Microbiome

RESEARCH

Open Access

An integrated microbiome- and metabolome-genome-wide association study reveals the role of heritable ruminal microbial carbohydrate metabolism in lactation performance in Holstein dairy cows

Chenguang Zhang[1,3], Huifeng Liu[1,3], Xingwei Jiang[1,3], Zhihong Zhang[3,4], Xinfeng Hou[3,4], Yue Wang[1,3], Dangdang Wang[1,3], Zongjun Li[1,3], Yangchun Cao[1,3], Shengru Wu[1,3*], Sharon A. Huws[2*] and Junhu Yao[1,3*]

Abstract

Background Despite the growing number of studies investigating the connection between host genetics and the rumen microbiota, there remains a dearth of systematic research exploring the composition, function, and metabolic traits of highly heritable rumen microbiota influenced by host genetics. Furthermore, the impact of these highly heritable subsets on lactation performance in cows remains unknown. To address this gap, we collected and analyzed whole-genome resequencing data, rumen metagenomes, rumen metabolomes and short-chain fatty acids (SCFAs) content, and lactation performance phenotypes from a cohort of 304 dairy cows.

Results The results indicated that the proportions of highly heritable subsets ($h^2 \geq 0.2$) of the rumen microbial composition (55%), function (39% KEGG and 28% CAZy), and metabolites (18%) decreased sequentially. Moreover, the highly heritable microbes can increase energy-corrected milk (ECM) production by reducing the rumen acetate/propionate ratio, according to the structural equation model (SEM) analysis (CFI = 0.898). Furthermore, the highly heritable enzymes involved in the SCFA synthesis metabolic pathway can promote the synthesis of propionate and inhibit the acetate synthesis. Next, the same significant SNP variants were used to integrate information from genome-wide association studies (GWASs), microbiome-GWASs, metabolome-GWASs, and microbiome-wide association studies (mWASs). The identified single nucleotide polymorphisms (SNPs) of rs43470227 and rs43472732 on *SLC30A9* (Zn^{2+} transport) ($P < 0.05$/nSNPs) can affect the abundance of rumen microbes such as *Prevotella_sp.*, *Prevotella_sp._E15-22*, *Prevotella_sp._E13-27*, which have the oligosaccharide-degradation enzymes genes, including the GH10, GH13, GH43, GH95, and GH115 families. The identified SNPs of chr25:11,177 on *5s_rRNA* (small ribosomal RNA) ($P < 0.05$/nSNPs) were linked to ECM, the abundance alteration of *Pseudobutyrivibrio_sp.* (a genus that was also showed to be linked to the ECM production via the mWASs analysis), GH24 (lysozyme), and 9,10,13-TriHOME (linoleic acid metabolism).

*Correspondence:
Shengru Wu
wushengru2013@163.com
Sharon A. Huws
s.huws@qub.ac.uk
Junhu Yao
yaojunhu2004@sohu.com
Full list of author information is available at the end of the article

Moreover, ECM, and the abundances of *Pseudobutyrivibrio* sp., GH24, and 9,10,13-TRIHOME were significantly greater in the GG genotype than in the AG genotype at chr25:11,177 ($P < 0.05$). By further the SEM analysis, GH24 was positively correlated with *Pseudobutyrivibrio* sp., which was positively correlated with 9,10,13-triHOME and subsequently positively correlated with ECM (CFI = 0.942).

Conclusion Our comprehensive study revealed the distinct heritability patterns of rumen microbial composition, function, and metabolism. Additionally, we shed light on the influence of host SNP variants on the rumen microbes with carbohydrate metabolism and their subsequent effects on lactation performance. Collectively, these findings offer compelling evidence for the host-microbe interactions, wherein cows actively modulate their rumen microbiota through SNP variants to regulate their own lactation performance.

Keywords Dairy cow, Host genetics, Ruminal metagenome, Ruminal metabolome, Heritability, GWAS, Lactation performance

Introduction

With an annual global per capita consumption of dairy milk which has been more than 140 kg/year, it has become an important and irreplaceable high-quality animal protein product [1]. However, in comparison with that in developed countries, the per capita consumption of dairy products in China is limited, accounting for only 1/3 of the world average. However, considering that the consumption is increasing [2], sustainably increasing high-quality milk yield has become one of the most important outcomes for dairy cows. Notably, increased high-quality milk yield can decrease the ratio of feed to dairy production, reduce the cost of producing milk, and ultimately determine the price at which people obtain milk [3]. Furthermore, sustainably increasing milk yield will reduce the methane output of cows when expressed as /kg milk is produced (methane intensity) [4]. Classic genetics studies have suggested a heritability of approximately 0.4 for cow milk yield [5]. Genome-wide association studies (GWASs) focused on milk performance have identified several natural variations, including single nucleotide polymorphisms (SNP) of several key genes, such as the well-known *DGAT1* gene located ~ 1.8 Mb on chr14, and promoted the genetic selection of high lactation performance dairy cows [5–10], which are beneficial for meeting the increased demand for milk. However, as a complex trait influenced by multiple factors, studies on lactation performance from only a genetic breeding perspective are limited. With the growth of the world's population and the increase in per capita milk consumption driven by increased individual economic growth and urbanization, coupled with the challenges of climate change, there is an urgent need to link the key factors affecting lactation, to deeply analyze the regulatory mechanisms of lactation, and to develop more refined breeding and nutritional intervention strategies.

It is increasingly recognized that the rumen harbors complex microbial communities that play vital roles in producing short-chain fatty acids (SCFAs), which provided 60–70% of the metabolizable energy for promoting milk yield by affecting feed utilization and milk quality [4]. The host has recently been proven to affect the gut microbiota composition, function, and their metabolism processes and metabolites. For instance, the 2.3 kb deletion in the ABO blood type gene leads to a reduction in the N-acetyl galactosamine (GalNAc) concentration in the intestine of domestic pigs, resulting in a decrease in the abundance of GalNAc-utilizing bacteria [11]. Recent studies have further revealed that GalNAc, influenced by ABO blood types, can selectively enrich the gut microbiota with specific metabolic function gene clusters [12]. Numerous studies have also focused on the significant impact of host SNP variants on the composition of the rumen microbiota in ruminants [13]. The identified heritable microbiota constituents could in part determine methane production and lactation performance [14]. However, in comparison with the widely suggested loci that regulate the gut microbiome composition in monogastric mammals, such as humans and pigs [11, 12, 15–17], only a few studies have focused on the identification of specific variants related to milk yield-associated ruminal microbial genera [18, 19]. Furthermore, considering that the ruminal microbiota is highly diverse but that microbiome function is more conserved than that of other microbiota [20], the lack of identification of heritable microbial functions that connect microbiota composition and phenotypic changes limits our understanding of the relationship between host gene variation and rumen microbiome colonization. Hence, the study of the mechanisms of ruminant-rumen microbiota interactions is still in the initial phase, and understanding how host SNP variants affect the microbiome will be a critical step toward regulating the rumen microbiome to improve ruminant performance, including cows' lactation performance [14].

To better understand how cows' SNP variants influence the ruminal microbiome and subsequent lactation performance improvement, we investigated the ruminal

microbiome and metabolome and host genomic SNP information of 304 Holstein dairy cows by using shotgun metagenome sequencing, liquid chromatography–mass spectrometry (LC–MS), and genome resequencing. Utilizing these discoveries, we examined the heritability of both the composition and function of the ruminal microbiome, as well as the associated ruminal metabolome. Furthermore, we uncovered the contributions of these factors in explaining the variations in lactation performance. Furthermore, lactation performance can be regulated by host metabolism, the ruminal microbiota, and their metabolites. Hence, by identifying the same SNP variants identified in genome-wide association studies, microbiome-genome-wide association studies (m-GWASs), metabolome-genome-wide association studies (M-GWASs), and microbiome-wide association studies (mWASs), we investigated the interplay between host genetics, milk traits, and the rumen microbiome using comprehensive multiomic technologies, exploring the specific factors that collectively influence changes in lactation performance. To the best of our knowledge, this is the first study to delve into the interactions among host genetics, milk traits, and the composition, function, and metabolome of the ruminal microbiome.

Materials and methods

Animals, phenotypic data, and sample collection

The Holstein dairy cows that participated in the sample collection process were all fed on the same dairy farm. The sampling location was located in Leyuan Animal Husbandry, Hebei Province (N, 37.0766°; E, 115.4008°). We selected 304 lactating cows from 5906 lactating cows based on the principle of similar lactation periods and parity and were fed the same diet. The physiological status, dietary formula, and nutritional composition of the cows are shown in Table S1. Specifically, for the dairy cows included in the sampling, the parity was 2–3, the lactation time was 99.82 ± 52.93 days, and the diet consisted of 45% forage and 55% concentrate (Table S1). For each individual cow, we collected rumen fluid and blood samples from 8:00 am to 10:00 am after the first milking every morning. The sample collection lasted for 15 days in October 2022 (the Institutional Animal Care and Use Committee at Northwest A&F University granted approval for the protocol). The daily milk production of each individual cow was recorded in the Fonton system (Fonton, Nanjing, China). The average milk yield (MY), milk fat (MF), milk protein (MP), and milk lactose (ML) during the 50–150 lactation period were used to calculate the final lactating phenotype. Data related to milk composition were collected with MilkoScan FT1 (FOSS, Hillerød, Denmark) from the monthly DHI (Dairy Herd Improvement) of the sampled farm. Next, we calculated the energy-corrected milk (ECM) yield based on the MY, MF, MP, and ML to characterize the lactation performance of the cows (Fig. S1A–E). We collected whole blood samples from the caudal vein using blood collection tubes containing ethylenediaminetetraacetic acid disodium salt (EDTA) for anticoagulant blood and then stored them at − 20 °C until further analysis. Rumen fluid samples were collected via esophageal tubing using oral stomach tubes. During sampling, the first 50 mL of ruminal fluid was discarded to avoid saliva contamination, and the next 50 mL rumen fluid was strained through four layers of sterile cheesecloth under an environment with constant flux of CO_2. After sampling from each cow, the rumen fluid was promptly packaged and temporarily stored in liquid nitrogen and finally stored at − 80 °C.

Ruminal SCFAs measurement

The concentrations of SCFAs including acetate, propionate, butyrate, acetate/propionate (A:P), and total acid (TA) were determined using gas chromatography (Agilent 7820A, Santa Clara, CA, USA) with a capillary column (AE-FFAP of 30 m × 0.25 mm × 0.33 μm, ATECH Technologies Co., Lanzhou, China) (Fig. S1F–J). Briefly, the thawed rumen fluid samples were centrifuged for 10 min at 13,500 × g at 4 °C. The supernatant was mixed with 200 μL of metaphosphate (25 w/v), incubated for 4 h, and then centrifuged for 15 min at 13,500 × g at 4 °C for protein and impurity precipitation. Then, crotonic acid was added to the supernatant as an internal standard. The final supernatant was transferred to a gas phase bottle through a filter. The supernatant of the gas phase bottle was analyzed using gas chromatography with a capillary column. The SCFA concentration detection program was performed as previously described [21, 22].

Metagenome sequencing

For metabolomic analysis, rumen fluid samples were collected from the same 304 cows. The repeat bead-beating plus column method was used to extract genomic DNA from rumen fluid samples [23] with the Mag-Bind® Stool DNA Kit M4015 (Omega Biotek, Norcross, GA, USA). The rumen microbial DNA extract was fragmented using a Covaris M220 (Gene Company Limited, China) to achieve an average size of approximately 400 bp. Paired-end library construction was carried out using a Rapid DNA Sequencing Kit (NEXTFLEX) (Bioo Scientific, Austin, TX, USA). Adapters containing the complete sequence of the sequencing primer hybridization sites were ligated to the blunt ends of the fragments. Paired-end sequencing was performed on an Illumina NovaSeq (Illumina, San Diego, CA, USA) with a NovaSeq 6000 S4 Reagent Kit v1.5 at Majorbio Biopharm Technology Co., Ltd. (Shanghai, China).

