# Sequence-based genome-wide association study reveals host genomic regions and candidate genes influencing the fecal microbiota of Holstein cows

L. Brulin, M.-P. Sanchez, Z. Cai, S. Ducrocq, G. Even, S. Martel, S. Merlin, C. Audebert, J. Estellé, G. Sahana, et al.

▶ **To cite this version:**

L. Brulin, M.-P. Sanchez, Z. Cai, S. Ducrocq, G. Even, et al.. Sequence-based genome-wide association study reveals host genomic regions and candidate genes influencing the fecal microbiota of Holstein cows. Journal of Dairy Science, 2025, TBC (TBC), ⟨10.3168/jds.2024-26203⟩. ⟨hal-05120249⟩

**HAL Id: hal-05120249**
**https://hal.inrae.fr/hal-05120249v1**

Submitted on 19 Jun 2025

**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.

Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

ARTICLE IN PRESS—UNCORRECTED PROOF

J. Dairy Sci. TBC
https://doi.org/10.3168/jds.2024-26203
© TBC, The Authors. Published by Elsevier Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

# Sequence-based genome-wide association study reveals host genomic regions and candidate genes influencing the fecal microbiota of Holstein cows

**L. Brulin,[1,2]* M.-P. Sanchez,[2] Z. Cai,[3] S. Ducrocq,[1,4] G. Even,[1,4] S. Martel,[1,4] S. Merlin,[1,4] C. Audebert,[1,4] J. Estellé,[2] G. Sahana,[3] and P. Croiseau[2]**
[1]GD Biotech – Gènes Diffusion, Lille, 59000, France
[2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France
[3]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, 8000, Denmark
[4]PEGASE-Biosciences, Institut Pasteur de Lille, Lille, 59019, France

## ABSTRACT

In recent decades, the digestive tract microbiota of livestock has been extensively studied, revealing associations with host phenotypes, including production- and health-related traits. The effect of host genetics on gut microbes has been documented in several species; however, in dairy cattle, the specific genomic regions that influence microbial communities remain relatively unexplored. This study aimed to conduct a sequence-based GWAS and a gene-based association study to identify the genomic regions and candidate genes affecting fecal microbiota diversity and composition in a population of 1,875 commercial Holstein cows. From the sequence-based GWAS conducted on 116 fecal microbiota taxonomic levels, 6 QTL were significantly associated with the abundances of *Paeniclostridium*, an unclassified genus from the Paludibacteraceae family, *Sutterella*, *Turicibacter*, and *Akkermansia* genera, as well as the associated family Akkermansiaceae. These QTL explained between 2.0% and 25.5% of the phenotypic variances of the taxa abundances. Conversely, no genomic variants were found significant for either the α- or the β-diversity of the fecal microbiota. A gene-based association study subsequently conducted on the sequence-based GWAS results revealed significant effects of 90 genes across the bovine genome, effecting the relative abundances of some fecal taxa. Many of these genes were located within the major histocompatibility complex and enriched in immune response pathways. By combining GWAS with gene-based association studies, we specifically identified an association between the ABO gene and the fecal abundance of *Akkermansia* and Akkermansiaceae. The study represents a significant step forward in understanding the genetic determinism of the complex interactions between the fecal microbiota and their host. It provides new insights into the biological mechanisms underlying host-microbiota interaction in dairy cattle and unveils strong associations between host genomic regions and fecal microbiota in a commercial population. This study holds promise for large-scale breeding strategies to shape the fecal microbiota in Holstein cows and benefit from the host-microbiota interactions.
**Key words:** dairy cattle, whole genome-wide association study, 16S analysis, fecal microbiota

## INTRODUCTION

Over the past decades, numerous studies have highlighted the significant roles of microbes inhabiting the digestive tract of livestock in influencing host phenotypes. In ruminants, microbiota from various digestive tract compartments play a crucial role in feed digestion. The rumen bacterial community, located at the beginning of the digestion process, ferments the diet, enabling the valorization of nutrients that the host alone cannot metabolize. This substantial influence on digestion largely explains the links between rumen bacteria and various traits, including feed efficiency (Monteiro et al., 2022; Martinez-Boggio et al., 2024), methane emissions (Zhang et al., 2020; Martínez-Álvaro et al., 2022), and milk traits (Xue et al., 2018; Liu et al., 2022a). Additionally, microbial fermentation occurs in the lower parts of the ruminant gastrointestinal tract, where microorganisms have also been associated with feed efficiency (Fan et al., 2021; Monteiro et al., 2022), dairy production (Monteiro et al., 2022; Brulin et al., 2024b), and health issues (Zhang et al., 2019). It is well established that gastrointestinal microbiotas are influenced by factors such as diet (Kim et al., 2014; Klevenhusen et al., 2017) and

Received December 20, 2024.
Accepted April 29, 2025.
*Corresponding author: l.brulin@genesdiffusion.com

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-25. Nonstandard abbreviations are available in the Notes.

ARTICLE IN PRESS—UNCORRECTED PROOF



age (Jami et al., 2013; Zhang et al., 2019; Cendron et al., 2020). However, several studies across various ruminant species have also demonstrated that their diversity and composition are partly heritable (Wallace et al., 2019; Martinez Boggio et al., 2022; Martínez-Álvaro et al., 2022). In dairy cattle, Brulin et al. (2024a) showed that the fecal microbiota of Holstein cows are associated with low to moderate $h^2$ estimates ($h^2$ between 0.08 and 0.32). Furthermore, many fecal bacteria abundances exhibited genetic correlations with other taxa abundances, diversity indices, and host phenotypes.

Given that digestive microbiotas are partly heritable and influenced by host genetics, the mechanisms connecting the host genome to the gastrointestinal microbial community are gaining interest. To this end, GWAS have been performed on various species to identify the underlying QTL for taxa abundances. In cattle and sheep, various genomic variants were significantly associated with the abundance of several fecal (Fan et al., 2021) and ruminal taxa (Li et al., 2019; Martinez Boggio et al., 2022). However, no consensus has been reached regarding the regions of the bovine genome that influence digestive microbiotas, and only a few candidate genes have been proposed. The absence of shared findings may be attributed to the complex nature of microbiota-based traits and limited sample sizes, aside from external factors differentiating the studies (e.g., animal diet, laboratory protocols). Therefore, a sequence-based GWAS appears to be an appealing solution for enhancing detection power and identifying potential candidate genes.

Given that some taxa constituting the fecal microbiota are heritable, we hypothesized that future modulation of these microbiota could be achieved through genomic selection. Furthermore, these microbiota show numerous associations with host phenotypes, and their ease of collection makes it relevant to the dairy industry. Therefore, in this study, we sought to broaden our understanding of the genetic link between dairy cattle and their microbiota by identifying the genomic regions and genes that influence the fecal microbiota of Holstein cows. To achieve this, we first carried out a sequence-based GWAS to finely map QTL associated with fecal microbiota traits. These GWAS results were then analyzed using a gene-based association study to identify the most likely candidate genes and enriched gene sets.

## MATERIALS AND METHODS

### Microbiota Samples

A population of 1,875 lactating Holstein cows (DIM between 5 and 715) from 144 commercial farms, previously used in Brulin et al. (2024a), was considered for this study. As described in Brulin et al. (2024a), ~50 g of feces from each cow was collected by technicians from Gènes Diffusion (Douai, France) using digital rectal stimulation. Samples were immediately preserved in 30 mL sterile tubes containing 15 mL of a custom-made ethanol-based preservation solution, similar to the one described by Marotz et al. (2021). The fecal samples were transported and stored at 4°C in the Gènes Diffusion research laboratory (Institut Pasteur de Lille, France). Samples were divided into 18 laboratory batches, and subsequent analyses, including DNA extraction, *16S* rRNA gene sequencing, bioinformatic processing, and taxonomic assignment, were performed as detailed in Brulin et al. (2024a). Briefly, following the sequencing and quality control step, sequences from each batch were grouped into amplicon sequence variants (**ASV**) using the DADA2 algorithm (Callahan et al., 2016), creating a count table. The count tables of each batch were merged, and chimeras were removed. A total of 37,138 ASV were identified and assigned to all taxonomic ranks from phyla to species with the SILVA v.138 database (Quast et al., 2013). The ASV were categorized into 474 genera, 208 families, and 33 phyla, with a predominance of the Firmicutes phylum (54.9% of all sequences).

### Adjustment of Microbiota Diversities and Taxa Abundances into Microbiota-Based Phenotypes

Two types of variables were constructed to consider both the diversity and composition of the fecal microbiota. For diversity, α- and β-diversity indices were estimated with the vegan R package (v.2.6-6; Oksanen et al., 2022) for the count table that was rarefied to 5,000 reads using the phyloseq R package (v.1.46.0; McMurdie and Holmes, 2013) to limit the effect of sequencing depth. Specifically, α-diversity was estimated with the Shannon index, and the β-diversity was assessed using the first 2 axes of nonmetric multidimensional scaling (**NMDS**) of the Bray-Curtis dissimilarity matrix (i.e., NMDS1 and NMDS2). To account for composition in taxa abundances, the centered-log ratio (**CLR**) abundances of 41 ASV, 40 genera, and 26 families, previously identified as heritable (Brulin et al., 2024a), were considered. The CLR abundances were calculated on the nonrarefied count table using the microbiome R package (v.1.24.0; Lahti and Shetty, 2017). For both categories of phenotypes, after removing outliers (mean ± 4 SD), CLR abundances were adjusted for nongenetic effects using the Wombat software (Meyer, 2007) with the *–blup* specific option, using the following mixed animal model

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

where **y** is the vector of microbiota-based phenotypes (diversity indices and taxa CLR abundances); **β** is the vector of nongenetic fixed effects with **X** as the incidence matrix; $\pm \sim \mathbf{N}\left(\mathbf{0},\ \mathbf{A}\tilde{A}^2_{\pm}\right)$ is the vector of random animal genetic effects, **A** being the relationship matrix and $\tilde{A}^2_{\pm}$ the genetic variance; and $\mathbf{e} \sim \mathbf{N}\left(\mathbf{0},\mathbf{I}\tilde{A}^2_{\mathbf{e}}\right)$ is the vector of random residual effects, with **I** as the identity matrix and $\tilde{A}^2_{e}$ the residual variance. For both diversity indices and taxa CLR abundances, fixed effects included herd-lab batch group, parity (4 levels: first, second, third, and ≥fourth parity) and lactation stage (24 classes: 15-d stages from 5 to 320 DIM, 45-d stages from 320 to 365 DIM, 40-d stage from 365 to 405 DIM, and ≥405 DIM). Additionally, for taxa CLR abundances, sequencing depth was included in the fixed effects of the model. All fixed effects were considered, as they effected the overall diversity or specific taxa abundances (Brulin et al., 2024a,b).

### Genotyping and Imputation

All cows were genotyped using the Illumina BovineSNP50 (50k) BeadChip (Illumina Inc., San Diego, CA). Quality control filters were applied to variants, including individual call rate >95%, SNP call rate >90%, minor allele frequency (**MAF**) >1% in the French Holstein breed, and genotype frequencies in Hardy-Weinberg equilibrium with $P > 10^{-4}$. After quality control, 53,469 autosomal SNPs remained.

Subsequently, the genotypes of the sampled cows were imputed to whole genome sequencing (**WGS**) in 2 steps for increased imputation accuracy (Daetwyler et al., 2014). First, genotypes were imputed from 50k to 777k high-density (**HD**) SNPs using the FImpute software (Sargolzaei et al., 2014), utilizing a reference population of 804 Holstein animals genotyped with the Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA). Second, allele dosages were imputed to WGS using the Minimac software (Howie et al., 2012) using 3,414 *Bos taurus* animals (including 1,414 Holstein individuals) originating from the merge of a local small genomic variants database (Boussaha et al., 2016) with the ninth run of the 1000 Bull Genomes Project (Daetwyler et al., 2014; Bouwman et al., 2018) as a reference. Our local database was produced by analyzing WGS for 571 animals using the guidelines recommended by the 1000 Bull Genomes project (Hayes and Daetwyler, 2019). Briefly, raw paired read sequences were first trimmed of adapters and low-quality bases (qscore <20) at the beginning and end. Reads with mean qscore <20 or length <35 bp were then filtered out using the Trimmomatic software v.0.38 (Bolger et al., 2014). Sequence alignments were performed using the Burrows-Wheeler Alignment tool (BWA 0.7.17; Li; unpublished data) with the *mem* option and default parameters for mapping reads to the ARS-UCD1.2 bovine reference genome. Potential PCR duplicates, which can adversely effect the variant calls, were removed using the MarkDuplicates tools from the Picard package version v2.18.2 (Broad Institute, 2019). Base quality recalibration was performed according to GATK best practices guidelines (Van der Auwera and O'Connor, 2020). Only properly paired reads with a mapping quality of at least 30 (−q = 30) were retained. The resulting BAM files were used to produce a GVCF file for each sample, and variant calling was performed from all GVCF files using the GATK GenotypeGVCFs options (https://gatk.broadinstitute.org/hc/en-us/articles/360037057852-GenotypeGVCFs).

