M., Marchini J., Abecasis G.R., 2012.** Nat Genet. 44(8), 955-959.
**Jami E., Whilte B.A., Mizrahi I., 2014.** PLoS ONE, 9 (1): e85423
**Jiang C., Hou X., Gao X., Liu P., Guo X., Hu G., Li Q., Huang C., Li G., Fang W., Mai W., Wu C., Xu Z., Liu P., 2023.** BMC Microbiol. 23:180.
**Lahti L., Shetty S., 2017. v.1.24.0.**
**https://github.com/microbiome/microbiome**
**Li F., Li C., Chen Y., Liu J., Zhang C., Irving B., Fitzsimmons C., Plastow G., Guan L.L., 2019.** Microbiome. 7:92.
**Lynch J.B., Gonzales E.L., Choy K., Faull K.F., Jewell T., Arellano A., Liang J., Yu K.B., Paramo J., Hsiao E.Y., 2023.** Nat Commun. 14:3669
**Martinez Boggio G., Meynadier A., Buitenhuis AJ., Marie-Etancelin C., 2022.** Genet Sel Evol. 54, 77-96
**Meyer K., 2007.** J Zhejiang Univ Sci B. 8(11), 815-821.
**Monteiro H.F., Zhou Z., Gomes M.S., Peixoto P.M.G., Bonsaglia E.C.R., Canisso I.F., Weimer B.C., Lima F.S.. 2022**. Sci Rep., 12, 4904-4917
**Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner F.O., 2013.** Nucleic Acids Research. 41, 590-596.
**Qin Y., Havulinna A.S., Liu Y., Jousilahti P., Ritchie S.C., Tokolyi A., Sanders J.G., Valsta L., Brożyńska M., Zhu Q., Tripathi A., Vázquez-Baeza Y., Loomba R., Cheng S., Jain M., Niiranen T., Lahti L., Knight R., Salomaa V., Inouye M., Méric G., 2022.** Nat. Genet. 54, 134-142
**Sahana G., Cai Z., Sanchez M.P., Boichard D., 2023.** J. Dairy Sci. 106, 5218-5241
**Sargolzaei M., Chesnais J.P., Schenkel F.S., 2014.** BioMed Central. 15, 478-490
**Wang Y., Nan X., Zhao Y., Jiang L., Wang H., Zhang F., Hua D., Liu J., Yang L., Yao J., Xiong B., 2022.** J. Dairy Sci. 105, 7668-7688.
**Wolter M., Grant E.T., Boudaud M., Pudlo N., Pereira G.V., Eaton K.A., Martens E.C. Desai M.S., 2023.** bioRxiv
**Xue M., Sun H., Wu X., Guan L.L., Liu J. 2018.** Appl Environ Microbiol. 84: e00970-18
**Yang H., Wu J., Huang X., Zhou Y., Zhang Y., Liu M., Liu Q., Ke S., He M., Fu H., FangS., Xiong X., Jiang X., Chen Z., Wu Z., Gong H., Tong X., HuangY., Ma J., Gao J., Charlier C., Coppieters W., Shagam L., Zhang Z., Ai H., Yang B., Georges M., Chen C., Huang L., 2022.** Nature. 606, 358-367
**Yang J., Lee S.H., Goddard M.E., Visscher P.M., 2010.** Am J Hum Genet. 88, 76-82