The fastp [24] (https://github.com/OpenGene/fastp, version 0.20.0) was used for quality control of the raw Illumina reads (reads with a length < 50 bp, a quality value < 20 or N bases were trimmed). The host (cow) genome reads were aligned and removed by using the Burrows-Wheeler-Alignment Tool (BWA) [25] (http://biobwa.sourceforge.net, version 0.7.9a). The metagenomic data were assembled using MEGAHIT [26] (https://github.com/voutcn/megahit, version 1.1.2), which employs succinct de Bruijn graphs. Open reading frames (ORFs) were predicted from each assembled contig using Prodigal [27] and MetaGene [28] (http://metagene.cb.k.u-tokyo.ac.jp/). ORFs with a length of ≥ 100 bp were extracted and translated into amino acid sequences using the NCBI translation table. To construct a nonredundant gene contigs, CD-HIT [29] (http://www.bioinformatics.org/cd-hit/, version 4.6.1) was utilized with a 90% sequence identity and a 90% coverage threshold. The SOAPaligner [30] (http://soap.genomics.org.cn/, version 2.21) was employed to align high-quality reads to nonredundant gene contigs, enabling the calculation of gene abundance with 95% identity, and the gene abundance in each sample was calculated as reads per kilobase per million mapped reads (RPKM). The relative abundance of a species in single sample was calculated based on the ratio of its RPKM to the sum of RPKM of all detected species in this sample, which was used for ranking microbes.

The representative sequences from the nonredundant gene contigs were aligned to the NR database by Diamond [31] (http://www.diamondsearch.org/index.php, version 0.8.35). Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation was conducted using Diamond [31] against the KEGG database (http://www.genome.jp/keeg/). Carbohydrate-active enzyme annotation was conducted using hmmscan (http://HMMER.janelia.org/search/hmmscan) against the Carbohydrate-Active enZYmes (CAZys) database (http://www.cazy.org/). All these databases had an E-value cut-off of $1e^{-5}$ while annotating ORFs.

After the taxa of each KEGG function-related ORFs was determined, the statistical composition based on the relative contribution (%) of species assigned to the KEGG function groups was determined using nonparametric Kruskal–Wallis analysis of variance ($P < 0.05$) followed by multiple comparisons with Bonferroni correction [32].

Metabolomic analysis

For metabolomic analysis, rumen fluid samples were collected from the same 304 cows. The process of sample pre-processing and LC–MS/MS detection followed the previously described protocol [33]. Briefly, protein precipitation of rumen fluid samples was achieved by adding methanol/acetonitrile (1:1, v/v) buffer and subjecting them to ultrasonic bath treatment (Kunshan Ultrasonic Instrument Co. Ltd., China). The resulting concentrated product was then subjected to LC–MS analysis using an UHPLC system (Q-Exactive, Thermo Fisher Scientific, USA). Chromatographic separations were performed on an ACQUITY UPLC HSS T3 column (100 mm × 2.1 mm, 1.8 μm) (Waters Co., USA). LC–MS data were collected in both positive and negative ionization modes using an electrospray ionization source.

Supervised orthogonal partial least-squares discriminant analysis (OPLS-DA) was performed using metaX [34]. The metabolites with high heritability were mapped to the KEGG pathways using the KEGG database (http://www.genome.jp/kegg/). Significantly enriched pathways were identified using Fisher's exact test, with the scipy.stats Python package (https://docs.scipy.org/doc/scipy/) utilized for this analysis.

Whole-genome resequencing

For whole-genome resequencing, blood samples were collected from the same 304 cows. Genomic DNA from the host was isolated from whole blood samples using a whole blood Genomic DNA Extraction Kit (BIOWEFIND Company, Wuhan, China). The 304 host-quality DNA samples were subjected to whole-genome resequencing on the DNBSEQ-T7 platform (MGI-Shenzhen, China), and 150 bp paired-end reads were generated. The fastp [24] software was used for quality control of the raw FASTQ reads. The clean FASTQ reads were mapped to the cow reference genome by BWA [25] with the command "bwa mem –M" and converted to the BAM format using SAMtools (http://github.com/samtools/samtools) with the command "samtools view -bS." The duplicate reads were subsequently sorted and labelled for PCR duplication by the Genome Analysis Toolkit [35] (GATK, https://software.broadinstitute.org/gatk/) with the commands "gatk SortSam" and "gatk MarkDuplicates." Variant detection was performed based on chromosomal information using the command "gatk HaplotypeCaller -L." The chromosomes of all the samples were merged, and the genotype files were generated with the commands "gatk CombineGVCFs" and "gatk GenotypeGVCFs." All 30 chromosomes were subsequently merged with the command "gatk MergeVcfs." The low-quality variants were filtered out with the commands "QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0." The PLINK (https://zzz.bwh.harvard.edu/plink/) was used for quality control for 10,494,909 unfiltered SNP variants generated by the above steps with the following parameters: "–geno 0.1, –maf 0.05, –mind 0.1."

A total of 2,337,054 SNPs distributed across 30 chromosomes and 304 dairy cows were ultimately obtained for analysis.

Estimation of phenotypic and rumen microbial heritability

We calculated the heritability (h^2) for the phenotype (lactation performance (MY, MF, MP, ML, and ECM), the rumen SCFAs (acetate, propionate, butyrate, A:P ratio, and TA), rumen microbes with relative abundances exceeding 0.01% at the species level, rumen microbial pathways at KEGG level 3, and rumen microbial CAZy module at the class and family levels, and all rumen microbial KEGG enzymes. Briefly, the heritability (h^2) represents the contribution of host SNPs to the composition of rumen microbiota and metabolites or the similarity of rumen microbes and metabolites among significantly related individuals [36]. Moreover, the centered log-ratio transformation (CLR) was performed to standardize metagenomic and metabolomic data, using population structure (PC1-3) of host, lactation times, and parity as covariates to correct heritability. Genome-based restricted maximum likelihood (GREML) analysis was used to estimate the heritability (h^2)-based genetic relationship matrix (GRM) by GCTA [37] (https://yanglab.westlake.edu.cn/software/gcta) with the command "–reml". The model is as follows:

$$y = Xa + Wb + e$$

where y represents the phenotypic, metabolomic, and rumen metagenomic data; a represents the fixed effects, including population structure (PC1-3) of host, lactation times, and parity as covariates; b represents the additive genetic effects following a distribution of N (0, $\mathbf{G}\sigma_a^2$), where G represents the GRM and σ_a^2 represents the additive genetic variance; and e represents the residual effects following a distribution of N (0, $\mathbf{I}\sigma_E^2$), where I represents an identity matrix and σ_E^2 represents the residual variance. X and W represent the incidence matrices for a and b, respectively. The heritability of the given data was tested using h^2, with a threshold value of 0.2.

Identification of significant SNPs based on GWASs using mixed linear model (MLM)

To establish a mixed linear model (MLM) to identify significant SNPs for highly heritable phenotypes, metabolites, and microbes, a kinship matrix was established using GEMMA [38] (https://github.com/genetics-statistics/GEMMA) with the command "-gk," and the population structure was established using PLINK with the command "—pca." The significant SNPs were detected using a mixed linear model with the GEMMA command "-lmm 1." The population structure (PC1-3), lactation times, and parity were used as covariates to correct the results. The equation was as follows:

$$Y = X\alpha + Z\beta + W\mu + e$$

where Y represents the phenotype, highly heritable metabolites, and microbes; $X\alpha$ represents the population structure, lactation times, and parity of the fixed effect; $Z\beta$ represents the SNP of the marker effect; $W\mu$ represents the kinship matrix of the random effect; and e represents the residual.

Subsequently, we set the genome-wide significance threshold based on a significance level of 1/nSNPs (1/2337054 = 4.28E-07, -log10(P) = 6.37) for significant associations and 0.05/nSNPs (0.05/2337054 = 2.14E-08, -log10(P) = 7.67) for extremely significant associations. We used the Variant Effect Predictor (https://www.ensembl.org/vep) for gene annotation.

Construction of the co-occurrence network

Co-occurrence network analysis was performed based on the Spearman correlation among the rumen microbes with relative abundances exceeding 0.01% at the species level using the R package ggClusterNet [39]. The degree centrality, closeness centrality, and betweenness centrality of microbes in the network were calculated and ranked using the R package ggClusterNet [39]. Briefly, the degree centrality refers to the number of connections or edges that a node has in a network. It measures the number of direct neighbors or connections that a node has. In a directed network, there can be inward and outward degrees, indicating the number of incoming and outgoing connections for a node. The betweenness centrality is a measure of a node's centrality in a network based on its position in connecting other nodes. It quantifies the number of times a node acts as a bridge or intermediary along the shortest path between other nodes in the network. Nodes with high betweenness centrality have a significant influence on the flow of information or interactions between other nodes in a network. And the closeness centrality measures how close or easily reachable a node is to the other nodes in a network. It is calculated as the inverse of the sum of the shortest path distances between a node and all other nodes in the network. A node with a higher closeness centrality is considered to be more central as it can reach other nodes more efficiently.

The association between the rumen microbiotamatrix and phenotypematrix by using the Mantel test

We first constructed a rumen microbial α diversity matrix with ACE, Chao1, Shannon, and β diversity matrices with PC1, PC2, and PC3. The Mantel test was used [40] to study the relationship between α-diversity or

β-diversity and phenotype using Spearman correlation (9999 permutations) with the R package ggcor. Moreover, we constructed a rumen SCFA matrix with rumen acetate, propionate, butyrate, A:P ratio, and total acid to explore the correlation between MY, rumen SCFAs, and the 50 most abundant rumen microbes at the species level using the Mantel test.

The causal relationships among highly heritable subsets of the rumen microbiota, rumen SCFAs, and MY according to structural equation model (SEM)

We used the 5 microbes with the highest and lowest heritability among the rumen microbes with the top 50 highest relative abundances at the species level to generate a high-heritability latent variable and a low-heritability latent variable, respectively, in a structural equation model (SEM) analysis. The goodness-of-fit of the SEM was checked using the root-mean-square error of approximation (RMSEA), the chi-squared test (chisq), and the comparative fit index (CFI). SEM was conducted using the R package lavaan [41].

Microbiota-wide association studies (mWAS)

The mWAS between rumen microbes with a relative abundance exceeding 0.01% at the species level, and the ECM and A:P ratio were determined in R. The approach was performed as described previously [42, 43].

Random forest

The key microbe markers discovered through the mWAS were further validated using the random forest method. Specifically, based on the ECM and rumen A:P ratio, 304 cows were divided into 5 groups from low to high scores. Microbe importance was ranked by the percentage decrease in the prediction accuracy of the model that occurred when the microbes were removed. To estimate the minimum number of top ranking discriminative taxa required for prediction, tenfold cross validation was implemented using the "rfcv" function in the "randomForest" package [44] and was applied over 100 iterations. The random forest algorithm was conducted with the R package "randomForest" with a default mtry parameter of p/3 where *p* was the number of input microbe species.

Results

Overview of the core rumen microbiota that affects the ECM

In this study, the ECM is the milk yield corrected by the milk composition and can more fully reflect the lactation performance of cows. The 304 cows were assigned to the low ($n=102$), medium ($n=101$), or high ($n=101$) group based on their ECM, with average ECMs of 31.8, 42.6, and 52.4 kg/d, respectively. Moreover, we found that MY was positively related to propionate and negatively correlated with rumen A:P ratio. ECM was negatively related to only rumen A:P ratio (Fig. S1K).