Subsequently, we phased our genomic dataset with the Beagle v.5.4 software (Browning et al., 2021; https://faculty.washington.edu/browning/beagle/beagle.html) using the 1,000 GB phased dataset as an input reference panel. Finally, we merged our phased local database with the ninth 1,000 GB phased dataset, selecting only biallelic variants with a MAF of 5% and a minimum genotype quality of 100. This resulted in a final dataset of 25,050,323 high-quality variants. The precision of imputation from HD to WGS was assessed using $R^2$ values calculated with Minimac software (Howie et al., 2012). For this analysis, 13,420,634 variants were retained, with $R^2$ ≥0.20 and MAF ≥0.005. The low $R^2$ threshold ($R^2$ ≥0.20) was chosen to address the underestimation of prediction accuracy by Minimac software (Nguyen et al., 2024).

### Whole Genome Sequence Association Analyses

Single-trait association analyses between the 13,420,634 variants and adjusted microbiota-based phenotypes were performed using the *–mlma* option of GCTA software (v.1.26.0; Yang et al., 2011), applying a mixed linear model with the variant to be tested

$$\mathbf{y} = 1\mu + \mathbf{W}\delta + \mathrm{u} + \mathrm{e}, \qquad [2]$$

where **y** is the vector of adjusted microbiota-based phenotypes (CLR abundance of taxa or diversity indices adjusted for all nongenetic effects as proposed in Equation 1); μ is the overall mean; δ is the additive fixed effect of the variant to be tested for association; **W** is the vector of imputed allele dosages, continuously varying from 0 to 2; and $\boldsymbol{\mu} \sim \mathbf{N}\left(\mathbf{0},\mathbf{G}\tilde{A}^2_{\mathbf{u}}\right)$ is the vector of random polygenic effects, **G** being the genomic relationship matrix constructed with the 50k genotypes, and $\tilde{A}^2_{u}$ the polygenic variance, estimated on the null model (y = 1μ + u + e) and then fixed while testing for the association between

ARTICLE IN PRESS—UNCORRECTED PROOF



each variant and the phenotype. Additionally, $\mathbf{e} \sim \mathbf{N}\left(\mathbf{0},\mathbf{I}\tilde{\mathrm{A}}_{\mathrm{e}}^{2}\right)$ is the vector of random residual effects, with $\mathbf{I}$ as the identity matrix and $\tilde{\mathrm{A}}_{e}^{2}$ the residual variance. A *t*-statistic test was employed for testing the association, consisting of dividing the variant effect estimate by its SE.

Considering the large number of variants, a correction for multiple testing was applied with a $-\log_{10}(P)$ threshold of 7.3, corresponding to 1 million independent tests, as Sahana et al. (2023) recommended for conducting GWAS in dairy cattle. This threshold is highly conservative, particularly given the high linkage disequilibrium among neighboring SNPs due to the low effective population size in dairy cattle populations (Doekes et al., 2018; Makanjuola et al., 2020; Gautason et al., 2021). Quantitative trait locus CI were defined based on the most distant variants included in the upper third of the QTL peak (Visscher et al., 1996; Sanchez et al., 2017). The percentage of phenotypic variance explained by each QTL was calculated as follows:

$$\%\tilde{\mathrm{A}}_{\mathrm{p}}^{2} = 100\left(\frac{2\mathrm{p}\left(1-\mathrm{p}\right)\pm^{2}}{\tilde{\mathrm{A}}_{\mathrm{p}}^{2}}\right), \quad [3]$$

with $\tilde{\mathrm{A}}_{\mathrm{p}}^{2}$ as the phenotypic variance of the trait (diversity indices or taxa CLR abundances), p and α the frequency and the estimated substitution effect of the variant with the most significant effect in the QTL region, respectively. Precise information on QTL, such as the number of genes, was obtained using the Ensembl ARS-UCD1.3 database (Harrison et al., 2024).

### Gene-Based Association Study

Gene-based association study analyses were performed using the MAGMA software v.1.10 (de Leeuw et al., 2015), which employs a multiple regression model to identify genes most likely associated with the traits. We obtained the cow gene information file (ARS-UCD1.3) from Ensembl Genes 112 (Harrison et al., 2024), containing 36,075 genes across the 29 autosomes. Using MAGMA software (–annotate nonhuman), SNPs were annotated to the genes, considering a 10-kbp window around each gene. A total of 7,694,794 SNPs (57.34%) were annotated to at least 1 gene, and 33,743 genes were mapped with at least 1 SNP. Linkage disequilibrium between SNPs was accounted for using 1,187 Holstein animals from the ninth 1000 Bull Genome (Daetwyler et al., 2014; Bouwman et al., 2018). All *P*-values were adjusted for multiple testing with a false discovery rate (**FDR**) of 5% (q-values), and only genes with a q-value ≤0.05 were declared significantly associated with a trait. All genes were analyzed with the Ensembl ARS-UCD1.3 database (Harrison et al., 2024), and tissue enrichment was checked with the Cattle GTEX database (Liu et al., 2022).

### Gene-Set Enrichment Analysis

Considering all candidate genes identified from our gene-based association study, we conducted a gene-set enrichment analysis using ShinyGo software (v.0.80; Ge et al., 2020). We examined gene ontology (**GO**) terms and pathways from the Kyoto Encyclopedia of Genes and Genomes (**KEGG**). Only gene sets with a Benjamini-Hochberg adjusted *P*-value <0.05 were deemed significant. All other parameters were kept at their default settings.

## RESULTS

### Whole Genome Sequence Association Analyses

A total of 110 single-trait GWAS were performed using imputed whole genome sequences. Three GWAS were related to the diversity indices, considering 1 parameter for the α-diversity index (i.e., Shannon index) and 2 parameters for the β-diversity (NMDS1 and NMDS2). No variant was found to be significantly associated with these diversity-linked variables (Figure 1; Table 1).

The 107 other GWAS analyses focused on heritable taxa abundances, divided into 41 ASV (representing 15.4% of all sequences), 40 genera (33.3%), and 26 families (46.6%). Considering a *P*-value threshold of $-\log_{10}(P)$ >7.3, 6 QTL were identified (Figure 1; Table 1). Five QTL were significantly associated with the abundances of 5 genera: *Paeniclostridium* ($h^2$ = 0.21; relative abundance [**RA**] = 1.97%), *Akkermansia* ($h^2$ = 0.21; RA = 0.92%), an unclassified genus from Paludibacteraceae family ($h^2$ = 0.16; RA = 1.28%), *Sutterella* ($h^2$ = 0.14; RA = 0.08%), and *Turicibacter* ($h^2$ = 0.08; RA = 0.56%; Figure 1; Table 1). The phenotypic variance explained by these QTL ranged from 2.0% for *Turicibacter* to 25.5% for *Sutterella* (Table 1). Additionally, a sixth QTL (i.e., QTL3) explained 2.5% of the phenotypic variance of the Akkermansiaceae family, which was only composed of the *Akkermansia* genus. This QTL was located on BTA 11, overlapping QTL2. Notably, although they did not present the most significant associations, these 2 QTL exhibited the highest number of significantly associated variants (Table 1). The smallest CI of QTL location was found for QTL4 and the unclassified genus from the Paludibacteraceae family, spanning 494,389 bp, whereas the longest was observed for QTL1 and *Paeniclostridium* abundance, with 2.94 Mbp.

Among each QTL, the most significant variant had a $-\log_{10}(P)$ ranging from 7.3 (unclassified genus from

ARTICLE IN PRESS—UNCORRECTED PROOF



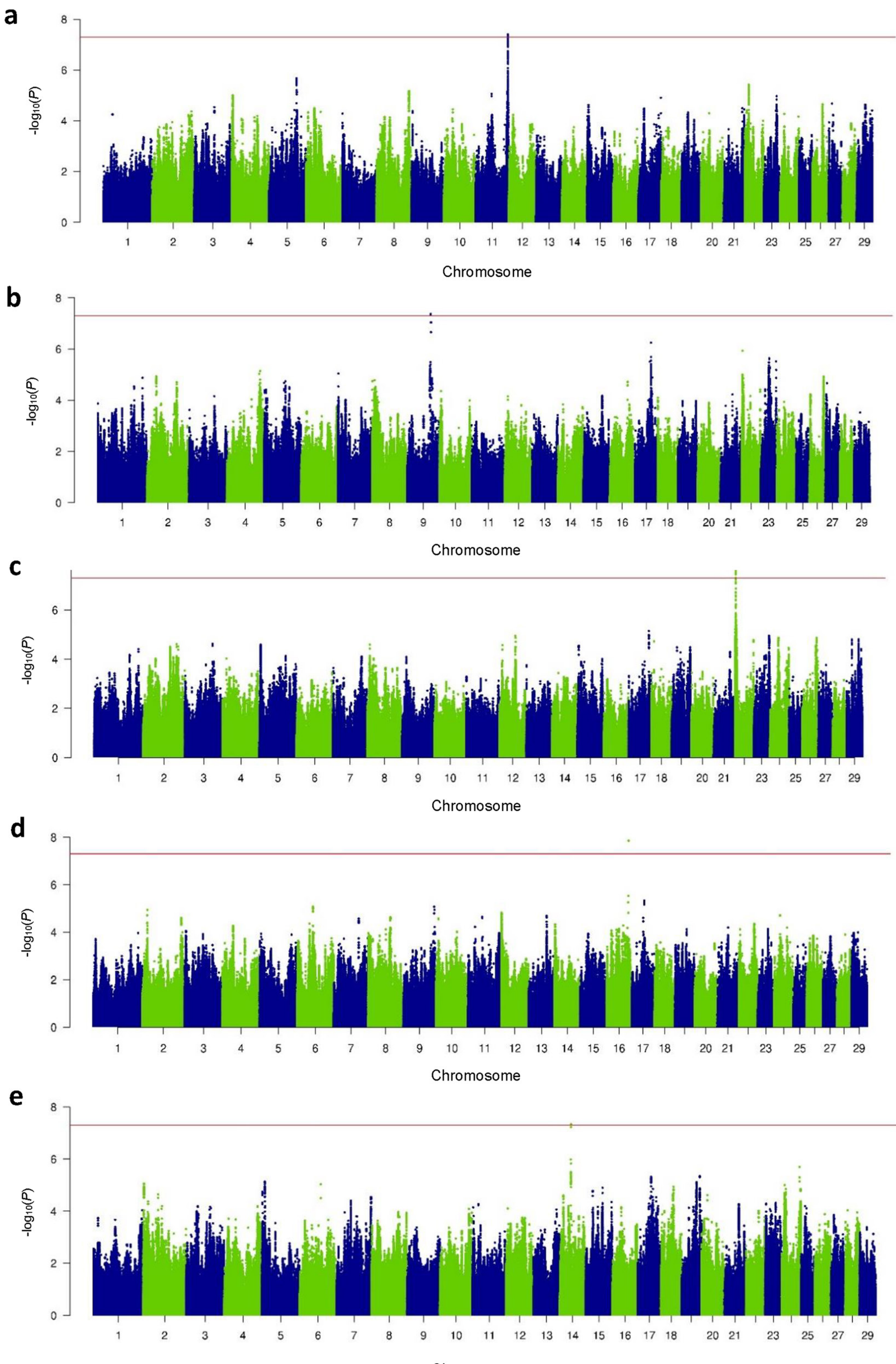


**Figure 1.** Manhattan plots with the $-\log_{10}(P)$ values plotted against the position of the variants on BTA for taxa abundances. (a) *Akkermansia*, (b) *Paeniclostridium*, (c) an unclassified genus from the Paludibacteraceae family, (d) *Sutterella*, and (e) *Turicibacter*. The red line represents the significance threshold of 7.3.