For microbial diversity, at the α-diversity level, the Chao and Ace indices of the low-ECM group were significantly greater than those of the high-ECM group ($P<0.05$) (Fig. S2A). At the β diversity level, the low, medium, and high groups were not significantly clustered separately in the PCoA and NMDS coordinate systems (Fig. S2B). Next, the correlations between the α diversity (Chao, Shannon, and Simpson), β diversity (PC1, PC2, and PC3) of the rumen microbiota and the rumen SCFAs, lactation performance were evaluated using the Mantel test.

Overall, β diversity was significantly correlated with MY and ECM (Mantel's $P<0.05$) (Fig. 1A), while α diversity was significantly correlated with MY, ECM, propionate, and butyrate (Mantel's $P<0.05$) (Fig. 1A).

At the domain level, the high-ECM group exhibited significantly higher abundances of bacteria and viruses compared to the low-ECM group, while the high-ECM group showed significantly lower abundances of archaea and eukaryotes ($P<0.05$) (Fig. 1B). At the species level, *Prevotella_sp*, *Prevotella_lacticifex*, *Prevotella_mizrahii*, *Eubacterium_sp*, and *Succinivibrionaceae_bacterium* were the markers of the high-ECM group based on lefse analysis (Fig. S2C).

In terms of rumen microbial CAZy at the class level, the high-ECM group exhibited a significantly higher abundance of GH and CBM models, whereas the GT model had significantly lower abundance in the high-ECM group ($P<0.05$) (Fig. 1B). At the KEGG level 1, the high-ECM group showed significantly higher abundance of "metabolism" and "genetic information processing," while "environmental information processing,"

(See figure on next page.)

Fig. 1 Relationships between the rumen microbiota and the phenotype of dairy cows. **A** The relationship between the microbial diversity matrix and phenotype matrix was determined based on Mantel's test. The α diversity indices included ACE, Chao1, and Shannon indices. The β diversity indices included PC1, PC2, and PC3 from the PCoA. MY: milk yield, MF: milk fat, MP: milk protein, ML: milk lactose, ECM: energy-corrected milk, A:P: acetate/propionate, TA: total acid. **B** Differences in the rumen microbiome at the domain, KEGG level 1, and CAZy family levels among the low (31.8 kg/d), medium (42.6 kg/d), and high (52.4 kg/d) ECM groups. **C** The relationship between the phenotype matrix and the top 50 microbial matrices at the species level was determined based on Mantel's test. Lactation included milk yield, milk fat, milk protein, and milk lactose. The ruSCFAs included acetate, propionate, butyrate, A:P, and total acid. **D** The co-occurrence networks of the microbes at the species level with relative abundances greater than 0.01%

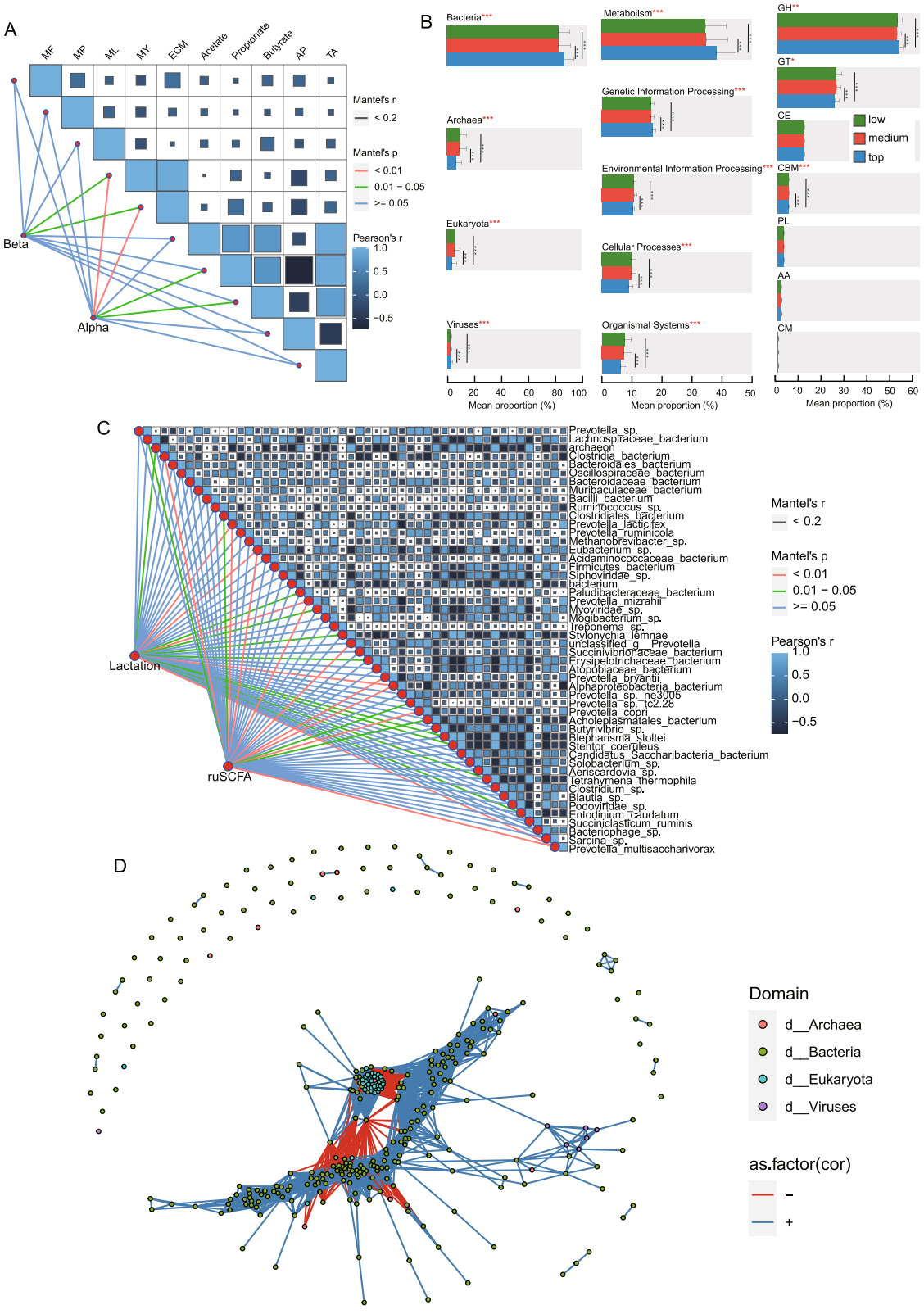

Fig. 1 (See legend on previous page.)

"human diseases," "cellular processes," and "organismal systems" were significantly lower abundance in the high-ECM group ($P < 0.05$) (Fig. 1B).

Furthermore, through the Mantel test, we identified rumen microbes significantly related to the phenotype at the species level. *Prevotella_lacticifex*, *Eubacterium_sp.*, *Prevotella_mizrahii*, *Alphaproteobacteria_bacterium*, *Prevotella_multisaccharivorax*, *Prevotella_bacterium*, *Elusimicrobia_bacterium*, *Prevotella_sp._AGR2160*, *Oribacterium_sp.*, and *Pseudobutyrivibrio_sp* were significantly correlated with rumen SCFAs and lactation performance (Fig. 1D). Next, we established a network of microbes with relative abundances exceeding 0.01% at the species level to identify the core microbes (Table S2). We found 276 microbes involved in the interaction, while 61 microbes existed independently in the rumen microbiota (Fig. 1E). *Alphaproteobacteria_bacterium* was the microbe with the highest degree and betweenness, *butyrivibrio_fibrisolvens*, *Acidaminococcaceae_bacterium*, *Succiniclasticum_ruminis*, *Schwartzia_sp.*, *Methanobrevibacter_ruminantium*, and *Pseudobutyrivibrio_sp.* were the microbes with the highest closeness (Table S2).

GWAS identified host genetics that affect rumen SCFAs and lactation performance

Among lactation performance, MY, ECM, MP, and ML had high heritability ($h^2 \geq 0.2$). Among rumen SCFAs, A:P ratio and TA had high heritability ($h^2 \geq 0.2$) (Fig. 2A). Among the associations between 2,337,054 SNPs (Fig. S3A) and phenotype (lactation performance and rumen SCFAs) using genetic relationship (Fig. S3B) as the random effect and population structure (Fig. S3C) as the covariate, a total of 57 significant associations were obtained, including 25% with an intro variant, 50% with an intergenic variant, 21% with a downstream gene variant, and 4% with an upstream gene variant. Forty-three SNP variants significantly affected lactation performance (-log10(P) > 6.39), which were annotated to the genes *RNF220*, *TSPAN9*, *RDH12*, *UGGT2*, *CDH4*, *EPG5*, and *5s_rRNA* (Fig. 2B and D). Therein, chr12:28,170,430, chr12:73,181,457, chr13:55,209,376, chr21:49,278,895, chr25:16,922, chr25:11,147, chr25:11,153, chr25:11,168, chr25:11,177, chrX:28,969,015, chrX:28,969,039, chrX:28,969,042, chrX:28,969,055, and chrX:28,969,077 had extremely significant impacts on lactation performance (-log10(P) > 7.67) (Table S3). Fourteen SNP variants significantly affected rumen SCFAs (-log10(P) > 6.39), which were annotated to the *CDH13* and *CD99 genes* (Fig. 2C and E). Therein, chrX:84,668,063 had an extremely significant impact on lactation performance (-log10(P) > 7.67) (Table S3).

Identification of highly heritable microbes and their regulatory SNPs via mGWAS

Among 337 microbes with relative abundance exceeding 0.01% at species level, 170 heritable species belonged to the Bacteria domain (most), while 3 heritable species belonged to the Archaea domain (least) ($h^2 \geq 0.2$) (Fig. 3A) (Table S2). Among the microbes with the top 100 most abundant species, 44 were highly heritable microbes, 14 of which were Prevotellaceae. *Methanobrevibacter_millerae, Methanocorpusculum_sp.,* and *Methanobrevibacter_thaueri* belong to the Archaea domain. *Myoviridae_sp., CrAss-like_virus_sp., Podoviridae_sp., Bacteriophage_sp.,* and *Siphoviridae_sp.* belonged to the virus domain (Table S2). For rumen microbial function at KEGG level 1, 72 highly heritable pathways were associated with "metabolism" (most), while only 12 highly heritable pathways were associated with "cellular processes" and "environmental information processing" (least) ($h^2 \geq 0.2$) (Table S4). For rumen microbial function at the CAZy family level, 92 highly heritable CAZy modules were associated with GH (most), while 3 highly heritable CAZy modules were associated with AA (least) ($h^2 \geq 0.2$) (Table S5). Next, we used microbes with relative abundances exceeding 0.01% to observe the relationship between microbial heritability and the ecological niche in the network (Fig. 1D). We found that the heritability of rumen microbes was positively related to closeness and betweenness ($P < 0.05$) (Fig. 3B).

Among the associations between 2,337,054 chromosomal genetic variants and 44 highly heritable microbial features from rumen microbes with the top 100 most abundant genera at the species level, 104 significant associations with 25 microbes were obtained, including 50% of the variants, 29% of the intergenic variants, 10% of the downstream gene variants, and 5% of the upstream gene variants (Fig. S4B), which were annotated to 26 genes, such as the *SLC30A9* and *5s_rRNA* (-log10(P) > 6.39) genes (Fig. 3C). Among these, chr1:776,938, 776,939, chr6:60,792,419, chr12:2,637,985, chr15:2,302,734, and chrX:28,987,601 had extremely significant impacts on highly heritable microbes (-log10(P) > 7.67) (Table S6).

The characteristics of highly heritable subsets in the rumen microbiota

To further study the differences in metabolic functions between highly heritable and less heritable microbes, we focused on the top 50 species according to relative abundance, of which 20 were highly heritable microbes ($h^2 \geq 0.2$), and 30 were lowly heritable microbes ($h^2 < 0.2$) (Table S2). According to the relative contribution (%) of the rumen microbes for metabolic pathway enrichment at KEGG level 3 (Table S7), among the top 50 microbes with relative abundance, 43 microbes contributed more than 1% to 143 KEGG metabolic pathways, in which included 15 highly heritable microbes and 28 low heritable microbes. Next, we focused on the microbes with the highest contribution to each pathway, the highly heritable

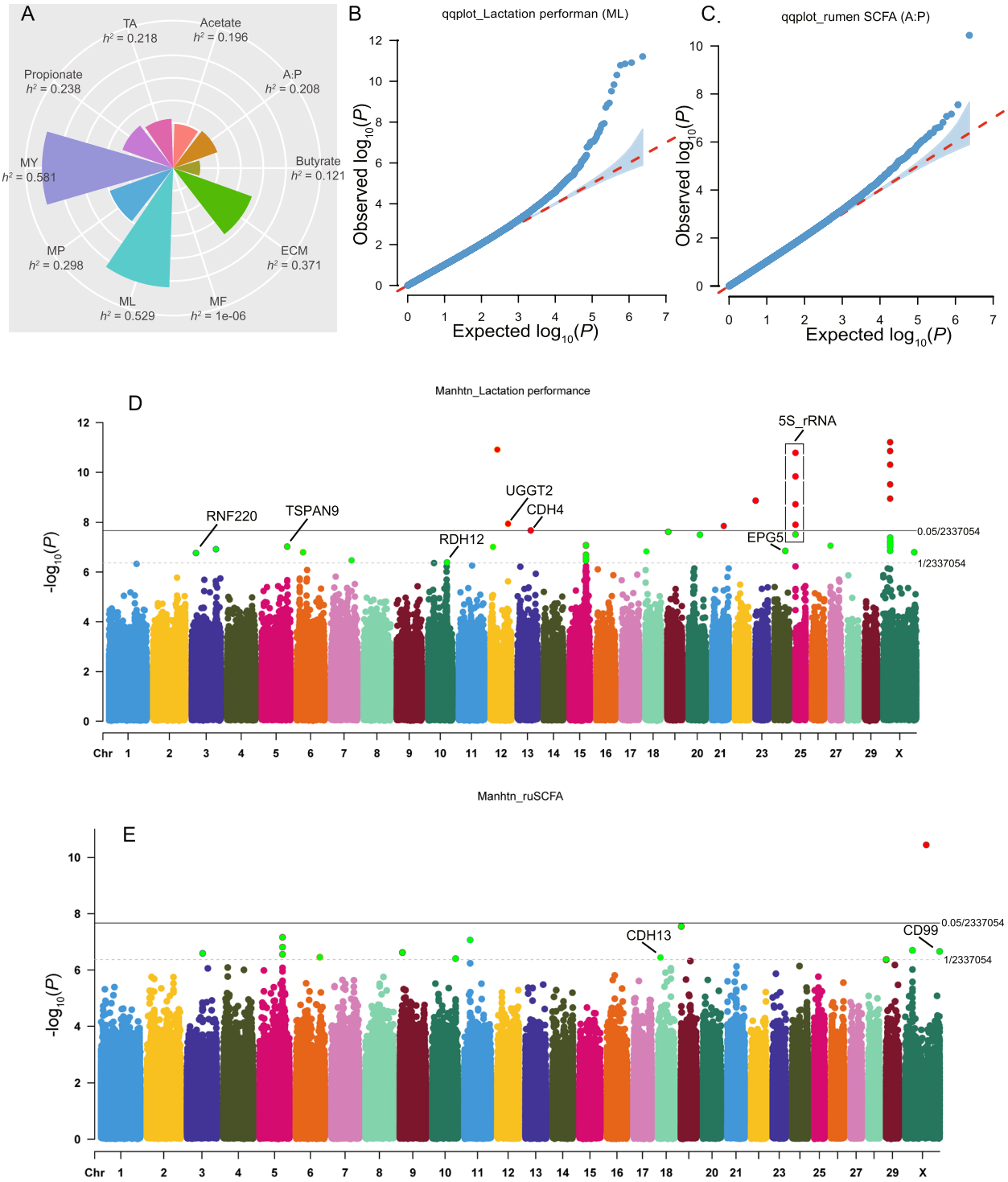

Fig. 2 The heritability and significant variants of the phenotype of dairy cows. **A** The heritability of lactation performance and rumen SCFAs. MY: milk yield, MF: milk fat, MP: milk protein, ML: milk lactose, ECM: energy-corrected milk, A:P: acetate/propionate, TA: total acid. **B** The Q–Q plot of lactation performance. **C** Q–Q plot of rumen SCFAs. **D** Manhattan plot of lactation performance. The significance threshold was 1/nSNP = 4.28E-07. The extremely significant threshold was 0.05/nSNP = 2.14E-08. **E** Manhattan plot of rumen SCFAs. The significance threshold was 1/nSNP = 4.28E-07. The extremely significant threshold was 0.05/nSNP = 2.14E-08

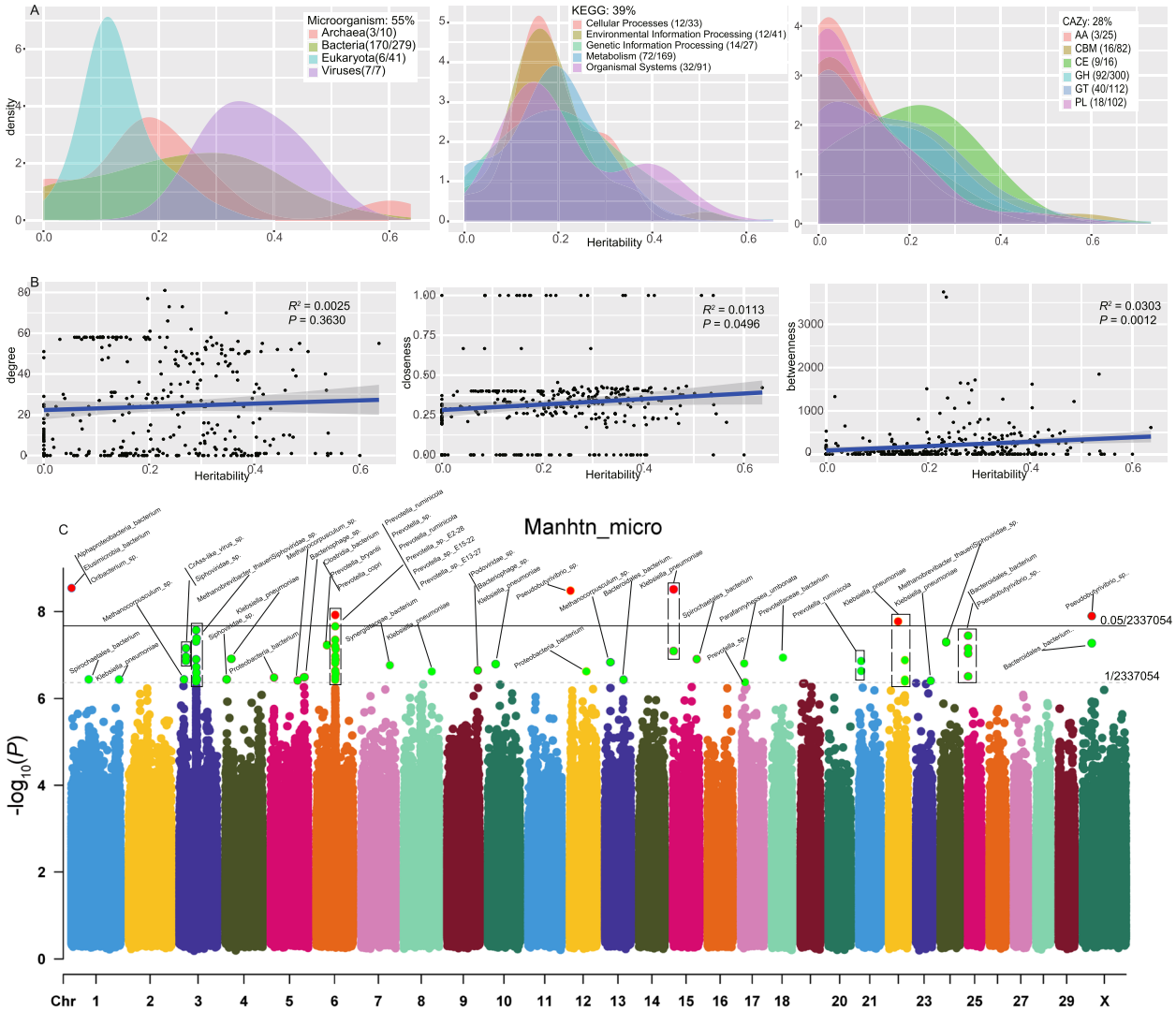

Fig. 3 The heritability and significant variants of the rumen microbiota of dairy cows. **A** The proportion of highly heritable microbes at the species level at the domain level, the proportion of highly heritable KEGG pathways at level 3 at level 1, and the proportion of highly heritable CAZy modules at the family level at the class level were calculated. **B** The linear relationship between node attributes (degree, closeness, and betweenness) and heritability. $P < 0.05$ was considered to indicate a linear relationship. **C** Manhattan plot of highly heritable rumen fat subsets from the top 100 microbes at the species level. The significance threshold was 1/nSNP = 4.28E-07. The extremely significant threshold was 0.05/nSNP = 2.14E-08

microbes ($h^2 \geq 0.2$) *Prevotella_sp.*, *Bacteroidaceae_bacterium*, and *Clostridia_bacterium* had the highest relative contributions to 116 pathways, while the less heritable microbes ($h^2 < 0.2$) *Tetrahymena_thermophila*, *archaeon*, *Muribaculaceae_bacterium*, *Treponema_sp.*, *Ruminococcus_sp.*, *Oscillospiraceae_bacterium*, *Methanobrevibacter_sp.*, *Candidatus_Saccharibacteria_bacterium*, and *Lachnospiraceae_bacterium* had the highest relative contribution to only 27 pathways (Fig. 4A). It can be seen that there were fewer highly heritable microbes involved in metabolism pathways, but their contribution for them was greater.

To further observe the role of highly heritable subsets of the rumen microbiota in the host phenotype, 5 microbes with the highest heritability (*Methanocorpusculum_sp.*, *Prevotella_mizrahii*, *Prevotella_multisaccharivorax*, *Parafannyhessea_umbonata*, *Pseudobutyrivibrio_sp.*) were considered a highly heritable latent variable (hh), and 5 microbes (*Clostridiales_bacterium*, *Firmicutes_bacterium*, *Methanosphaera_stadtmanae.*, *Blautia_sp.*, *Sarcina_sp.*) with the lowest heritability were considered a low heritable latent variable (lh) from rumen microbes at the species level with the top 100 relative abundances to explain rumen SCFA and ECM variations via SEM

(Fig. 4B). We found that hh variables, rather than lh variables, can increase the ECM by reducing the rumen A:P ratio (CFI = 0.898, RMSEA = 0.196).

Carbohydrate metabolism characteristics of highly heritable microbes

Highly heritable microbial subsets could decrease the rumen A:P ratio. Hence, we further focused on the Carbohydrate-Active Enzyme Database. At the class level, the GH ($h^2 = 0.21$) and GT ($h^2 = 0.21$) families had high heritability (Fig. 5A). Hence, the highly heritable CAZy modules from the GH and GT were focused on. For the highly heritable CAZy modules, 110 significant associations with the 31 modules were identified (Fig. S5B). These modules were annotated to the genes *ZFP90*, *ERC1*, *TMEM33*, *SLC30A9*, *RFTN1*, *HHIP*, *ARHGAP44*, *KCNJ12*, *5S_rRNA*, and *MYOM2* (-log10(*P*) > 6.39) (Fig. S5A). Among these, chr6:60,748,009 and chr17:13,553,198 had extremely significant impacts on highly heritable CAZy modules (-log10(*P*) > 7.67) (Table S8). The SNP variants (rs43470227 and rs43472732) related to the oligosaccharides-degradation in the CAZy module (GH67, GH13_38, GH95, GH43_10, GH115, and GH10) overlapped with SNP variants related to multiple *Prevotella* species (*Prevotella_sp.*, *Prevotella_sp._E15-22*, *Prevotella_sp._E13-27*) (Table S6 and S8). Moreover, *Prevotella* species (e.g., *Prevotella_sp.*) were the main microbes contributing to these CAZy module genes (Fig. S5C).