**Table 1.** Features of the QTL significantly associated with the abundances of fecal taxa[1]

| | QTL region | | | | | | Variant with the most significant effect | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | Taxa | BTA | From (bp) | To (bp) | # variants ($-\log_{10}(P) \geq 7.3$) | # genes | Id | Position (bp) | $-\log_{10}(P)$ | b | SE | $\%\sigma_p$ | Annotation | MAF | $R^2$ |
| 1 | *Paeniclostridium* | 9 | 73,845,265 | 76,787,603 | 2 | 36 | rs379744821 | 75,786,168 | 7.4 | −6.26 | 1.14 | 4.38 | Intergenic region | 0.017 | 0.374 |
| 2 | *Akkermansia* | 11 | 103,501,870 | 104,240,306 | 99 | 33 | rs208249234 | 104,174,750 | 7.4 | 1.67 | 0.30 | 2.26 | Downstream *ABO* | 0.340 | 0.915 |
| 3 | Akkermansiaceae | 11 | 103,501,870 | 104,240,306 | 119 | 33 | rs381382605 | 104,173,769 | 8.0 | 1.71 | 0.30 | 2.51 | Downstream *ABO* | 0.335 | 0.895 |
| 4 | Unclassified Paludibacteraceae | 14 | 35,347,692 | 35,842,081 | 1 | 10 | rs452257070 | 35,842,081 | 7.3 | −7.78 | 1.42 | 5.90 | Intron *ENSBTAG00000060358* | 0.012 | 0.312 |
| 5 | *Sutterella* | 16 | 69,370,397 | 71,115,512 | 3 | 24 | rs525096899 | 69,370,397 | 8.64 | −17.86 | 2.99 | 25.49 | Intergenic region | 0.015 | 0.207 |
| 6 | *Turicibacter* | 22 | 1,871,656 | 3,025,335 | 15 | 9 | rs110084268 | 2,552,694 | 8.0 | 1.19 | 0.21 | 2.02 | Intergenic region | 0.444 | 0.966 |

[1] b = additive fixed effect of the variant on the centered-log ratio abundance of the taxa; MAF = minor allele frequency; $R^2$ = accuracy of imputation; $\%\sigma_p$ = percentage of phenotypic variance explained by the QTL.

the Paludibacteraceae family) to 8.6 (*Sutterella*). These 2 variants also presented the strongest additive effect (**b**), with values of −7.78 and −17.86, respectively. The effect of the most significant variant of QTL1 (BTA 9) on the CLR abundance of *Paeniclostridium* (b = −6.26) was also higher than the effects of the most significant variants on QTL2, 3, and 6, which varied between 1.19 and 1.71. However, in addition to being associated with lower SE, these variants presented a better imputation accuracy ($R^2 \geq 0.90$), and their MAF were closer to 0.50, implying a better representation in the population. Conversely, the most significant variants on QTL4 and 5 both presented small MAF of 0.012 and 0.015, respectively, and weak $R^2$ values of 0.31 and 0.21, respectively (Table 2).

The QTL CI encompassed from 9 to 36 genes (Supplemental Table S1, see Notes). The most significant variants of QTL1, 5, and 6 were located in intergenic regions, whereas those associated with QTL2 and 3, related to *Akkermansia* and Akkermansiaceae, were located in the downstream region of the *ABO* gene. Other significant variants displayed notable annotations within *ABO*, including 2 stop gain variants, 1 missense variant, 1 splice region variant, and 1 synonymous variant (Table 2). The most significant variant associated with the abundance of an unclassified genus from the Paludibacteraceae family (QTL4) was annotated within an intron of the long noncoding RNA *ENSBTAG00000060358*, near the *KCNB2* gene.

## *Identification of Candidate Genes Associated With Fecal Microbiota*

The MAGMA gene-based association approach was used to search for candidate genes using the GWAS results described previously (Tables 3 and 4) to increase detection power. This method highlighted additional genome regions associated with the composition of the fecal microbiota, including regions not detected in the initial GWAS. Specifically, we identified 90 candidate genes across 17 autosomes for 5 ASV (12% of all heritable ASV), 8 genera (20% of all heritable genera), and 8 families (38% of all heritable families). Most of the genes identified were associated with the abundances of families and genera, with 63 and 60 genes, respectively (Table 3).

Forty-one genes were significantly associated with the abundance of multiple taxa across all taxonomic ranks (Table 4). Some genes were linked to taxa from the same lineage, such as *Akkermansia* and Akkermansiaceae, which were associated with 6 genes located on BTA 10, 11, and 19. Conversely, others were associated with taxa from different taxonomic groups; for example, the *Parasutterella* genus ($h^2$ = 0.21; RA = 0.65%) and the Spirochaetaceae family ($h^2$ = 0.08; RA = 0.91%) were

ARTICLE IN PRESS—UNCORRECTED PROOF



**Table 2.** Number of candidate genes significantly associated with the different heritable taxonomic levels and diversity indices

| Functional annotation | Taxa | % variant annotated |
|---|---|---|
| 3 prime UTR variant | *Akkermansia* | 20.2 |
| | Akkermansiaceae | 18.5 |
| Downstream gene variant | *Akkermansia* | 77.8 |
| | Akkermansiaceae | 73.1 |
| Intergenic region | *Akkermansia* | 55.5 |
| | Akkermansiaceae | 52.9 |
| | Unclassified Paludibacteraceae | 100 |
| | *Paeniclostridium* | 100 |
| | *Sutterella* | 100 |
| | *Turicibacter* | 100 |
| Intron variant | *Akkermansia* | 20.2 |
| | Akkermansiaceae | 23.5 |
| | Unclassified Paludibacteraceae | 100 |
| Missense variant | *Akkermansia* | 1.0 |
| | Akkermansiaceae | 0.008 |
| Splice region variant | *Akkermansia* | 1.0 |
| | Akkermansiaceae | 0.008 |
| Stop-gained | *Akkermansia* | 2.0 |
| | Akkermansiaceae | 0.02 |
| Synonymous variant | *Akkermansia* | 1.0 |
| | Akkermansiaceae | 0.02 |
| Upstream gene variant | *Akkermansia* | 43.4 |
| | Akkermansiaceae | 44.5 |
| | *Turicibacter* | 26.7 |

both associated with *TRIM15* and *TRIM26* located on BTA 23. Despite the absence of a QTL on BTA 23, this autosome contained the largest number of genes significantly associated with taxa abundances in the gene-based approach, in a region between 26.9 and 34.9 Mbp (Figure 2). No gene was found to be associated with the diversity indices (i.e., Shannon index and the first 2 axes of NMDS constructed on the Bray-Curtis dissimilarity matrix), which is consistent with the single marker-based association results.

Interestingly, the CI of QTL2 and 3 linked with *Akkermansia* and Akkermansiaceae abundances, respectively (Table 1), contained 3 candidate genes associated with both taxa abundances (Table 4): *ABO* ($q_{Akkermansia}$ = 0.0012; $q_{Akkermansiaceae}$ = 0.0003), *U6* ($q_{Akkermansia}$ = 0.0009; $q_{Akkermansiaceae}$ = 0.0002), and *DIPK1B* ($q_{Akkermansia}$ = 0.046; $q_{Akkermansiaceae}$ = 0.037). However, the gene-based approach did not identify significant candidate genes in the other QTL regions (Figure 2).

These 92 genes were then considered for the enrichment analyses, as explained in the Materials and Methods section. Among these, 39 genes were not recognized by the ShinyGo software because they were either non-protein-coding genes or uncharacterized genes. Among the 53 genes with functional annotations, we identified significant enrichment in 1 KEGG pathway and 63 GO terms.

The KEGG pathway type I diabetes mellitus (fold enrichment = 26.4; q = 0.017) was found significantly enriched, with 3 genes (out of 54) involved: *LTA*, *TNF*, and *BoLA/MHC-I*, all located on BTA 23. Another KEGG pathway, namely, the NF-κB signaling pathway, showed a tendency toward enrichment (fold enrichment = 13.6; q = 0.061), involving 3 genes (*LTA* and *TNF* from BTA 23, and *MALT1* from BTA 24) out of 105 genes in the pathway.

Regarding GO terms enrichment, 63 biological processes were significantly enriched considering the 53 candidate genes (Figure 3). The most enriched GO term was positive regulation of chronic inflammatory response (GO:0002678; fold enrichment = 237.6; q = 0.003), with 2 genes (*LTA* and *TNF*) involved out of the 4 genes in this biological process. The most significant GO term was defense response to other organisms (GO:0098542; fold enrichment = 7.6; q = 0.0008), followed by immune response (GO:0006955; fold enrichment = 5.3; q = 0.001) with 10 genes (out of 625) and 12 genes (out of 1,077) associated, respectively. For both terms, only *MALT1* was located on BTA 24, whereas the remaining associated genes were mapped on BTA 23. Additionally, significant enrichments were observed for 2 biological processes: the immune system process (GO:0002376) and response to stress (GO:0006950), both involving 13 genes. No molecular functions or cellular components were significantly enriched considering the list of genes.

## DISCUSSION

### Bovine Fecal Microbiota Associated With the Host Genome

This study represents a pioneering application of sequence-based GWAS to the fecal microbiota composition in dairy cattle. By leveraging imputed whole genome sequence data, we identified several bovine genomic regions associated with the abundance of specific bacterial taxa. Five QTL were significantly associated with 5 genera (*Akkermansia*, *Paeniclostridium*, an unclassified genus from the Paludibacteraceae family, *Sutterella*, and *Turicibacter*) commonly found in the bovine digestive

**Table 3.** Number of candidate genes significantly associated with the different heritable taxonomic levels and diversity indices[1]

| Type of microbiota-based traits | Number of traits analyzed | Number of candidate genes |
|---|---|---|
| ASV | 41 | 8 |
| Genus | 40 | 60 |
| Family | 26 | 63 |
| Diversity | 3 | 0 |
| Total | 110 | 90 |

[1]ASV = amplicon sequence variants.

ARTICLE IN PRESS—UNCORRECTED PROOF



**Table 4.** Gene-based analyses results: genes significantly associated with ASV, genus, or family abundances[1]

| Gene | BTA | Start | End | ASV | Genus | Family |
|---|---|---|---|---|---|---|
| *PMS1* | 2 | 6,450,096 | 6,602,554 | | Unknown Peptococcaceae | |
| *OTUD3* | 2 | 132,788,458 | 132,841,991 | *UCG-005*_g_ASV39 | | |
| *ENSBTAG00000060639* | 3 | 41,852,206 | 41,968,931 | *Turicibacter*_g_ASV17 | | |
| *ENSBTAG00000057475* | 3 | 114,525,242 | 114,601,748 | | Unknown Bacteroidales RF16 group | |
| *WC1* | 5 | 102,757,946 | 102,829,482 | | *Alloprevotella* | Erysipelotrichaceae |
| *RHNO1* | 5 | 106,866,984 | 106,896,915 | | | Rikenellaceae |
| *ENSBTAG00000062745* | 7 | 15,162,042 | 15,186,277 | | *Parasutterella* | |
| *ENSBTAG00000069017* | 7 | 45,576,861 | 45,609,017 | | Unknown Bacteroidales RF16 group | |
| *ENSBTAG00000062061* | 7 | 45,577,084 | 45,605,831 | | Unknown Bacteroidales RF16 group | |
| *PCDH1* | 7 | 52,924,572 | 52,968,313 | | | Peptococcaceae |
| *PCDH12* | 7 | 52,998,893 | 53,031,354 | | Unknown Peptococcaceae | Peptococcaceae |
| *RNF14* | 7 | 53,025,711 | 53,078,836 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000067538* | 8 | 6,587,218 | 6,613,100 | | *Parasutterella* | |
| *ENSBTAG00000056934* | 8 | 85,120,194 | 85,156,493 | | | Rikenellaceae |
| *ENSBTAG00000070117* | 10 | 23,698,382 | 23,772,513 | | Unknown Peptococcaceae | Peptococcaceae |
| *TRAV24* | 10 | 23,710,087 | 23,730,599 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000059019* | 10 | 23,715,903 | 23,736,475 | | Unknown Peptococcaceae | Peptococcaceae |
| *OR4K15* | 10 | 27,343,751 | 27,364,695 | | *Akkermansia* | |
| *ENSBTAG00000050207* | 10 | 27,356,790 | 27,377,764 | | *Akkermansia* | Akkermansiaceae |
| *SPMIP9/ TEX37* | 11 | 47,530,445 | 47,555,141 | | *Roseburia* | |
| *ENSBTAG00000047449* | 11 | 49,215,723 | 49,238,192 | Lachnospiraceae_f_ASV13 | | |
| *DPM2* | 11 | 98,590,692 | 98,613,745 | | | Erysipelatoclostridiaceae |
| *EEIG1* | 11 | 98,593,817 | 98,653,064 | | | Erysipelatoclostridiaceae |
| *DIPK1B* | 11 | 104,096,376 | 104,137,755 | | *Akkermansia* | Akkermansiaceae |
| *ABO* | 11 | 104,167,182 | 104,224,809 | | *Akkermansia* | Akkermansiaceae |
| *U6* | 11 | 104,171,329 | 104,191,435 | | *Akkermansia* | Akkermansiaceae |
| *PCSK2* | 13 | 37,496,073 | 37,754,161 | *Gastranaerophilales*_o_ASV80 | | |
| *ENSBTAG00000049520* | 13 | 37,742,435 | 37,766,353 | *Gastranaerophilales*_o_ASV80 | | |
| *NXT1* | 13 | 42,009,420 | 42,032,739 | | | Rikenellaceae |
| *GZF1* | 13 | 42,018,972 | 42,046,112 | | | Rikenellaceae |
| *NAPB* | 13 | 42,029,083 | 42,086,724 | | | Rikenellaceae |
| *ENSBTAG00000061029* | 13 | 42,067,022 | 42,094,723 | | | Rikenellaceae |
| *CSTL1* | 13 | 42,076,442 | 42,099,820 | | | Rikenellaceae |
| *U6* | 13 | 47,185,637 | 47,205,743 | | | Rikenellaceae |
| *SNX31* | 14 | 63,696,889 | 63,800,182 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000049445* | 17 | 29,575,880 | 29,596,182 | | | Prevotellaceae |
| *ENSBTAG00000060723* | 19 | 20,527,140 | 20,548,842 | | *Akkermansia* | Akkermansiaceae |
| *ENSBTAG00000063662* | 19 | 20,528,970 | 20,560,928 | | *Akkermansia* | Akkermansiaceae |
| *SNORA72* | 19 | 20,533,611 | 20,553,742 | | *Akkermansia* | |
| *TCAIM* | 22 | 16,261,889 | 16,319,668 | | Unknown Bacteroidales RF16 group | Bacteroidales RF16 group |
| *TSBP1* | 23 | 26,914,545 | 27,010,146 | | *Paeniclostridium* | |
| *CYP21A2* | 23 | 27,316,529 | 27,340,230 | | | Spirochaetaceae |
| *BAG6* | 23 | 27,635,741 | 27,667,411 | | | Peptococcaceae |
| *PRRC2A* | 23 | 27,647,894 | 27,681,044 | | Unknown Peptococcaceae | Peptococcaceae |
| *SNORA38* | 23 | 27,660,553 | 27,680,683 | | Unknown Peptococcaceae | Peptococcaceae |
| *NCR3* | 23 | 27,663,927 | 27,724,142 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000063748* | 23 | 27,669,810 | 27,713,235 | | Unknown Peptococcaceae | Peptococcaceae |
| *AIF1* | 23 | 27,675,449 | 27,698,367 | | Unknown Peptococcaceae | Peptococcaceae |
| *LST1* | 23 | 27,696,392 | 27,723,236 | | Unknown Peptococcaceae | Peptococcaceae |
| *TNF* | 23 | 27,706,180 | 27,729,078 | | Unknown Peptococcaceae | Peptococcaceae |
| *LTA* | 23 | 27,710,176 | 27,731,739 | | | Peptococcaceae |
| *ENSBTAG00000031913* | 23 | 27,775,529 | 27,807,772 | | *Paeniclostridium* | |