To provide a clearer explanation of the impact of highly heritable CAZy modules on rumen microbial metabolism, the rumen metabolome was examined (Table S9). A total of 2536 metabolites were detected in the rumen, 447 (18%) of which had high heritability ($h^2 \geq 0.2$) (Fig. 5B). The low, medium, and high groups of ECM were not significantly clustered separately in the OPLS-DA coordinate systems (Fig. S6A). Moreover, the Mantel test revealed that the highly heritable subset ($h^2 \geq 0.2$) rather than the less heritable subset ($h^2 < 0.2$) of the top 50 rumen metabolites was associated with ECM (Fig. S6B). Hence, we further focused on the Human Metabolome Database (HMDB) compound classification and enriched KEGG pathways of the top 50 highly heritable metabolites in the rumen. The most highly heritable metabolites were classified into "lipids and lipid-like molecules" (56.67%) (Figure S6C) and were significantly enriched in "linolenic acid metabolism," "alpha-linolenic acid metabolism," "cutin, suberine, and wax biosynthesis," and "arachidonic acid metabolism" (Fig. S6D). Next, 141 significant associations with the 30 highly heritable metabolites were identified (Table S10, Fig. S6F). These SNP variants related to metabolites were annotated to the genes *ZNF831*, *UTP15*, *SHC3*, *MROH8*, *MFSD4B*, *MANBAL*, *EDN3*, *DTYMK*, *C4A*, and *5s_rRNA* (-log10(*P*) > 6.39) (Fig. S6E).

Furthermore, we used the highly heritable CAZy modules from the GH and GT classes with the top 100 (31 modules) to associate with the rumen SCFAs, the highly heritable metabolites with the top 50 (16 metabolites). Except for GT92 and GT8, all the CAZy modules were negatively related to rumen A:P ratio and positively related to propionate ($P < 0.05$) but not related to acetate ($P > 0.05$) (Fig. 5C).

Finally, the enzymes involved in SCFA synthesis were focused on. Most of the highly heritable enzymes involved in "glycolysis/gluconeogenesis," "pyruvate metabolism," "butanoate metabolism," and "propanoate metabolism" were negatively correlated with the rumen A:P ratio (Table S11). The highly heritable enzymes involved in glycolysis/gluconeogenesis and the pyruvate pathway enhanced the synthesis of pyruvate. The highly heritable enzymes involved in "pyruvate metabolism," "butanoate metabolism," and "propanoate metabolism" enhanced the synthesis of pyruvic acid to propionate while weakening the synthesis of butyrate and acetate (Fig. 5D).

Heritable characteristics of the rumen microbiota related to rumen propionate and milk yield based on the mWAS

The highly heritable subsets of the rumen microbiota could increase the ECM by decreasing the rumen A:P ratio based on the characteristics of high-abundance microbes. To further verify these results, we conducted mWAS on 337 microbes with relative abundances exceeding 0.01% at the species level using the rumen A:P

(See figure on next page.)

Fig. 4 The characteristics of highly heritable rumen microbes of dairy cows. **A** The relationship between rumen microbes (with high and low heritability) and KEGG pathway enrichment at level 3 in "metabolism". The 20 highly heritable microbes and 30 with lowly heritable microbes within the top 50 with relative abundance were selected at the species level. The microbes were connected to metabolism pathways based on relative contribution (%). The color of the lines was determined based on relative contribution and heritability. Gray: contribution < 1%, blue: low heritable microbes had the highest contribution for pathways, red: high heritable microbes had the highest contribution for pathways. **B** The effect of highly heritable and weakly heritable microbes on the ECM determined by A:P via SEM. The 5 microbes with the highest and lowest heritability were selected based on heritability at the species level. Highly heritable microbes are integrated into highly heritable latent variables. Lowly heritable microbes are integrated into a lowly heritable latent variable. Red arrows represent positive paths, and green arrows represent negative paths

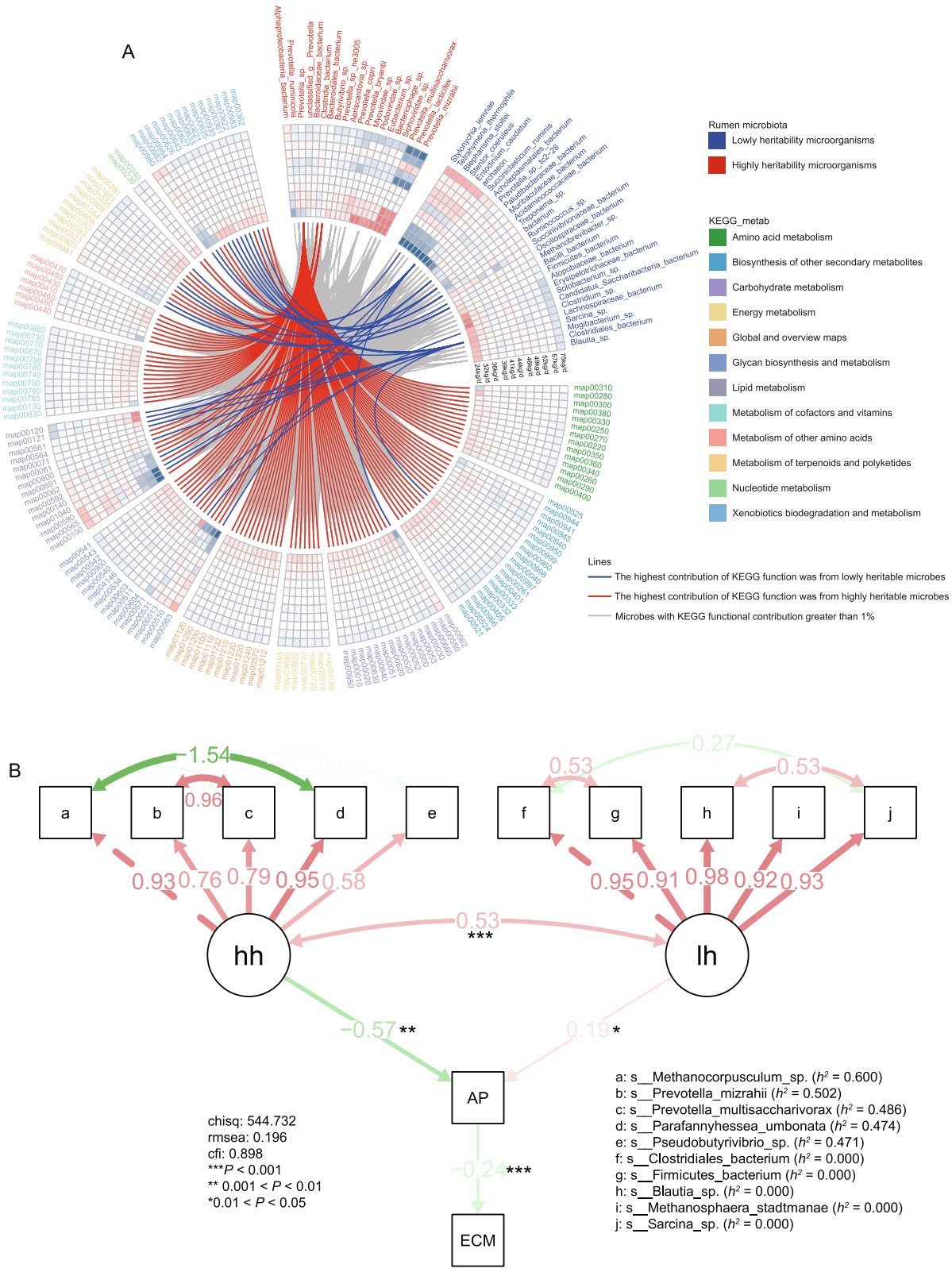

Fig. 4 (See legend on previous page.)

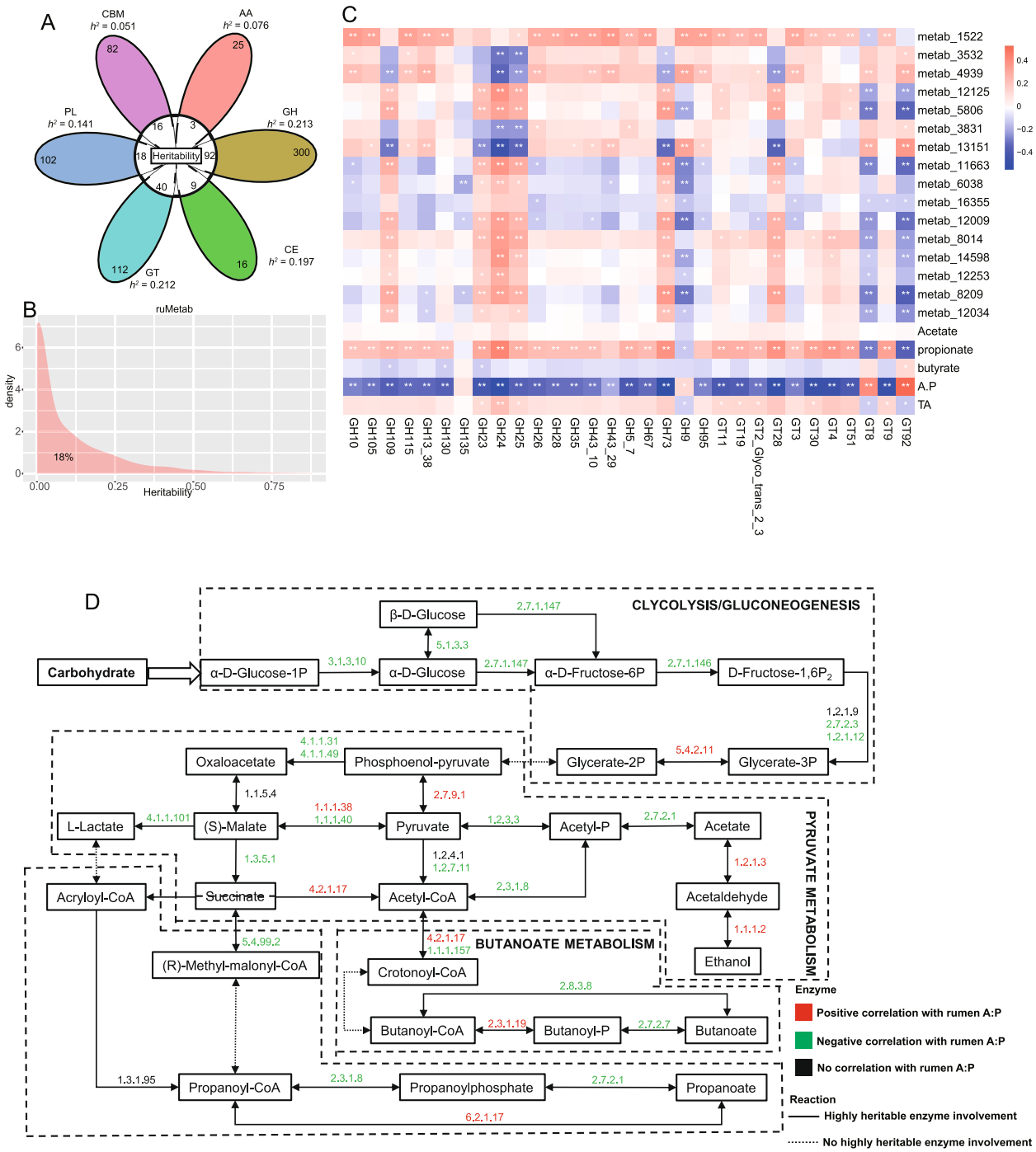

Fig. 5 The characteristics of highly heritable rumen enzymes and metabolites in dairy cows. **A** The heritability of CAZy modules at the class level. **B** The proportion of highly heritable metabolites in the rumen. **C** The relationship between highly heritable CAZy modules from the top 100 modules at the family level and highly heritable metabolites from the top 50 metabolites, SCFAs. **D** The relationships between high-heritability enzymes involved in "butanoate metabolism", "glycolysis/gluconeogenesis", "pentose phosphate pathway", "propionate metabolism", "pyruvate metabolism" and the rumen A:P ratio

ratio and ECM as the phenotypes (Fig. 6A, Table S2). We found that 252 microbes were significantly related to the ECM, and 296 microbes were significantly related to the rumen A:P ratio. Therein, 247 microbes were significantly related to the ECM and the rumen A:P ratio (marked species), accounting for 83% and 98%, respectively, of the rumen A:P ratio and ECM-related microbes (Fig. 6B). Moreover, compared with those of microbes (nonmarker) that were not related to the rumen A:P ratio or the ECM ($P > 0.05$), the proportions of highly heritable microbes among the marker species were greater (Fig. 6C).