*Continued*

ARTICLE IN PRESS—UNCORRECTED PROOF



**Table 4 (Continued).** Gene-based analyses results: genes significantly associated with ASV, genus, or family abundances[1]

| Gene | BTA | Start | End | ASV | Genus | Family |
|---|---|---|---|---|---|---|
| *ENSBTAG00000005182* | 23 | 27,819,542 | 27,840,480 | | *Parasutterella* | |
| *ENSBTAG00000001476* | 23 | 27,820,086 | 27,909,645 | | *Parasutterella* | |
| *ENSBTAG00000054754* | 23 | 27,820,623 | 27,840,699 | | *Parasutterella* | |
| *NRM* | 23 | 28,309,981 | 28,333,011 | | | Spirochaetaceae |
| *C23H6orf136* | 23 | 28,339,054 | 28,363,973 | | | Peptococcaceae |
| *TRIM26* | 23 | 28,767,621 | 28,797,323 | | *Parasutterella* | Spirochaetaceae |
| *TRIM15* | 23 | 28,792,214 | 28,823,725 | | *Parasutterella* | Spirochaetaceae |
| *TRIM10* | 23 | 28,806,455 | 28,835,684 | | | Spirochaetaceae |
| *ENSBTAG00000069438* | 23 | 30,052,891 | 30,076,680 | | | Peptococcaceae |
| *ENSBTAG00000055332* | 23 | 30,055,214 | 30,076,467 | | Unknown Peptococcaceae | Peptococcaceae |
| *U6* | 23 | 30,059,139 | 30,079,235 | | Unknown Peptococcaceae | Peptococcaceae |
| *H1-6* | 23 | 31,831,419 | 31,855,499 | | Unknown Peptococcaceae | Peptococcaceae |
| *H4C3* | 23 | 31,836,920 | 31,857,571 | | Unknown Peptococcaceae | Peptococcaceae |
| *HFE* | 23 | 31,846,985 | 31,874,326 | | Unknown Peptococcaceae | Peptococcaceae |
| *H1-2* | 23 | 31,865,797 | 31,888,229 | | Unknown Peptococcaceae | Peptococcaceae |
| *H3C6* | 23 | 31,872,681 | 31,893,164 | | Unknown Peptococcaceae | Peptococcaceae |
| *H2BC3* | 23 | 31,874,096 | 31,895,726 | | Unknown Peptococcaceae | Peptococcaceae |
| *H2AC5P* | 23 | 31,874,099 | 31,894,488 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000058034* | 23 | 31,881,310 | 31,903,600 | | Unknown Peptococcaceae | Peptococcaceae |
| *H4C6* | 23 | 31,894,189 | 31,915,695 | | Unknown Peptococcaceae | Peptococcaceae |
| *H3C6* | 23 | 31,897,552 | 31,924,844 | | Unknown Peptococcaceae | Peptococcaceae |
| *H1-1* | 23 | 31,902,689 | 31,923,752 | | Unknown Peptococcaceae | Peptococcaceae |
| *ENSBTAG00000066000* | 23 | 31,913,093 | 31,937,486 | | Unknown Peptococcaceae | Peptococcaceae |
| *C23H6orf62* | 23 | 33,044,916 | 33,077,994 | | *Parasutterella* | |
| *U6* | 23 | 34,865,975 | 34,886,083 | | | Spirochaetaceae |
| *U6* | 23 | 34,892,642 | 34,912,750 | | | Spirochaetaceae |
| *MALT1* | 24 | 57,764,963 | 57,850,221 | | | Rikenellaceae |
| *ENSBTAG00000052092* | 26 | 42,380,257 | 42,502,598 | *Clostridium sensu stricto 1*_g_ASV82 | | |
| *bta-mir-10225a* | 26 | 42,420,932 | 42,440,990 | *Clostridium sensu stricto 1*_g_ASV82 | | |
| *bta-mir-10225a* | 26 | 42,427,008 | 42,447,066 | *Clostridium sensu stricto 1*_g_ASV82 | | |
| *ENSBTAG00000063476* | 27 | 6,155,629 | 6,176,953 | | Unknown Gastranaerophilales | |
| *ENSBTAG00000056541* | 27 | 6,162,625 | 6,185,783 | | Unknown Gastranaerophilales | |
| *ENSBTAG00000053110* | 27 | 6,173,690 | 6,198,322 | | Unknown Gastranaerophilales | |
| *ENSBTAG00000060533* | 27 | 42,003,959 | 42,052,183 | | *Alloprevotella* | Prevotellaceae |
| *ENSBTAG00000063225* | 27 | 42,034,872 | 42,056,204 | | *Alloprevotella* | Prevotellaceae |
| *RPL15* | 27 | 42,041,425 | 42,065,369 | | *Alloprevotella* | Prevotellaceae |
| *NKIRAS1* | 27 | 42,045,586 | 42,085,734 | | *Alloprevotella* | Prevotellaceae |
| *ENSBTAG00000059958* | 28 | 23,355,281 | 23,375,319 | | Unknown Gastranaerophilales | |

[1]ASV = amplicon sequence variant.

ARTICLE IN PRESS—UNCORRECTED PROOF



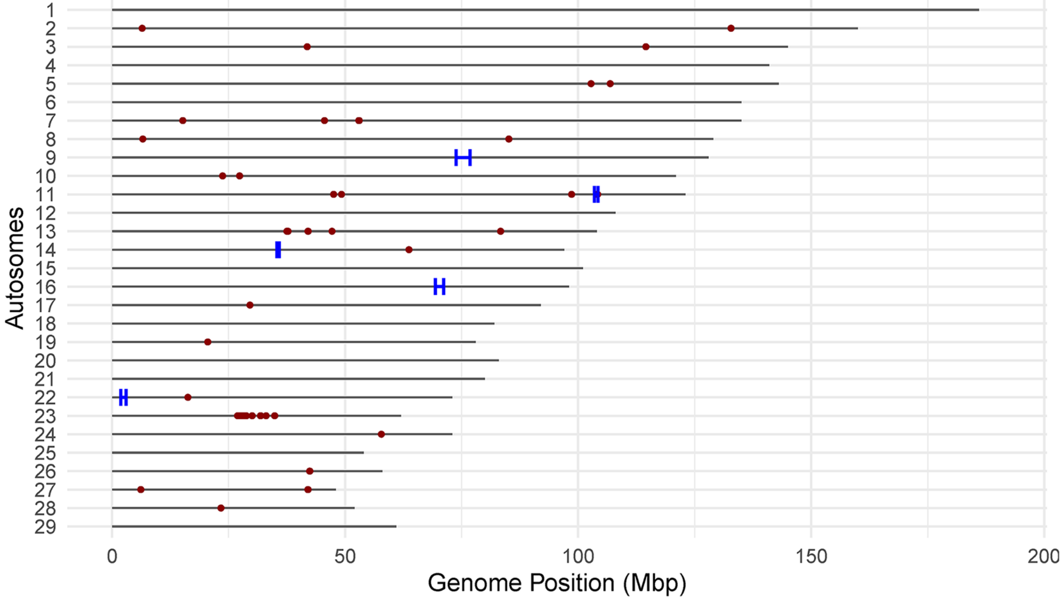


**Figure 2.** Regions of the bovine genome being significantly associated with the abundance of at least 1 taxon. Blue segments represent CI of the QTL detected by GWAS, and red dots indicate the position of genes identified in the gene-based approach.

tract, with $h^2$ estimates ranging from 0.08 to 0.21 (Brulin et al., 2024a). Another QTL, with a CI similar to the QTL linked to Akkermansia variability, was observed for its associated family, Akkermansiaceae ($h^2$ = 0.22). This result was expected, as *Akkermansia* is the sole representative genus of the Akkermansiaceae family. However, the higher significance and effect at the family level, compared with the genus level, reinforce the findings and highlight the advantages of working at higher taxonomic levels. This approach can leverage the effect of a reduced number of taxonomic groups on the CLR transformation. Nonetheless, working at higher taxonomic levels imposes limitations in microbiota analyses. The vast diversity of taxa within a family or phylum complicates the mapping of host genes and makes interpretation challenging. This complexity was illustrated in our study, where only 1 family (represented by a single genus) was associated with a genome region.

In young beef cattle, variants associated with the fecal abundance of *Akkermansia* have also been identified by Fan et al. (2020) but in other genomic regions dedicated to genes encoding mucin. Similarly, in another study on beef calves, Fan et al. (2021) identified 19 significant SNPs (out of ~90k variants) that differed from our findings, in terms of QTL positions and associated taxa (*Roseburia*, *Oscillospira*, *Ruminoclostridium* and *Mailhella* genera, Peptococcaceae family and Spirochaetes phylum). This discrepancy may be attributed to breed differences, as beef and dairy cattle have distinct microbial communities (Bainbridge et al., 2016; Noel et al., 2019), and the low detection power of GWAS when applied to a limited population size. Additionally, as shown by Fan et al. (2020, 2021), *Akkermansia* and *Sutterella* are both breed-dependent taxa, with fecal composition fluctuating with the breed composition. The breed effect could be one factor explaining the differences between beef and dairy cattle, along with other external factors, such as animal age (i.e., preweaning animals vs. lactating cows) and nutritional and management practices.

In comparison, more studies have investigated the host genome regions that may influence rumen microbiota composition. In Holstein cows, Golder et al. (2023) identified significant variants associated with 5 phyla and 14 (out of 29) families, many of which presented multiple QTL. In a more recent study, Martinez-Boggio et al. (2024) observed a larger number of associations between genomic variants and 33 ruminal ASV. In the rumen of beef cattle, Li et al. (2019a) observed associations between host genomic regions and various bacterial taxa, including 6 genera, 3 families, 1 order, 2 classes, and 2 phyla, totaling 19 significant variants. Abbas et al. (2020) identified 8 QTL associated with multiple taxa at the OTU, family, and phylum levels. However, among all these studies on the rumen microbiota, only Martinez-Boggio et al. (2024) observed genomic variants near 1 of our QTL. Specifically, the 2 variants (BovineHD1100027712 and ARS-BFGL-NGS-34624) in the vicinity of our QTL on BTA 11 (<3 Mbp) were associated with the abundances of *Akkermansia* and Akkermansiaceae. None of the genomic regions identified in Abbas

ARTICLE IN PRESS—UNCORRECTED PROOF

**Figure 3.** Fold enrichment values of gene ontology (GO) terms significantly enriched among the genes significantly associated with taxa abundances. The color of the bars represents the $-\log_{10}(P)$ adjusted for the Benjamini-Hochberg multiple testing correction (false discovery rate [FDR] = 5%). The size of the circles reflects the number of genes associated with each GO term, with larger circles indicating a greater number of associated genes.

et al. (2020), Golder et al. (2023), or Li et al. (2019b) align with our results on Holstein feces or the observations of Fan et al. (2021) on young beef cattle feces. This discrepancy may be due to insufficient detection power or biases related to microbiota data processing (i.e., laboratory protocols, bioinformatics procedures, and broadly, the 16S rRNA gene sequencing that targets only a segment of the gene). However, this limited overlap may also indicate that whereas both fecal (Brulin et al., 2024a) and rumen microbiotas (Li et al., 2019; Abbas et al., 2020) are driven by host genetics, they are governed by different host genomic regions. This suggests that the genetic mechanisms underlying fecal and rumen microbiota composition could differ. Additionally, the close genomic regions between Martinez-Boggio et al. (2024) and our study indicate that some genetic regions might affect both communities in different ways. These findings reflect the significant compositional differences observed between the ruminal and fecal microbiotas in dairy cows (Monteiro et al., 2022). Nevertheless, the variability in genomic regions observed underscores the complex interactions between bacterial communities and the host genome, driven by a multitude of genetic mechanisms, making microbiota composition a complex polygenic trait.

In the rumen microbiota, Martinez-Boggio et al. (2024) colocalized genomic regions associated with taxa abundance, with regions influencing the feed efficiency performance of Holstein cows. Given that genetic correlations between host performances and taxa abundances

ARTICLE IN PRESS—UNCORRECTED PROOF



were estimated in Brulin et al. (2024a), we conducted a comparative analysis of the genes and QTL positions identified in our study with those reported by Tribout et al. (2020), who used sequence-based GWAS to pinpoint QTL and significant variants for traits under selection in various French dairy breeds, including Holstein. Interestingly, the most significant variant (rs207607053) influencing fat content on BTA 11 in Holstein cattle was near the QTL associated with *Akkermansia* abundance in feces. This finding is notable, given that *Akkermansia* is a commensal genus from the digestive tract, which is often used as a probiotic. Various studies, such as Wolter et al. (2024), have demonstrated *Akkermansia*'s protective effect against pathogens in high-fiber diets.

Although only 1 QTL colocalized with a genomic region influencing a trait under selection, the QTL identified in this study were associated with taxa relevant to host health. For instance, the *Paeniclostridium* genus, identified as one of the most heritable genera in Holstein feces (Brulin et al., 2024a), includes pathogenic species involved in various infections in cattle (Wang et al., 2022; Jiang et al., 2023a). *Sutterella* has been described as a proinflammatory genus that triggers the host's immune response (Jiang et al., 2023a). Conversely, the Paludibacteraceae family, for which we could not identify a specific genus, tends to be enriched in healthy animals (Gu et al., 2023b) and can metabolize lignocellulose, enabling the host to access more VFA. *Turicibacter*, associated with the lowest $h^2$ ($h^2 = 0.08$), can modify host-produced bile acids and lipid metabolism, potentially aiding diet valorization (Lynch et al., 2023). Therefore, even if some identified genomic regions did not colocalize with those of traits under selection, they could influence nonselected phenotypes, such as immunity-linked traits.

However, some of the identified QTL, such as the one linked to *Sutterella* abundance, must be carefully considered, as they gather a limited number of variants with low MAF. Despite testing a broad range of phenotypes, including 113 taxa abundances and 3 diversity metrics, the number of QTL is relatively modest. This finding is consistent with previous studies conducted in humans (Davenport et al., 2015; Hughes et al., 2020; Kurilshikov et al., 2021), pigs (Crespo-Piazuelo et al., 2019), and ruminant studies focusing on the rumen (Li et al., 2019; Martinez Boggio et al., 2022) and calf feces (Fan et al., 2021). This limited number of QTL mirrors the complex interactions between the host and its microbiota, primarily influenced by environmental factors (Rothschild et al., 2018). Although our study benefited from a large sample size for a microbiota analysis, it remains relatively modest for GWAS. The use of WGS increased the detection power (Daetwyler et al., 2014), but genomic regions with smaller effects may not have been identified due to the study population size.

We did not identify QTL linked to diversity indices (i.e., Shannon index [α-diversity], and NMDS1 and NMDS2 [β-diversity]). Although our results are consistent with those of Li et al. (2019) in the rumen of beef cattle, we could have expected certain genomic regions to promote either heterogeneous colonization of the intestines or the dominance of specific taxa. The absence of variants with significant effects on diversity indices could be attributed to a large number of genes with small effects, undetectable due to the relatively small sample size. It could also suggest that α- or β-diversity metrics may not be directly influenced by the host genome, despite the estimation of a significant $h^2$ for β-diversity by Brulin et al. (2024).

### Candidate Genes Associated With Fecal Microbiota

To increase detection power and mitigate sample size limitations, we supplemented our GWAS with gene-based association studies. This approach identified 92 candidate genes across the genome associated with 22 taxa, including 5 ASV, 8 genera, and 8 families. However, considering upper taxonomic ranks, such as family, can be controversial due to the diversity of taxa they encompass, as previously discussed.

Among studies investigating genomic associations with gastrointestinal microbiota, only Martinez-Boggio et al. (2024) identified 10 variants near our candidate genes (<3 Mbp). Interestingly, 2 different regions, one at the end of BTA 11 and another on BTA 19, were associated with the abundance of certain *Prevotella* ASV in the rumen in their study, as well as with the abundance of *Akkermansia* and Akkermansiaceae in the feces in our study. Several other overlapping genomic regions were found on BTA 3 (associated with a ruminal *CAG-352* ASV and an unclassified fecal genus from *Bacteroidales RF16 group*), on BTA 11 (linked to a ruminal *F082* ASV and the fecal *Roseburia* abundance), on BTA 13 (2 regions associated with a ruminal *Acetitomaculum* ASV and a fecal ASV from *Gastranaerophilales* and the Rikenellaceae family, respectively), on BTA 14 (associated with both the ruminal abundance of a Rikenellaceae *RC9 gut group* ASV and the fecal abundances of the Peptococcaceae family and its unclassified genus), and on BTA 28 (influencing a ruminal *p-251-o5* ASV and the fecal abundance of an unclassified genus from the *Gastranaerophilales* order). It is relevant to note that working at the gene level enabled us to detect more genomic regions than in a standard GWAS study, which facilitated the identification of a greater number of regions shared with other studies. However, the number of common genes and regions remained limited compared with the number of microbiota phenotypes analyzed, reinforcing the hy-

ARTICLE IN PRESS—UNCORRECTED PROOF



pothesis that distinct genomic mechanisms underlie the composition of fecal and ruminal microbiotas.

Our gene-based association studies highlighted 41 genes associated with multiple taxa. Many of these genes were linked to 2 taxa from the same taxonomic group, exemplified by Peptococcaceae ($h^2$ = 0.09; RA = 0.73%) and its unclassified genus ($h^2$ = 0.09; RA = 0.73%) on BTA 7, 10, 14, and 23, or *Alloprevotella* ($h^2$ = 0.06; RA = 1.50%) and Prevotellaceae ($h^2$ = 0.06; RA = 7.67%) on BTA 27. Conversely, other genes, such as *WC1*, *TRIM26*, and *TRIM15*, were linked to phylogenetically distant taxa. For instance, the *WC1* gene on BTA 5 was associated with both *Alloprevotella* and Erysipelotrichaceae ($h^2$ = 0.07; RA = 0.74%). Notably, the *WC1* gene influences the expression of bovine γδ T cells involved in immunity and protection against pathogens such as *Mycobacterium bovis* (Bhat et al., 2023). These interactions were interesting as *Alloprevotella* is a beneficial bacterium enriched in the gut of healthy calves (Chen et al., 2022; Wu et al., 2024) and promotes host health and fermentation efficiency in goats (Chen et al., 2024a). Conversely, Erysipelotrichaceae, which contains pathogenic species, is negatively associated with residual feed intake in cattle (McGovern et al., 2020) and linked with inflammation and disorders in the human gut (Kaakoush, 2015; Nagao-Kitamoto et al., 2016).

The gene-based association analyses performed with MAGMA identified a specific genomic region on BTA 23, spanning from 26.9 to 34.9 Mbp, which harbors 38 genes. This region was primarily associated with the Peptococcaceae family and 1 affiliated unclassified genus (25 and 21 genes, respectively), as well as the Spirochaetaceae family ($h^2$ = 0.08; RA = 0.91%), and the genera *Paeniclostridium* and *Parasutterella* ($h^2$ = 0.21; RA = 0.65%). However, GWAS failed to detect significantly associated variants in this region (the most significant variants presented a $-\log_{10}(P)$ of 7.02 for Peptococcaceae, 6.48 for the unclassified Peptococcaceae, 6.28 for Spirochaetaceae, 6.17 for *Parasutterella,* and 5.64 for *Paeniclostridium*). Interestingly, many genes significantly associated with these taxa are part of the bovine MHC, a genomic region known for its involvement in the immune response. The MHC encodes cell surface glycoproteins that present antigens from pathogens to T-lymphocytes, initiating the immune response (Sommer, 2005). Numerous studies have demonstrated the effect of MHC polymorphisms on microbiota composition across various species (Bolnick et al., 2014; Kubinak et al., 2015; Khan et al., 2019). In cattle, Derakhshani et al. (2018) reported an association between the bovine MHC and the microbiota of colostrum and milk during the first week of lactation. Our findings are further supported by gene-set enrichment analyses that showed significant enrichment of genes involved in immunity-linked GO terms. The most enriched term was defense response to other organisms, totaling 10 genes (on BTA 11, 23, 24, and 27) out of 625 genes belonging to this GO term. Additionally, 13 genes out of the 53 mapped genes were associated with the immune response process and response to stress. Many of the genes involved in the host's immune response were located on BTA 23, with a high concentration in the MHC region, which poses a risk of false positives due to the proximity of these genes. However, our results are reinforced by the identification of immunity-linked genes out of BTA 23, such as *MALT1* on BTA 24. Notably, the *MALT1* gene plays an essential role in the activation of the *NF-κB* regulation pathway, which was found to be significantly enriched in our KEGG pathway enrichment analysis and is known to interact with the bacterial community (Karrasch and Jobin, 2008; Johannessen et al., 2013). Certain genes have also been associated with specific gut disorders in ruminants, including *PCDH1* (Veshkini et al., 2024), whose expression is induced by *Cryptosporidium parvum* in calves. Other genes are known to be associated with gut disorders in other species, such as *PCSK2*, which is a type 2 diabetes susceptibility gene (Chang et al., 2015), and *DMP2* and *CLST1*, whose expression is linked to the occurrence of colorectal cancers (Wei et al., 2021; Zou et al., 2022). It is important to note that our enrichment analyses were limited by incomplete knowledge about some genes and their functions, as 39 genes (42%) were not successfully recognized by ShinyGO. Among these, aside from some uncharacterized genes, we identified several long noncoding genes that might influence the host's response to pathogens through interactions with the *NF-κB* pathway (Gupta et al., 2019).

Our findings suggest that the genetic makeup under the host's immune response may shape the composition of the fecal microbiota of dairy cows. Host genetics might influence gut colonization by different bacterial species, either by promoting or limiting their growth or by facilitating the provision of nutrients to certain taxa (Tomkovich and Jobin, 2016). These results align with the observation of Fan et al. (2021), who also reported enrichment of genes involved in the immune response in the fecal microbiota composition of young suckling cattle.

Additionally, as mentioned previously, we identified associations between immunity-related genes and microbial taxa, as well as taxa known to influence host health status. Apart from *Alloprevotella*, some taxa may have specific benefits for cattle, such as *Roseburia*. Indeed, this butyrate-producing genus is often used as a probiotic due to its roles in effecting colonic motility, effects on immunity maintenance, and antiinflammatory properties (Tamanai-Shacoori et al., 2017). *Parasutterella* is another potentially beneficial genus identified as the

ARTICLE IN PRESS—UNCORRECTED PROOF



most heritable taxon in Brulin et al. (2024a). Conversely, some of the identified taxa can be considered as potentially pathogenic. For instance, the *Alloprevotella* genus and the Erysipelatoclostridiaceae family have been associated with ileal inflammation and oxidative stress in Holstein cattle (Gu et al., 2023a; Jiang et al., 2023b), whereas *Clostridium sensu stricto 1* is enriched in diarrheic calves (Tang et al., 2023). Beyond health-related associations, we did not identify genes influencing traits under selection, such as production or fertility. However, Tribout et al. (2024) highlighted variants on BTA 14 influencing milk yield, fat content, and protein content of Holstein cows, located near the *SNX31* gene. Interestingly, in our study, this same gene was linked to the abundance of the Peptococcaceae family and its unclassified genus, suggesting a potential association between fecal microbiota composition and milk production traits.