In order to further verify the relationship between the microbes filtered by mWAS and rumen fermentation and lactation performance, we used the random forest algorithm to rank the importance of 247 marker species, and identified the top 20 biomarkers based on the mean decrease accuracy (MDA) index. For the ECM, *Rickettsiales_bacterium, Lachnospiraceae_bacterium_NK3A20, Olsenella_sp., [Clostridium]_aminophilum, Parameciu m_primaurelia, Sharpea_azabuensis, Prevotella_sp._AGR2160, Prevotella_sp._Rep29, Paramecium_tetraurelia, [Eubacterium]_cellulosolvens, Parabacteroides_merdae, Prevotella_sp._P5-126, Bacteroides_thetaiotaomicron, Selenomonas_bovis, Naegleria_fowleri, Pseudobutyrivibrio_sp., Oscillibacter_sp., Chryseobacterium_sp., Verrucomicrobia_bacterium,* and *Erysipelotrichaceae_bacterium* were key biomarkers that affect the lactation performance of dairy cows (Fig. S7A and Table S2). For rumen A:P ratio, *Sodaliphilus_pleomorphus, Candidatus_Methanomethylophilaceae_archaeon, Prevotella_bryantii, Deltaproteobacteria_bacterium, Veillonellaceae_bacterium, Prevotellaceae_bacterium, Acidaminococcus_fermentans, Clostridium_sp._28_17, Dialister_sp., Kiritimatiellae_bacterium, Clostridium_sp._CAG:793, Cyanobacteria_bacterium_UBA11991, Prevotella_mizrahii, Lachnoclostridium_sp., Prevotella_albensis, Succinivibrio_sp., Prevotella_histicola, Loktanella_sp., [Eubacterium]_cellulosolvens, Prevotella_sp._CAG:1092* were the key biomarkers affecting the rumen of dairy cows (Fig. S7B and Table S2). Moreover, biomarkers affecting the ECM and rumen A:P ratio were filtered by random forest, with high heritability microbes accounting for the majority (Fig. S7C).

Finally, the results of GWAS with lactation performance, rumen microbial compositions at the species level, rumen microbial CAZy module at the family level, and the rumen metabolome were integrated and found that chr25:11,177, located at *5s_RNA* genes, appeared in metagenome-GWAS signal sets of *Pseudobutyrivibrio* sp. (the marker species most strongly associated with rumen A:P and ECM based on mWAS) and GH24, metabolome-GWAS signal sets of the 9,10,13-TRIHOME, and lactation performance GWAS signal sets. Moreover, ECM, and the relative abundances of *Pseudobutyrivibrio* sp., GH24, 9,10,13-TRIHOME were significantly greater in the GG genotype than in the AG genotype at chr25:11,177 ($P < 0.05$) (Fig. S8A–D).

To further clarify this relationship, we first focused on the impact of GH24. We found that GH24 was related to 8 of the top 10 microbes with heritability, while it was related to only 1 of the bottom 10 microbes with heritability ($P < 0.05$, cor ≥ 0.6) (Fig. S8E). GH24 may be an important mediator of host regulating highly heritable microbes. By SEM analysis, GH24 was positively correlated with *Pseudobutyrivibrio* sp., which was positively correlated with 9,10,13-triHOME and subsequently positively correlated with ECM (CFI = 0.942, RMSEA = 0.162) (Fig. S8F).

Discussion

Lactation performance, as an important economic trait, has always received widespread attention. Previous studies have focused on the relationship between rumen microbial fermentation and lactation performance [33], especially under the same feeding management conditions. This indicates that the host's regulation for production performance may be mediated by rumen microbiota. Therefore, increasing research links host genes to the rumen microbes [14, 45, 46], but the role of highly heritable microbes in lactation performance has not been determined. Meanwhile, there is relatively little research on whether the host's influence on microbial composition will further reflect its impact on microbial function. Hence, we combined the rumen metagenome, rumen metabolome, and host genome data from 304 dairy cows with similar physiology and fed the same diet. This comprehensive approach allowed us to investigate the highly heritable aspects of rumen microbial composition, function, and metabolism (including SCFAs), and further

(See figure on next page.)

Fig. 6 Relationships among lactation performance, rumen microbial composition and function, and metabolites based on the mWAS. **A** Manhattan plot showing the microbes related to the ECM and the rumen A:P ratio based on microbiota-wide association studies (mWASs). **B** Venn diagram showing the proportion and quantity of microbes associated with the ECM and the rumen A:P ratio. **C** A density plot was generated to show the heritability of microbes associated with the ECM and the rumen A:P ratio. **D** A circular Manhattan plot showed that the same variant (chr25:11177 on 5 s_*rRNA*) of *Pseudobutyrivibrio* sp. (species-GWAS), GH24 (CAZy-GWAS), metab_11663 (9,10,13-TRIHOME) (metabolite-GWAS), and lactation performance (phenotype-GWAS)

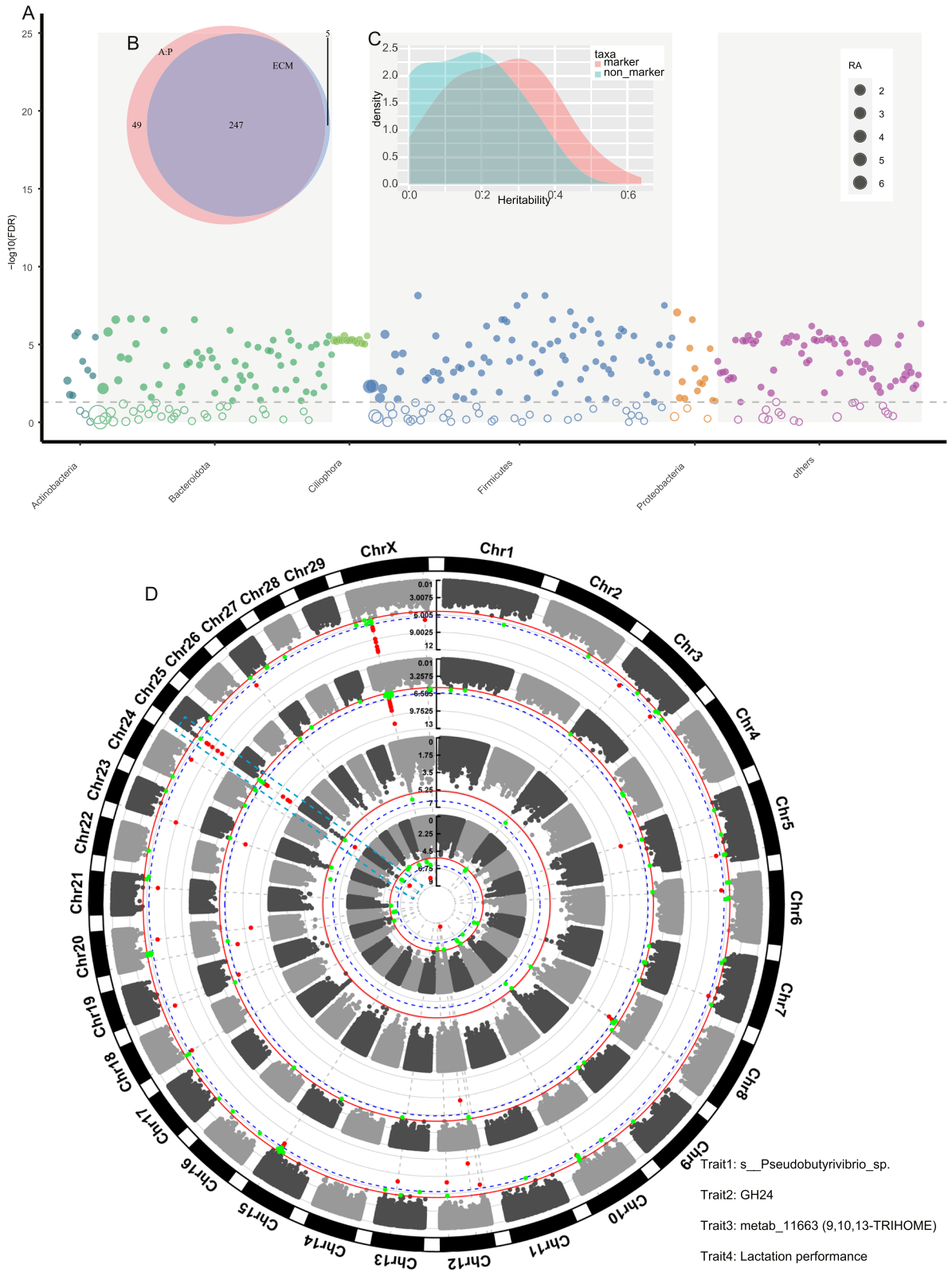

Fig. 6 (See legend on previous page.)

explore their impact on lactation performance. We found that (1) rumen fermentation is driven by microbes with high heritability rather than low heritability, which in turn affects the milk production performance of cows. Specifically, the high-heritability subset of the rumen microbiota can increase ECM production by reducing the rumen A:P ratio. Furthermore, highly heritable enzymes involved in the SCFA synthesis pathway can promote propionic fermentation. (2) Through GWAS analysis, we identified potential SNP variants that may affect cow milk production performance through the rumen microbiota. Specifically, significant rs43470227 and rs43472732 variants in *SLC30A9* regulated *Prevotella* species with oligosaccharides-degradation enzyme genes (Tables S6 and S8). The significant SNP variant chr25:11,177 on *5s_rRNA* was involved in the process by which *Pseudobutyrivibrio* enhances linoleic acid metabolism, thereby improving lactation performance.

With the development of metagenome and metabolome, research on the gastrointestinal microbiota is no longer limited to taxa level, with an increasing number of studies showing the importance of microbial function and metabolites. Hence, highly heritable rumen microbial composition, function, and metabolites are the focus of our study. Interestingly, the proportions of highly heritable subsets of the rumen microbial composition, function, and metabolites decreased sequentially. Furthermore, our results revealed a positive correlation between node attributes (betweenness and closeness) and heritability, which suggested that highly heritable rumen microbial subsets were located at hub ecological niches in co-occurrence networks. This phenomenon has also been reported in previous studies on the rumen microbiota [14, 46]. Notably, in the study, highly heritable rumen microbial subsets may play an important role in the host phenotype. Here, our results indicated that the rumen of cows with high lactation performance (ECM) had higher abundances of bacteria and viruses domain, "Metabolism" KEGG level 1, and GH family, which were also highly heritable variables. The rumen of cows with high milk protein yields was more abundant in *Prevotella* species [33]. In our study, we also found that multiple *Prevotella* species, such as *Prevotella_lacticifex*, *Prevotella_mizrahii*, and *Prevotella_multisaccharivorax*, were significantly correlated with rumen SCFAs and lactation performance according to the Mantel test, which were also highly heritable microbes. Prevotellaceae as one of the most abundant rumen core family can utilize dietary nutrients to produce SCFAs [47].

The function of microbes depends on their taxa [48]. Therefore, the host's influence on microbial composition may be further reflected in microbial function and metabolites. In the study, the GWAS analysis and heritability estimation of microbial functions are used to characterize the functional characteristics of highly heritable microbes, in order to explore whether the host-regulated microbes have functional similarities, thereby helping cows to efficiently utilize feed. The results revealed the significant contributions of highly heritable rumen microbes and their function in rumen fermentation. Notably, the conversion of glucose to acetate, propionate, and butyrate contributed 62%, 109%, and 78% of the original energy supply (2805 kJ/mol), respectively, to the overall energy provision for the body [49]. Therefore, rumen propionic fermentation is the most efficient way for ruminants to utilize energy. Here, we first focused on the role of host SNP variants in determining rumen fermentation type. The high heritability of rumen propionate and A:P ratio suggested that host genes may be involved in the process of rumen propionic fermentation. A study of broiler chickens revealed that the heritability of cecal propionate is less than 0.2, while that of butyrate is greater than 0.2 [50], which contradicted the findings of our study on the rumen. This difference suggested that the host and lumen may be key factors affecting lumen SCFA synthesis. Moreover, the partial SNP variants affecting rumen SCFAs and lactation performance were associated with the CDH (cadherin) gene in our study, which was significantly related to cell differentiation and rumen development [51, 52]. Rumen SCFAs are generated mainly from the metabolism of carbohydrates by rumen microbes. Here, highly heritable microbes not only participated in carbohydrate metabolism but also increased the ECM by reducing the A:P ratio of rumen SCFAs. However, lowly heritable subsets did not have this effect. The majority of highly heritable CAZy modules in the rumen were positively correlated with propionate but not with acetate. For example, GH13, which is mainly composed of α-amylase, can increase the synthesis of propionate [53]. In the pyruvate metabolism and SCFA synthesis pathways, highly heritable microbial enzymes can enhance the synthesis of propionate and weaken the synthesis of acetate and butyrate. The above phenomena indicate that highly heritable microbes and their functions can promote rumen propionic fermentation.

In this study, a locus overlap approach (with the same SNP variants) was used to determine whether host SNP variants could affect the rumen microbes with specific enzyme-encoding genes. Multiple *Prevotella* species (e.g., *Prevotella_sp.*) significantly related to the ECM had the same significant SNP variants (chr6:6:60,860,355 (rs43470227) and chr6:60,948,378 (rs43472732), *SLC30A9*) with multiple GH families (GH67, GH13_38, GH95, GH43_10, GH115, and GH10). Oligosaccharide-degradation enzymes from the GH10, GH13, GH43, GH95, and GH115 families have xylosidase, mannosidase, or

glucosidase activity and produced SCFAs [54, 55]. The *SLC30A9* gene belongs to the solute carrier family and is involved in zinc ion homeostasis and zinc ion transport. Zn^{2+}, an essential micronutrient and key cofactor in microbial metabolism, could significantly enhance the growth and metabolic capacity of microbes in vitro [56]. For lactating cows, a balance of Zn^{2+} intake could enhance the degradation of fibre and promote rumen SCFA production, bacterial growth, and microbial protein biosynthesis [57]. Moreover, Zn^{2+} was also the key cofactor for the synthesis of microbial enzymes [58]. Hence, *SLC30A9* may play an important role in the host regulation of the rumen microbes with specific enzyme genes, but this possibility requires well-designed experimental verification, such as gene knockout experiments.

Finally, information from GWASs, m-GWASs, M-GWASs, and mWASs were integrated to determine a host-microbe interaction for microbiota- CAZy- metabolism-lactation performance. First, consistent with the findings of previous GWASs for dairy cows, lactation performance was also a highly heritable trait in our research ($h^2 \geq 0.2$) [5]. Our study revealed the same genes that affected lactation performance as did previous GWASs for reproductive and production performance in cows, such as *TSPAN9* [59] on chr5 and *5s_rRNA* on chr25 [59–64]. The same SNP variants were found in the microbiome and phenotype. Next, we found that chr25:11,177 on *5s_rRNA* appeared in the association analysis for *Pseudobutyrivibrio_sp.*, GH24, 9,10,13-TRIHOME, and lactation performance. *5s_rRNA* is a small RNA with a length of 120 nt that has highly conserved secondary and tertiary structures in both prokaryotes and eukaryotes. *5S_rRNA* not only participates in protein translation regulation but is also associated with ribosome production. The liver undergoes metabolic reprogramming and activates the tumor suppressor p53, which subsequently results in decreased or aberrant ribosome production. This, in turn, leads to the reprogramming of cellular transcription [65]. Because the rumen microbiome and metabolome are located at chr25:11,177 (*5s_rRNA*), *Pseudobutyrivibrio* possesses the capability to produce a diverse array of hydrolytic enzymes, facilitating efficient digestion of forage. Several strains of *Pseudobutyrivibrio* (i.e., *Pseudobutyrivibrio xylanivorans Mz5T*) produce multiple xylanases that once accounted for the highest xylanolytic activity among the rumen bacteria tested thus far [66]. Moreover, *Pseudobutyrivibrio* species possess several extraordinary characteristics (i.e., active hydrolases, bacteriocin, and conjugated linoleic acid production), which make them probiotic for animals [67]. Interestingly, 9,10,13-TRIHOME is a "lipid and lipid-like molecule" metabolite involved in "linoleic acid metabolism." Linoleic acid is essential for normal brain development by influencing neurogenesis and synapse formation [68]. For lactating cows, supplementation with linoleic acid in the diet had a positive effect on maintaining energy balance and alleviating liver metabolic stress during lactation [69]. Moreover, linoleic acid intake could be related to animal characteristics (i.e., weight), as indicated by the presence of linoleic acid in milk fat [70]. GH24 contains lysozyme, which inhibits methane production and enhances fermentation in vitro in rumen fermentation experiments [71]. Here, we speculated that lysozyme may enhance the competitiveness of *Pseudobutyrivibrio* in the rumen microbiota, but we have not found any evidence of a relationship between lysozyme and *Pseudobutyrivibrio*. In summary, we suggest that, under the action of lysozyme (GH24), *Pseudobutyrivibrio* species increase the abundance of metabolites (9,10,13-TRIHOME) involved in linoleic acid metabolism and improve lactation performance while gaining niche advantages of *Pseudobutyrivibrio* in the rumen microbiota. The SNP variant chr25:11,177 on *5s_rRNA* is involved in this process. The above results provide a potential biological process for explaining the impact of *5s_rRNA* on cow production performance discovered in previous GWAS analyses related to lactation performance of dairy cow in other region [59].

There are several limitations to the present study. First, though a host-microbe interaction that was influenced by host SNP variants was identified in the present study, this is a limited cohort of cows with a focused sampling of the same herd of cows at a relatively centralized time point, which must be taken into account when interpreting the data. However, the consistency in the feeding and management conditions of the sampling cows can avoid the influences of environment and feeds on the ruminal microbiome, and better highlight the host genetic role in shaping rumen microbiota [46]. Second, our findings mainly focused on the metagenome and host genetic levels. Considering gene expression was a crucial step for the functional execution of microbes and host, future studies should pay more attention to the host and microbial transcriptome results, which can provide more information about the host-microbes interaction between the transcriptome of host and the functional and expressed microbial genes. To sum up, the effects of the host SNP variants on rumen microbiome needed further validation in a larger cohort of cows, so that these selected SNPs can be well used for breeding and selection of dairy cows. Notably, the genes annotated by multiple SNP variants related to microbes in our study were significantly related to the lactation performance and efficiency of cows. These genes included *5s_rRNA*, which is associated with lactation persistence [59], *DIAPH3* and *TSPAN11*, which are associated with feed efficiency [72, 73], and *UGGT2*, which is associated with milk fat yield [74]. That is to

say, these genes identified in our research may affect the production performance of cows by influencing rumen microbiota. Further, due to the formation of rumen microbiome, which may be proved to be affected by the host genetics SNPs variants, and further considering the CRISPR-Cas9 mediated gene editing breeding technology was well developed, our selected SNPs can be tested in the future study to verify the impact of host SNPs variants on ruminal microbiome formation.

Conclusion

In this study, we used lactation performance and the rumen SCFAs of 304 lactating cows as a phenotype and evaluated the correlation between rumen metagenomes and metabolomes and host resequencing. Our findings highlight the crucial involvement of highly heritable rumen microbial composition and function in the modulation of rumen fermentation patterns, with consequential impacts on the efficiency of ECM production. Furthermore, we discovered a host-microbe interaction that was influenced by host SNP variants, wherein the rumen microbiome composition, function, and metabolites exert a significant impact on cows' lactation performance. Specifically, our results showed that the influence of chr25:11,177 (*5s_rRNA*) on lactation performance may be mediated through the enhanced niche advantages of *Pseudobutyrivibrio* within the rumen microbiota, facilitated by the action of GH24 (lysozyme). This, in turn, leads to an increase in metabolites involved in linoleic acid metabolism (9,10,13-TRIHOME). In a meticulously controlled large-scale population with carefully regulated environment and diet, our study uniquely integrated the variations in composition, function, and metabolism of rumen microbiota with host SNP variations. This comprehensive approach allowed us to unravel the intricate interplay between the host and microbiota, revealing how dairy cows actively shape and select their rumen microbiota to regulate lactation performance. These findings establish a direct connection between the nutrition and genetics of dairy cows through the mediation of rumen microbiota, thereby providing a solid theoretical foundation for precision nutrition strategies in modern farming practices.

Abbreviations

A:P	Acetate and propionate ratio
BWA	Burrows-Wheeler-Alignment Tool
CAZy	Carbohydrate-Active enZYmes
CFI	Comparative fit index
chisq	Chi-squared test
CLR	Centered log-ratio transformation
ECM	Energy-corrected milk
GATK	Genome Analysis Toolkit
GRM	Genetic relationship matrix
GREML	Genome-based restricted maximum likelihood
GWAS	Genome-wide association studies
h^2	Heritability
hh	Highly heritable latent variable
KEGG	Kyoto Encyclopedia of Genes and Genomes
HMDB	The Human Metabolome Database
LC-MS	Liquid chromatography–mass spectrometry
lh	Lowly heritable latent variable
MDA	Mean decrease accuracy
MF	Milk fat
m-GWAS	Metabolome-genome-wide association studies
m-GWAS	Microbiome-genome-wide association studies
ML	Milk lactose
MLM	Mixed linear model
MP	Milk protein
mWAS	Microbiota-wide association studies
MY	Milk yield
OPLS-DA	Orthogonal partial least-squares discriminant analysis
ORF	Open reading frames
RMSEA	Root-mean-square error of approximation
SCFAs	Short-chain fatty acids
SEM	Structural equation model
SNPs	Single nucleotide polymorphisms
TA	Total acid

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01937-3.

Additional file 1: Fig. S1. Data characteristics of phenotype and their relationships. Fig. S2. The differences of rumen microbial diversity and compositions at species level among low, medium, and high group of ECM. Fig. S3. Individual genetic information of 304 dairy cows. Fig. S4. The heritability and genetic information of rumen microbiota. Fig. S5. The variants information of highly heritable CAZy modules at family level by GWAS. Fig. S6. The characteristic and variant information of rumen highly heritable metabolites. Fig. S7. The relationships among ECM, GH24, Pseudobutyrivibrio , and 9,10,13-TriHOME.