Nevertheless, none of the genes located within the CI of QTL1, 4, 5, and 6 presented a significant *P*-value in the gene-based association study. Surprisingly, the most significant variant of QTL4 (associated with the unclassified genus from the Paludibacteraceae family) is located in an intron of the long noncoding *ENSBTAG00000060358* gene on BTA 14. This gene is associated with *KCNB2*, which mediates transmembrane potassium transport in excitable membranes, such as smooth muscle cells. Similarly, the most significant variant in the QTL1 CI is located in an intergenic region surrounded by 8 RNA genes and 3 protein-coding genes, including *TNFAIP3*, *PERP*, and *ARFGEF3*. Interestingly, *TNFAIP3* is involved in the *NF-κB2* pathway, which is associated with immune response. In QTL5, the most significant variant was close to various noncoding RNA genes and the *PROX1* gene, whereas in QTL6, the variant was located between 2 RNA genes and 2 protein-coding genes, *CMC1* and *ENSBTAG00000060599*. Whereas there is still a lack of functional annotation and characterization of *ENSBTAG00000060599*, *CMC1* has been associated with the host's immune response in a mouse model, acting as a positive regulator of CD8+ T cells (Chen et al., 2024b), which is negatively correlated with the abundance of *Turicibacter* (Labarta-Bajo et al., 2020; Singh et al., 2023).

The integration of single-marker– and gene-based associations provided a finer view of the genomic regions influencing the composition of the fecal microbiota. Sometimes, they reached a consensus, such as at the end of BTA 11, where 3 genes (*ABO*, *U6*, and *DIPK1B*) were located within the 2 QTL associated with *Akkermansia* and its associated family, Akkermansiaceae. As mentioned previously, this genus is a beneficial gut commensal (Wolter et al., 2024) that presented a moderate $h^2$ in Holstein cows (Brulin et al., 2024a). Whereas previous studies have associated *Akkermansia* with the *TGIF1* gene in pigs (Crespo-Piazuelo et al., 2019) and *PLD2* in humans (Davenport et al., 2015), it has not been linked to the candidate genes we found in our study (i.e., *ABO*, *U6*, and *DIPK1B*). Whereas *DIPK1B* is a poorly characterized gene and *U6* a spliceosomal RNA, the *ABO* gene stands out as the most promising candidate. Despite being less studied in cattle, *ABO* is involved in blood group determination (Turcot-Dubois et al., 2007), and is mainly expressed in the gastrointestinal tract (Liu et al., 2022). In this study, the most significant variant was located downstream of *ABO,* and various mutations with a potential influence on the *ABO* protein function were identified: 2 stop-gained variants and 1 missense variant. The lack of knowledge about the bovine *ABO* gene and its protein limits our understanding of the mechanisms linking this gene and *Akkermansia* abundance in the bovine feces. However, in pigs and humans, *ABO* gene variation has been found to effect the composition of the fecal microbiota. Specifically, a deletion in *ABO* affected the abundance of Erysipelotrichaceae in pigs (Yang et al., 2022). In humans, the effect of the *ABO* gene on the digestive microbiota remains unclear, but its expression has been shown to effect various taxa abundances (Lopera-Maya et al., 2022; Qin et al., 2022).

### *Limitations and Perspectives of the Study*

In this study, we examined bacterial abundances across various taxonomic levels, from ASV to family. However, analyzing higher taxonomic ranks imposes limitations in microbiota analyses. The vast diversity of taxa within a family or phylum complicates the mapping of host genes and makes interpretation challenging. This complexity was illustrated in our GWAS study, where only 1 family—represented by a single genus—was associated with a genome region. However, using higher taxonomic ranks, such as genus or family levels, allowed us to form larger groups with shared characteristics. This strategy is primarily employed in 16S rRNA gene sequencing (ref). Currently, the 16S rRNA gene sequencing provides a comprehensive overview of bacterial community composition on a large scale. However, methods such as WGS or long-read sequencing would allow for finer characterization of the microbial community, once they become more feasible for large populations (i.e., considering sequencing cost or data storage).

Therefore, additional analyses are needed to validate these associations and elucidate the precise underlying biological processes and mechanisms. However, these findings provide valuable insights into the complex interplay between host genetics and fecal microbiota composition in dairy cattle. Notably, aside from sample size, using a large commercial population associated with a

ARTICLE IN PRESS—UNCORRECTED PROOF



diversity of farms helps bypass the strong genotype-by-environment effect on microbiota composition (Edwards et al., 2023). This also demonstrates the feasibility of conducting genetic analyses on fecal microbiota at a national level, without dedicated experimental farms or highly detailed information on animals.

## CONCLUSIONS

The use of sequence-based GWAS on fecal microbiota data from a relatively large population of commercial Holstein cows highlighted the complex links between the microbiota composition and the host genome. Despite the potentially highly polygenic nature of the microbiota, we successfully identified various regions of the Holstein cow genome associated with specific microbial taxa in a commercial population, even with limited information on cattle management and feeding conditions. This illustrates the benefits of conducting microbiota studies in production settings, enabling larger-scale investigations compared with experimental designs. Additionally, we unveiled a putative association between the genetic makeup underlying the host's immune response and the fecal microbiota composition and identified potential functional candidate genes along the cow genome, such as *ABO*.

## NOTES

This work has been financially supported by the French "Association Nationale de la Recherche et de la Technologie" (ANRT; Paris, France) that allows the PhD contract of Louise Brulin (CIFRE n°2021/0713). The authors thank both the breeders and the technicians who were involved in the collection of the samples, as well as Gérard Heumez (Gènes Diffusion, Douai, France) for his assistance in the data processing. The authors also thank Mekki Boussaha (Université Paris-Saclay, AgroParisTech, INRAE, GABI, Jouy-en-Josas, France) for his help and support in the imputation steps and the processing of the 1000 Bull Genomes Project sequences. The authors gratefully acknowledge the French National Database (CTIG; Jouy-en-Josas, France) and Valogene (Paris, France) for providing data; and the contribution of the 1000 Bull Genomes Project. The authors additionally thank the INRAE MIGALE Bioinformatics Facility (MIGALE, INRAE, 2020) and Migale Bioinformatics Facility (Jouy-en-Josas, France, doi: 10.15454/1.5572390655343293E12) for providing the computational resources needed for our analyses. GD Biotech expresses gratitude to the ANR-10-EQPX-07-01 LIGAN-PM-UMR8199 (Institut Pasteur de Lille, Lille, France) for its technical support regarding the MiSeq sequencing. The data used in this study are available upon request from the corresponding author. Supplemental material for this article is available at [URL]. This study was conducted on dairy cattle reared for commercial purposes in compliance with the French regulation (Code Rural et de la Pêche Maritime) and the European Council Directive 98/58/EC. The sampling of the animals was part of the routine animal manipulations carried out by duly authorized technicians from the Gènes Diffusion breeding company (Douai, France). Thus, according to the French legislation, no approval from the Institutional Animal Care and Use Committee or ethics committee was necessary. Additionally, the farmers involved in this study agreed to the use of their animals' samples for research purposes. LB, SD, GE, SMa, SMe, and CA are employed by GD Biotech/Gènes Diffusion company. The remaining authors have not stated any conflicts of interest.

**Nonstandard abbreviations used:** ASV = amplicon sequence variants; b = additive effect; CLR = centered-log ratio; FDR = false discovery rate; GO = gene ontology; HD = high-density; KEGG = Kyoto Encyclopedia of Genes and Genomes; MAF = minor allele frequency; NMDS = nonmetric multidimensional scaling; RA = relative abundance; WGS = whole genome sequencing.

## REFERENCES

Abbas, W., J. T. Howard, H. A. Paz, K. E. Hales, J. E. Wells, L. A. Kuehn, G. E. Erickson, M. L. Spangler, and S. C. Fernando. 2020. Influence of host genetics in shaping the rumen bacterial community in beef cattle. Sci. Rep. 10:15101. https://doi.org/10.1038/s41598-020-72011-9.

Bainbridge, M. L., L. M. Cersosimo, A.-D. G. Wright, and J. Kraft. 2016. Rumen bacterial communities shift across a lactation in Holstein, Jersey and Holstein × Jersey dairy cows and correlate to rumen function, bacterial fatty acid composition and production parameters. FEMS Microbiol. Ecol. 92:fiw059. https://doi.org/10.1093/femsec/fiw059.

Bhat, S. A., M. Elnaggar, T. J. Hall, G. P. McHugo, C. Reid, D. E. MacHugh, and K. G. Meade. 2023. Preferential differential gene expression within the WC1.1+ γδ T cell compartment in cattle naturally infected with *Mycobacterium bovis*. Front. Immunol. 14:1265038. https://doi.org/10.3389/fimmu.2023.1265038.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bolnick, D. I., L. K. Snowberg, J. G. Caporaso, C. Lauber, R. Knight, and W. E. Stutz. 2014. Major histocompatibility complex class IIb polymorphism influences gut microbiota composition and diversity. Mol. Ecol. 23:4831–4845. https://doi.org/10.1111/mec.12846.

Boussaha, M., P. Michot, R. Letaief, C. Hozé, S. Fritz, C. Grohs, D. Esquerré, A. Duchesne, R. Philippe, V. Blanquet, F. Phocas, S. Floriot, D. Rocha, C. Klopp, A. Capitan, and D. Boichard. 2016. Construction of a large collection of small genome variations in French dairy and beef breeds using whole-genome sequences. Genet. Sel. Evol. 48:87. https://doi.org/10.1186/s12711-016-0268-z.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, M. Dolezal, H. Pausch, R. F. Brøndum, P. J. Bowman, B. Thomsen, B. Guldbrandtsen, M. S. Lund, B. Servin, D. J. Garrick, J. Reecy, J. Vilkki, A. Bagnato, M. Wang, J. L. Hoff, R. D. Schna-

bel, J. F. Taylor, A. A. E. Vinkhuyzen, F. Panitz, C. Bendixen, L.-E. Holm, B. Gredler, C. Hozé, M. Boussaha, M.-P. Sanchez, D. Rocha, A. Capitan, T. Tribout, A. Barbat, P. Croiseau, C. Drögemüller, V. Jagannathan, C. Vander Jagt, J. J. Crowley, A. Bieber, D. C. Purfield, D. P. Berry, R. Emmerling, K.-U. Götz, M. Frischknecht, I. Russ, J. Sölkner, C. P. Van Tassell, R. Fries, P. Stothard, R. F. Veerkamp, D. Boichard, M. E. Goddard, and B. J. Hayes. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50:362–367. https://doi.org/10.1038/s41588-018-0056-5.

Broad Institute. 2019. Picard Toolkit.

Browning, B. L., X. Tian, Y. Zhou, and S. R. Browning. 2021. Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. 108:1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Brulin, L., S. Ducrocq, J. Estellé, G. Even, S. Martel, S. Merlin, C. Audebert, P. Croiseau, and M.-P. Sanchez. 2024a. The fecal microbiota of Holstein cows is heritable and genetically correlated to dairy performances. J. Dairy Sci. 107:11254–11268. https://doi.org/10.3168/jds.2024-25003.

Brulin, L., S. Ducrocq, G. Even, M. P. Sanchez, S. Martel, S. Merlin, C. Audebert, P. Croiseau, and J. Estellé. 2024b. Short communication: Bifidobacterium abundance in the faecal microbiota is strongly associated with milk traits in dairy cattle. Animal 18:101243 https://doi.org/10.1016/j.animal.2024.101243.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

Cendron, F., G. Niero, G. Carlino, M. Penasa, and M. Cassandro. 2020. Characterizing the fecal bacteria and archaea community of heifers and lactating cows through 16S rRNA next-generation sequencing. J. Appl. Genet. 61:593–605. https://doi.org/10.1007/s13353-020-00575-3.

Chang, T.-J., Y.-F. Chiu, W. H.-H. Sheu, K.-C. Shih, C.-M. Hwu, T. Quertermous, Y.-S. Jou, S.-S. Kuo, Y.-C. Chang, and L.-M. Chuang. 2015. Genetic polymorphisms of PCSK2 are associated with glucose homeostasis and progression to type 2 diabetes in a Chinese population. Sci. Rep. 5:14380. https://doi.org/10.1038/srep14380.

Chen, J., X. Zhang, X. Chang, B. Wei, Y. Fang, S. Song, D. Gong, D. Huang, Y. Sun, X. Dong, Y. Zhao, and Z. Zhao. 2024a. Multi-omics analysis reveals the effects of host-rumen microbiota interactions on growth performance in a goat model. Front. Microbiol. 15:1445223. https://doi.org/10.3389/fmicb.2024.1445223.

Chen, Y., J. Gao, M. Ma, K. Wang, F. Liu, F. Yang, X. Zou, Z. Cheng, and D. Wu. 2024b. The potential role of CMC1 as an immunometabolic checkpoint in T cell immunity. OncoImmunology 13:2344905. https://doi.org/10.1080/2162402X.2024.2344905.

Crespo-Piazuelo, D., L. Migura-Garcia, J. Estellé, L. Criado-Mesas, M. Revilla, A. Castelló, M. Muñoz, J. M. García-Casco, A. I. Fernández, M. Ballester, and J. M. Folch. 2019. Association between the pig genome and its gut microbiota composition. Sci. Rep. 9:8791. https://doi.org/10.1038/s41598-019-45066-6.

Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsegge, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46:858–865. https://doi.org/10.1038/ng.3034.

Davenport, E. R., D. A. Cusanovich, K. Michelini, L. B. Barreiro, C. Ober, and Y. Gilad. 2015. Genome-wide association studies of the human gut microbiota. PLoS One 10:e0140301. https://doi.org/10.1371/journal.pone.0140301.

de Leeuw, C. A., J. M. Mooij, T. Heskes, and D. Posthuma. 2015. MAGMA: Generalized gene-set analysis of GWAS data. PLOS Comput. Biol. 11:e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

Derakhshani, H., J. C. Plaizier, J. De Buck, H. W. Barkema, and E. Khafipour. 2018. Association of bovine major histocompatibility complex (BoLA) gene polymorphism with colostrum and milk microbiota of dairy cows during the first week of lactation. Microbiome 6:203. https://doi.org/10.1186/s40168-018-0586-1.

Doekes, H. P., R. F. Veerkamp, P. Bijma, S. J. Hiemstra, and J. J. Windig. 2018. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. Genet. Sel. Evol. 50:15. https://doi.org/10.1186/s12711-018-0385-y.

Edwards, J. A., U. B. Saran, J. Bonnette, A. MacQueen, J. Yin, T. Nguyen, J. Schmutz, J. Grimwood, L. A. Pennacchio, C. Daum, T. Glavina del Rio, F. B. Fritschi, D. B. Lowry, and T. E. Juenger. 2023. Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range. Curr. Biol. 33:1926–1938.e6. https://doi.org/10.1016/j.cub.2023.03.078.

Fan, P., B. Bian, L. Teng, C. D. Nelson, J. Driver, M. A. Elzo, and K. C. Jeong. 2020. Host genetic effects upon the early gut microbiota in a bovine model with graduated spectrum of genetic variation. ISME J. 14:302–317. https://doi.org/10.1038/s41396-019-0529-2.

Fan, P., C. D. Nelson, J. D. Driver, M. A. Elzo, F. Peñagaricano, and K. C. Jeong. 2021. Host genetics exerts lifelong effects upon hindgut microbiota and its association with bovine growth and immunity. ISME J. 15:2306–2321. https://doi.org/10.1038/s41396-021-00925-x.

Gautason, E., A. A. Schönherz, G. Sahana, and B. Guldbrandtsen. 2021. Genomic inbreeding and selection signatures in the local dairy breed Icelandic cattle. Anim. Genet. 52:251–262. https://doi.org/10.1111/age.13058.

Golder, H. M., J. Thomson, J. Rehberger, A. H. Smith, E. Block, and I. J. Lean. 2023. Associations among the genome, rumen metabolome, ruminal bacteria, and milk production in early-lactation Holsteins. J. Dairy Sci. 106:3176–3191. https://doi.org/10.3168/jds.2022-22573.

Gu, F., S. Zhu, J. Hou, Y. Tang, J.-X. Liu, Q. Xu, and H.-Z. Sun. 2023a. The hindgut microbiome contributes to host oxidative stress in postpartum dairy cows by affecting glutathione synthesis process. Microbiome 11:87. https://doi.org/10.1186/s40168-023-01535-9.

Gu, F., S. Zhu, Y. Tang, X. Liu, M. Jia, N. Malmuthuge, T. G. Valencak, J. W. McFadden, J.-X. Liu, and H.-Z. Sun. 2023b. Gut microbiome is linked to functions of peripheral immune cells in transition cows during excessive lipolysis. Microbiome 11:40. https://doi.org/10.1186/s40168-023-01492-3.

Gupta, P., S. Peter, M. Jung, A. Lewin, G. Hemmrich-Stanisak, A. Franke, M. von Kleist, C. Schütte, R. Einspanier, S. Sharbati, and J. Bruegge. 2019. Analysis of long non-coding RNA and mRNA expression in bovine macrophages brings up novel aspects of *Mycobacterium avium* subspecies paratuberculosis infections. Sci. Rep. 9:1571. https://doi.org/10.1038/s41598-018-38141-x.

Harrison, P. W., M. R. Amode, O. Austine-Orimoloye, A. G. Azov, M. Barba, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S. K. Bhurji, S. Boddu, P. R. Branco Lins, L. Brooks, S. B. Ramaraju, L. I. Campbell, M. C. Martinez, M. Charkhchi, K. Chougule, A. Cockburn, C. Davidson, N. H. De Silva, K. Dodiya, S. Donaldson, B. El Houdaigui, T. E. Naboulsi, R. Fatima, C. G. Giron, T. Genez, D. Grigoriadis, G. S. Ghattaoraya, J. G. Martinez, T. A. Gurbich, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Lodha, D. Marques-Coelho, G. Maslen, G. A. Merino, L. P. Mirabueno, A. Mushtaq, S. N. Hossain, D. N. Ogeh, M. P. Sakthivel, A. Parker, M. Perry, I. Piližota, D. Poppleton, I. Prosovetskaia, S. Raj, J. G. Pérez-Silva, A. I. A. Salam, S. Saraf, N. Saraiva-Agostinho, D. Sheppard, S. Sinha, B. Sipos, V. Sitnik, W. Stark, E. Steed, M.-M. Suner, L. Surapaneni, K. Sutinen, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, D. Ware, E. Wass, N. L. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakiachvili, B. Flint, S. Giorgetti, L. Haggerty, G. R. Ilsley, J. Keatley, J. E. Loveland, B. Moore, J. M. Mudge, G. Naamati, J. Tate, S. J. Trevanion, A. Winterbottom, A. Frankish, S. E. Hunt, F. Cunningham, S. Dyer, R. D. Finn, F. J. Martin, and A. D. Yates. 2024. Ensembl 2024. Nucleic Acids Res. 52(D1):D891–D899. https://doi.org/10.1093/nar/gkad1049.

Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. Annu. Rev. Anim. Biosci. 7:89–102. https://doi.org/10.1146/annurev-animal-020518-115024.

ARTICLE IN PRESS—UNCORRECTED PROOF




Howie, B., C. Fuchsberger, M. Stephens, J. Marchini, and G. R. Abecasis. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44:955–959. https://doi.org/10.1038/ng.2354.

Hughes, D. A., R. Bacigalupe, J. Wang, M. C. Rühlemann, R. Y. Tito, G. Falony, M. Joossens, S. Vieira-Silva, L. Henckaerts, L. Rymenans, C. Verspecht, S. Ring, A. Franke, K. H. Wade, N. J. Timpson, and J. Raes. 2020. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat. Microbiol. 5:1079–1087. https://doi.org/10.1038/s41564-020-0743-8.

Jami, E., A. Israel, A. Kotser, and I. Mizrahi. 2013. Exploring the bovine rumen bacterial community from birth to adulthood. ISME J. 7:1069–1079. https://doi.org/10.1038/ismej.2013.2.

Jiang, C., X. Hou, X. Gao, P. Liu, X. Guo, G. Hu, Q. Li, C. Huang, G. Li, W. Fang, W. Mai, C. Wu, Z. Xu, and P. Liu. 2023a. The 16S rDNA high-throughput sequencing correlation analysis of milk and gut microbial communities in mastitis Holstein cows. BMC Microbiol. 23:180. https://doi.org/10.1186/s12866-023-02925-7.

Jiang, Q., D. N. Sherlock, A. A. Elolimy, M. Vailati-Riboni, I. Yoon, and J. J. Loor. 2023b. Impact of a *Saccharomyces cerevisiae* fermentation product during an intestinal barrier challenge in lactating Holstein cows on ileal microbiota and markers of tissue structure and immunity. J. Anim. Sci. 101:skad309. https://doi.org/10.1093/jas/skad309.

Johannessen, M., F. Askarian, M. Sangvik, and J. E. Sollid. 2013. Bacterial interference with canonical NFκB signalling. Microbiology (Reading) 159:2001–2013. https://doi.org/10.1099/mic.0.069369-0.

Kaakoush, N. O. 2015. Insights into the role of Erysipelotrichaceae in the human host. Front. Cell. Infect. Microbiol. 5. https://doi.org/10.3389/fcimb.2015.00084.

Karrasch, T., and C. Jobin. 2008. NF-κB and the intestine: Friend or foe? Inflamm. Bowel Dis. 14:114–124. https://doi.org/10.1002/ibd.20243.

Khan, M. A. W., W. Z. Stephens, A. D. Mohammed, J. L. Round, and J. L. Kubinak. 2019. Does MHC heterozygosity influence microbiota form and function? PLoS One 14:e0215946. https://doi.org/10.1371/journal.pone.0215946.

Kim, M., J. Kim, L. A. Kuehn, J. L. Bono, E. D. Berry, N. Kalchayanand, H. C. Freetly, A. K. Benson, and J. E. Wells. 2014. Investigation of bacterial diversity in the feces of cattle fed different diets. J. Anim. Sci. 92:683–694. https://doi.org/10.2527/jas.2013-6841.

Klevenhusen, F., R. M. Petri, M.-T. Kleefisch, R. Khiaosa-ard, B. U. Metzler-Zebeli, and Q. Zebeli. 2017. Changes in fibre-adherent and fluid-associated microbial communities and fermentation profiles in the rumen of cattle fed diets differing in hay quality and concentrate amount. FEMS Microbiol. Ecol. 93:fix100. https://doi.org/10.1093/femsec/fix100.

Kubinak, J. L., W. Z. Stephens, R. Soto, C. Petersen, T. Chiaro, L. Gogokhia, R. Bell, N. J. Ajami, J. F. Petrosino, L. Morrison, W. K. Potts, P. E. Jensen, R. M. O'Connell, and J. L. Round. 2015. MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. Nat. Commun. 6:8642. https://doi.org/10.1038/ncomms9642.

Kurilshikov, A., C. Medina-Gomez, R. Bacigalupe, D. Radjabzadeh, J. Wang, A. Demirkan, C. I. Le Roy, J. A. Raygoza Garay, C. T. Finnicum, X. Liu, D. V. Zhernakova, M. J. Bonder, T. H. Hansen, F. Frost, M. C. Rühlemann, W. Turpin, J.-Y. Moon, H.-N. Kim, K. Lüll, E. Barkan, S. A. Shah, M. Fornage, J. Szopinska-Tokov, Z. D. Wallen, D. Borisevich, L. Agreus, A. Andreasson, C. Bang, L. Bedrani, J. T. Bell, H. Bisgaard, M. Boehnke, D. I. Boomsma, R. D. Burk, A. Claringbould, K. Croitoru, G. E. Davies, C. M. van Duijn, L. Duijts, G. Falony, J. Fu, A. van der Graaf, T. Hansen, G. Homuth, D. A. Hughes, R. G. Ijzerman, M. A. Jackson, V. W. V. Jaddoe, M. Joossens, T. Jørgensen, D. Keszthelyi, R. Knight, M. Laakso, M. Laudes, L. J. Launer, W. Lieb, A. J. Lusis, A. A. M. Masclee, H. A. Moll, Z. Mujagic, Q. Qibin, D. Rothschild, H. Shin, S. J. Sørensen, C. J. Steves, J. Thorsen, N. J. Timpson, R. Y. Tito, S. Vieira-Silva, U. Völker, H. Völzke, U. Võsa, K. H. Wade, S. Walter, K. Watanabe, S. Weiss, F. U. Weiss, O. Weissbrod, H.-J. Westra, G. Willemsen, H. Payami, D. M. A. E. Jonkers, A. Arias Vasquez, E. J. C. de Geus, K. A. Meyer, J. Stokholm, E. Segal, E. Org, C. Wijmenga, H.-L. Kim, R. C. Kaplan, T. D. Spector, A. G. Uitterlinden, F. Rivadeneira, A. Franke, M. M. Lerch, L. Franke, S. Sanna, M. D'Amato, O. Pedersen, A. D. Paterson, R. Kraaij, J. Raes, and A. Zhernakova. 2021. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat. Genet. 53:156–165. https://doi.org/10.1038/s41588-020-00763-1.

Labarta-Bajo, L., A. Gramalla-Schmitz, R. R. Gerner, K. R. Kazane, G. Humphrey, T. Schwartz, K. Sanders, A. Swafford, R. Knight, M. Raffatellu, and E. I. Zúñiga. 2020. CD8 T cells drive anorexia, dysbiosis, and blooms of a commensal with immunosuppressive potential after viral infection. Proc. Natl. Acad. Sci. USA 117:24998–25007. https://doi.org/10.1073/pnas.2003656117.

Lahti, L., and S. Shetty. 2017. Tools for microbiome analysis in R.

Liu, S., Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, E. Pairo-Castineira, K. D'Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, and L. Fang. 2022. A multi-tissue atlas of regulatory variants in cattle. Nat. Genet. 54:1438–1447. https://doi.org/10.1038/s41588-022-01153-5.

Lopera-Maya, E. A., A. Kurilshikov, A. van der Graaf, S. Hu, S. Andreu-Sánchez, L. Chen, A. V. Vila, R. Gacesa, T. Sinha, V. Collij, M. A. Y. Klaassen, L. A. Bolte, M. F. B. Gois, P. B. T. Neerincx, M. A. Swertz, R. Aguirre-Gamboa, P. Deelen, L. Franke, J. A. Kuivenhoven, E. A. Lopera-Maya, I. M. Nolte, S. Sanna, H. Snieder, M. A. Swertz, J. M. Vonk, C. Wijmenga, H. J. M. Harmsen, C. Wijmenga, J. Fu, R. K. Weersma, A. Zhernakova, and S. Sanna. 2022. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat. Genet. 54:143–151. https://doi.org/10.1038/s41588-021-00992-y.

Lynch, J. B., E. L. Gonzalez, K. Choy, K. F. Faull, T. Jewell, A. Arellano, J. Liang, K. B. Yu, J. Paramo, and E. Y. Hsiao. 2023. Gut microbiota *Turicibacter* strains differentially modify bile acids and host lipids. Nat. Commun. 14:3669. https://doi.org/10.1038/s41467-023-39403-7.

Makanjuola, B. O., F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

Marotz, C., K. J. Cavagnero, S. J. Song, D. McDonald, S. Wandro, G. Humphrey, M. Bryant, G. Ackermann, E. Diaz, and R. Knight. 2021. Evaluation of the effect of storage methods on fecal, saliva, and skin microbiome composition. mSystems 6:e01329-e20. https://doi.org/10.1128/msystems.01329-20.

Martínez-Álvaro, M., M. D. Auffret, C.-A. Duthie, R. J. Dewhurst, M. A. Cleveland, M. Watson, and R. Roehe. 2022. Bovine host genome acts on rumen microbiome function linked to methane emissions. Commun. Biol. 5:350. https://doi.org/10.1038/s42003-022-03293-0.

Martinez Boggio, G., A. Meynadier, A. J. Buitenhuis, and C. Marie-Etancelin. 2022. Host genetic control on rumen microbiota and its impact on dairy traits in sheep. Genet. Sel. Evol. 54:77. https://doi.org/10.1186/s12711-022-00769-9.

Martinez-Boggio, G., H. F. Monteiro, F. S. Lima, C. C. Figueiredo, R. S. Bisinotto, J. E. P. Santos, B. Mion, F. S. Schenkel, E. S. Ribeiro, K. A. Weigel, G. J. M. Rosa, and F. Peñagaricano. 2024. Investigating relationships between the host genome, rumen microbiome, and dairy cow feed efficiency using mediation analysis with structural equation modeling. J. Dairy Sci. 107:8193–8204. https://doi.org/10.3168/jds.2024-24675.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.

Meyer, K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 8:815–821. https://doi.org/10.1631/jzus.2007.B0815.

Monteiro, H. F., Z. Zhou, M. S. Gomes, P. M. G. Peixoto, E. C. R. Bonsaglia, I. F. Canisso, B. C. Weimer, and F. S. Lima. 2022. Rumen and lower gut microbiomes relationship with feed efficiency and produc-

ARTICLE IN PRESS—UNCORRECTED PROOF



tion traits throughout the lactation of Holstein dairy cows. Sci. Rep. 12:4904. https://doi.org/10.1038/s41598-022-08761-5.

Nagao-Kitamoto, H., S. Kitamoto, P. Kuffa, and N. Kamada. 2016. Pathogenic role of the gut microbiota in gastrointestinal diseases. Intest. Res. 14:127–138. https://doi.org/10.5217/ir.2016.14.2.127.

Nguyen, T. V., S. Bolormaa, C. M. Reich, A. J. Chamberlain, C. J. Vander Jagt, H. D. Daetwyler, and I. M. MacLeod. 2024. Empirical versus estimated accuracy of imputation: Optimising filtering thresholds for sequence imputation. Genet. Sel. Evol. 56:72. https://doi.org/10.1186/s12711-024-00942-2.

Noel, S. J., D. W. Olijhoek, F. Mclean, P. Løvendahl, P. Lund, and O. Højberg. 2019. Rumen and fecal microbial community structure of Holstein and Jersey dairy cows as affected by breed, diet, and residual feed intake. Animals (Basel) 9:498. https://doi.org/10.3390/ani9080498.

Oksanen, J., G.L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M.D. Caceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, D. McGlinn, M.-H. Ouellette, E. R. Cunha, T. Smith, A. Stier, C. J. F. T. Braak, and J. Weedon. 2022. vegan: Community Ecology Package.

Qin, Y., A. S. Havulinna, Y. Liu, P. Jousilahti, S. C. Ritchie, A. Tokolyi, J. G. Sanders, L. Valsta, M. Brożyńska, Q. Zhu, A. Tripathi, Y. Vázquez-Baeza, R. Loomba, S. Cheng, M. Jain, T. Niiranen, L. Lahti, R. Knight, V. Salomaa, M. Inouye, and G. Méric. 2022. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat. Genet. 54:134–142. https://doi.org/10.1038/s41588-021-00991-z.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41(D1):D590–D596. https://doi.org/10.1093/nar/gks1219.

Rothschild, D., O. Weissbrod, E. Barkan, A. Kurilshikov, T. Korem, D. Zeevi, P. I. Costea, A. Godneva, I. N. Kalka, N. Bar, S. Shilo, D. Lador, A. V. Vila, N. Zmora, M. Pevsner-Fischer, D. Israeli, N. Kosower, G. Malka, B. C. Wolf, T. Avnit-Sagi, M. Lotan-Pompan, A. Weinberger, Z. Halpern, S. Carmi, J. Fu, C. Wijmenga, A. Zhernakova, E. Elinav, and E. Segal. 2018. Environment dominates over host genetics in shaping human gut microbiota. Nature 555:210–215. https://doi.org/10.1038/nature25973.

Sahana, G., Z. Cai, M. P. Sanchez, A. C. Bouwman, and D. Boichard. 2023. Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle. J. Dairy Sci. 106:5218–5241. https://doi.org/10.3168/jds.2022-22694.

Sanchez, M.-P., A. Govignon-Gion, P. Croiseau, S. Fritz, C. Hozé, G. Miranda, P. Martin, A. Barbat-Leterrier, R. Letaïef, D. Rocha, M. Brochard, M. Boussaha, and D. Boichard. 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genet. Sel. Evol. 49:68. https://doi.org/10.1186/s12711-017-0344-z.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478.

Singh, A., H. Ashar, J. T. Butcher, and A. Ranjan. 2023. Age-associated changes in the gut microbiome impact efficacy of tumor immunomodulatory treatments. Exp. Gerontol. 181:112268. https://doi.org/10.1016/j.exger.2023.112268.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2:16. https://doi.org/10.1186/1742-9994-2-16.

Tamanai-Shacoori, Z., I. Smida, L. Bousarghin, O. Loreal, V. Meuric, S. B. Fong, M. Bonnaure-Mallet, and A. Jolivet-Gougeon. 2017. Roseburia Spp.: A marker of health? Future Microbiol. 12:157–170. https://doi.org/10.2217/fmb-2016-0130.

Tang, G., X. Wang, M. Cui, G. Ren, F. Yan, S. Wang, J. Yao, and X. Xu. 2023. The characteristics of diarrhea-related bacteria in suckling calves and their dynamic succession with ages. https://doi.org/10.21203/rs.3.rs-3411867/v1.

Tomkovich, S., and C. Jobin. 2016. Microbiota and host immune responses: A love–hate relationship. Immunology 147:1–10. https://doi.org/10.1111/imm.12538.

Tribout, T., P. Croiseau, R. Lefebvre, A. Barbat, M. Boussaha, S. Fritz, D. Boichard, C. Hoze, and M.-P. Sanchez. 2020. Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. Genet. Sel. Evol. 52:55. https://doi.org/10.1186/s12711-020-00575-1.

Turcot-Dubois, A.-L., B. Le Moullac-Vaidye, S. Despiau, F. Roubinet, N. Bovin, J. Le Pendu, and A. Blancher. 2007. Long-term evolution of the CAZY glycosyltransferase 6 (ABO) gene family from fishes to mammals—a birth-and-death evolution model. Glycobiology 17:516–528. https://doi.org/10.1093/glycob/cwm016.

Van der Auwera, G. A., and B. D. O'Connor. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. 1st ed. O'Reilly Media, Inc.

Veshkini, A., C. Kühn, F. Dengler, L. Bachmann, W. Liermann, C. Helm, R. Ulrich, C. Delling, and H. M. Hammon. 2024. *Cryptosporidium parvum* infection alters the intestinal mucosa transcriptome in neonatal calves: Impacts on epithelial barriers and transcellular transport systems. Front. Cell. Infect. Microbiol. 14:1495309. https://doi.org/10.3389/fcimb.2024.1495309.

Visscher, P. M., R. Thompson, and C. S. Haley. 1996. Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020. https://doi.org/10.1093/genetics/143.2.1013.

Wallace, R. J., G. Sasson, P. C. Garnsworthy, I. Tapio, E. Gregson, P. Bani, P. Huhtanen, A. R. Bayat, F. Strozzi, F. Biscarini, T. J. Snelling, N. Saunders, S. L. Potterton, J. Craigon, A. Minuti, E. Trevisi, M. L. Callegari, F. P. Cappelli, E. H. Cabezas-Garcia, J. Vilkki, C. Pinares-Patino, K. O. Fliegerová, J. Mrázek, H. Sechovcová, J. Kopečný, A. Bonin, F. Boyer, P. Taberlet, F. Kokou, E. Halperin, J. L. Williams, K. J. Shingfield, and I. Mizrahi. 2019. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci. Adv. 5:eaav8391. https://doi.org/10.1126/sciadv.aav8391.

Wang, Y., X. Nan, Y. Zhao, L. Jiang, H. Wang, F. Zhang, D. Hua, J. Liu, L. Yang, J. Yao, and B. Xiong. 2022. Discrepancies among healthy, subclinical mastitic, and clinical mastitic cows in fecal microbiome and metabolome and serum metabolome. J. Dairy Sci. 105:7668–7688. https://doi.org/10.3168/jds.2021-21654.

Wei, X.-L., T.-Q. Luo, J.-N. Li, Z.-C. Xue, Y. Wang, Y. Zhang, Y.-B. Chen, and C. Peng. 2021. Development and validation of a prognostic classifier based on lipid metabolism–related genes in gastric cancer. Front. Mol. Biosci. 8:691143. https://doi.org/10.3389/fmolb.2021.691143.

Wolter, M., E. T. Grant, M. Boudaud, N. A. Pudlo, G. V. Pereira, K. A. Eaton, E. C. Martens, and M. S. Desai. 2024. Diet-driven differential response of *Akkermansia muciniphila* modulates pathogen susceptibility. Mol. Syst. Biol. 20:596–625. https://doi.org/10.1038/s44320-024-00036-7.

Yang, H., J. Wu, X. Huang, Y. Zhou, Y. Zhang, M. Liu, Q. Liu, S. Ke, M. He, H. Fu, S. Fang, X. Xiong, H. Jiang, Z. Chen, Z. Wu, H. Gong, X. Tong, Y. Huang, J. Ma, J. Gao, C. Charlier, W. Coppieters, L. Shagam, Z. Zhang, H. Ai, B. Yang, M. Georges, C. Chen, and L. Huang. 2022. ABO genotype alters the gut microbiota by regulating GalNAc levels in pigs. Nature 606:358–367. https://doi.org/10.1038/s41586-022-04769-z.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Zhang, G., Y. Wang, H. Luo, W. Qiu, H. Zhang, L. Hu, Y. Wang, G. Dong, and G. Guo. 2019. The association between inflammaging and age-related changes in the ruminal and fecal microbiota among lactating Holstein cows. Front. Microbiol. 10:1803. https://doi.org/10.3389/fmicb.2019.01803.

Zhang, Q., G. Difford, G. Sahana, P. Løvendahl, J. Lassen, M. S. Lund, B. Guldbrandtsen, and L. Janss. 2020. Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. ISME J. 14:2019–2033. https://doi.org/10.1038/s41396-020-0663-x.

ARTICLE IN PRESS—UNCORRECTED PROOF



Zou, J., H. Zhang, Y. Huang, W. Xu, Y. Huang, S. Zuo, Z. Li, and H. Zhou. 2022. Multi-omics analysis of the tumor microenvironment in liver metastasis of colorectal cancer identified FJX1 as a novel biomarker. Front. Genet. 13:960954. https://doi.org/10.3389/fgene.2022.960954.

## ORCIDS

L. Brulin, https://orcid.org/0009-0006-8264-1064
M.-P. Sanchez, https://orcid.org/0000-0002-1371-5342
Z. Cai, https://orcid.org/0000-0002-9579-3415
S. Ducrocq, https://orcid.org/0009-0004-6898-156X
C. Audebert, https://orcid.org/0000-0002-6509-7574
G. Sahana, https://orcid.org/0000-0001-7608-7577