Additional file 2: Table S1. Diet and Animal cohort information. Table S2. The information (classification, relative abundance, heritability, node attribute, and FDR for A:P and ECM) of rumen microbes with relative abundance exceeding 0.01% at species level of 304 dairy cows. Table S3. GWAS for lactation performance and rumen SCFA. Table S4. The information (classification, relative abundance, and heritability) of rumen microbial pathways at KEGG level3 of 304 dairy cows. Table S5. The information (classification, relative abundance, and heritability) of CAZy modules at family level of 304 dairy cows. Table S6. GWAS for rumen highly heritability subsets from top 100 microbes at species level. Table S7. The relative contribution (%) matrix between lowly, highly heritable microbes and KEGG pathways of "Metabolism." Table S8. GWAS for rumen highly heritability subsets from top 100 CAZy module at family level. Table S9. The information (classification, relative quantitation, heritability) of rumen microbes with relative abundance exceeding 0.01% at species level of 304 dairy cows. Table S10. GWAS for rumen highly heritability subsets from top 50 metabolites. Table S11. The heritability of enzymes involved in metabolic pathways and their relationship with rumen A:P.

Acknowledgements

The authors would like to express our appreciation to all the members of the Leyuan Animal Husbandry for their help in experimental farm. And we also thank the High Performance Computing Platform of Northwest A&F University.

Authors' contributions

Conception and design: JY, SW, CZ. Sample collection: CZ, HL, XJ, ZZ, and HX. Development of methodology: CZ, SW, JY. Acquisition of data: CZ. Analysis and interpretation of the data: CZ, SW, SH. Manuscript writing and revision: CZ, SW, and SH. Review of the manuscript: All the authors. Lead contact author: SRW. The author(s) read and approved the final manuscript.

Funding
The raw sequencing data used and described in this study have been deposited into CNGB Sequence Archive (CNSA) (https://db.cngb.org/cnsa/) of China National GeneBank DataBase (CNGBdb) with accession number CNP0005323 (Metagenome data), CNP0005324 (whole-genome resequencing data), and CNP0005479 (Metabolome data). All data have now been publicly available since 21st March 2024.
The private link that the reviewers can use to access data is provided as follows:
Metagenome data:
http://db.cngb.org/cnsa/project/CNP0005323_2afb08a6/reviewlink/
Whole-genome resequencing data:
http://db.cngb.org/cnsa/project/CNP0005324_1354668f/reviewlink/
Metabolome data:
https://db.cngb.org/cnsa/project/CNP0005479_d4c88f7b/reviewlink/
All information is included in the manuscript or supporting files.

Data availability
The readers can contact the corresponding authors as needed to request raw data.
The raw sequencing data used and described in this study have been deposited into CNGB Sequence Archive (CNSA) (https://db.cngb.org/cnsa/) of China National GeneBank DataBase (CNGBdb) with accession number CNP0005323 (Metagenome data, link: https://db.cngb.org/search/project/CNP0005323/), CNP0005324 (Whole genome resequencing data, link: https://db.cngb.org/search/project/CNP0005324/), and CNP0005479 (Metabolome data, link: https://db.cngb.org/search/project/CNP0005479/). All information is included in the manuscript or supporting files.

Declarations

Ethics approval and consent to participate
This experiment was conducted at the Animal Research and Technology Centre of Northwest A&F University (Yangling, Shaanxi, China). All analyses were performed in accordance with the guidelines recommended by the Administration of Affairs Concerning Experimental Animals (Ministry of Science and Technology, China, revised 2004). The protocol was approved by the Institutional Animal Care and Use Committee of Northwest A&F University.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Animal Science and Technology, Northwest A&F University, 22 Nt, Xinong Road, Yangling, Shaanxi, China. [2]Institute of Global Food Security, School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Belfast, Northern Ireland BT9 5DL, UK. [3]Key Laboratory of Livestock Biology, Northwest A&F University, 22 Nt, Xinong Road, Yangling, Shaanxi, China. [4]JUNLEBAO-Northwest A&F University Cooperation Dairy Research Institute, Leyuan Animal Husbandry, JUNLEBAO Company, Shijiazhuang, Hebei, China.

Received: 8 March 2024 Accepted: 20 September 2024
Published online: 12 November 2024

References

1. FAO. Food Outlook Biannual report on global food markets. Rome: Food and Agriculture Organization of the United Nations; 2022. https://doi.org/10.4060/cb9427en.
2. Rask KJ, Rask N. Economic development and food production–consumption balance: a growing global challenge. Food Policy. 2011;36(2):186–96. https://doi.org/10.1016/j.foodpol.2010.11.015.
3. Tricarico JM, Kebreab E, Wattiaux MA. MILK symposium review: sustainability of dairy production and consumption in low-income countries with emphasis on productivity and environmental impact*. J Dairy Sci. 2020;103(11):9791–802. https://doi.org/10.3168/jds.2020-18269.
4. Xue M-Y, Xie Y-Y, Zhong Y, Ma X-J, Sun H-Z, Liu J-X. Integrated meta-omics reveals new ruminal microbial features associated with feed efficiency in dairy cattle. Microbiome. 2022;10(1):32. https://doi.org/10.1186/s40168-022-01228-9.
5. Schennink A, Stoop WM, Visker MHPW, Heck JML, Bovenhuis H, Van Der Poel JJ, Van Valenberg HJF, Van Arendonk JAM. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Animal Genet. 2007;38(5):467–73. https://doi.org/10.1111/j.1365-2052.2007.01635.x.
6. Wang T, Li J, Gao X, Song W, Chen C, Yao D, Ma J, Xu L, Ma Y: Genome-wide association study of milk components in Chinese Holstein cows using single nucleotide polymorphism. Livestock Science. 2020;233103951. https://doi.org/10.1016/j.livsci.2020.103951.
7. Coppieters W, Riquet J, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M. A QTL with major effect on milk yield and composition maps to bovine Chromosome 14. Mamm Genome. 1998;9(7):540–4. https://doi.org/10.1007/s003359900815.
8. Heyen DW, Weller JI, Ron M, Band M, Beever JE, Feldmesser E, Da Y, Wiggans GR, VanRaden PM, Lewin HA. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol Genomics. 1999;1(3):165–75. https://doi.org/10.1152/physiolgenomics.1999.1.3.165.
9. Conte G, Mele M, Chessa S, Castiglioni B, Serra A, Pagnacco G, Secchiari P. Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian rown cattle. J Dairy Sci. 2010;93(2):753–63. https://doi.org/10.3168/jds.2009-2581.
10. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 2002;12(2):222–31. https://doi.org/10.1101/gr.224202.
11. Yang H, Wu J, Huang X, Zhou Y, Zhang Y, Liu M, Liu Q, Ke S, He M, Fu H, Fang S, Xiong X, Jiang H, Chen Z, Wu Z, Gong H, Tong X, Huang Y, Ma J, Gao J, Charlier C, Coppieters W, Shagam L, Zhang Z, Ai H, Yang B, Georges M, Chen C, Huang L. ABO genotype alters the gut microbiota by regulating GalNAc levels in pigs. Nature. 2022;606(7913):358–67. https://doi.org/10.1038/s41586-022-04769-z.
12. Zhernakova DV, Wang D, Liu L, Andreu-Sánchez S, Zhang Y, Ruiz-Moreno AJ, Peng H, Plomp N, Del Castillo-Izquierdo Á, Gacesa R, Lopera-Maya EA, Temba GS, Kullaya VI, van Leeuwen SS, Aguirre-Gamboa R, Deelen P, Franke L, Kuivenhoven JA, Nolte IM, Sanna S, Snieder H, Swertz MA, Visscher PM, Vonk JM, Xavier RJ, de Mast Q, Joosten LAB, Riksen NP, Rutten JHW, Netea MG, Sanna S, Wijmenga C, Weersma RK, Zhernakova A, Harmsen HJM, Fu J, Lifelines Cohort S. Host genetic regulation of human gut microbial structural variation. Nature. 2024;625(7996):813–21. https://doi.org/10.1038/s41586-023-06893-w.
13. Zhang C, Liu H, Sun L, Wang Y, Chen X, Du J, Sjöling Å, Yao J, Wu S. An overview of host-derived molecules that interact with gut microbiota. iMeta. 2023;2(2):e88. https://doi.org/10.1002/imt2.88.
14. Wallace RJ, Sasson G, Garnsworthy PC, Tapio I, Gregson E, Bani P, Huhtanen P, Bayat AR, Strozzi F, Biscarini F, Snelling TJ, Saunders N, Potterton SL, Craigon J, Minuti A, Trevisi E, Callegari ML, Cappelli FP, Cabezas-Garcia EH, Vilkki J, Pinares-Patino C, Fliegerová KO, Mrázek J, Sechovcová H, Kopečný J, Bonin A, Boyer F, Taberlet P, Kokou F, Halperin E, Williams JL, Shingfield KJ. Mizrahi I: a heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. Sci Adv. 2019;5(7):eaav8391. https://doi.org/10.1126/sciadv.aav8391.
15. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D, Clark AG. Host genetic variation impacts microbiome composition across human body sites. Genome Biol. 2015;16(1):191. https://doi.org/10.1186/s13059-015-0759-1.
16. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova DV, Bonder MJ, Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon J-Y, Kim H-N, Lüll K, Barkan E, Shah SA, Fornage M, Szopinska-Tokov J, Wallen ZD, Borisevich D, Agreus L, Andreasson A, Bang C, Bedrani L, Bell JT, Bisgaard H, Boehnke M, Boomsma DI, Burk RD, Claringbould A, Croitoru K, Davies GE, van Duijn CM, Duijts L, Falony G, Fu J, van der Graaf A, Hansen T, Homuth G, Hughes DA, Ijzerman RG, Jackson MA, Jaddoe VWV, Joossens M, Jørgensen T, Keszthelyi D, Knight R, Laakso M, Laudes M, Launer LJ, Lieb W, Lusis AJ, Masclee AAM, Moll HA, Mujagic Z, Qibin Q, Rothschild

D, Shin H, Sørensen SJ, Steves CJ, Thorsen J, Timpson NJ, Tito RY, Vieira-Silva S, Völker U, Völzke H, Võsa U, Wade KH, Walter S, Watanabe K, Weiss S, Weiss FU, Weissbrod O, Westra H-J, Willemsen G, Payami H, Jonkers DMAE, Arias Vasquez A, de Geus EJC, Meyer KA, Stokholm J, Segal E, Org E, Wijmenga C, Kim H-L, Kaplan RC, Spector TD, Uitterlinden AG, Rivadeneira F, Franke A, Lerch MM, Franke L, Sanna S, D'Amato M, Pedersen O, Paterson AD, Kraaij R, Raes J, Zhernakova A. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156–65. https://doi.org/10.1038/s41588-020-00763-1.
17. Xu F, Fu Y, Sun T-y, Jiang Z, Miao Z, Shuai M, Gou W, Ling C-w, Yang J, Wang J, Chen Y-m, Zheng J-S. The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for complex human diseases. Microbiome. 2020;8(1):145. https://doi.org/10.1186/s40168-020-00923-9.
18. Golder HM, Thomson J, Rehberger J, Smith AH, Block E, Lean IJ. Associations among the genome, rumen metabolome, ruminal bacteria, and milk production in early-lactation Holsteins. Journal of Dairy Science. 2023;106(5):3176–91. https://doi.org/10.3168/jds.2022-22573.
19. Zang X-W, Sun H-Z, Xue M-Y, Zhang Z, Plastow G, Yang T, Guan LL, Liu J-X, Metcalf JL: Heritable and nonheritable rumen bacteria are associated with different characters of lactation performance of dairy cows. mSystems. 2022;7(5):https://doi.org/10.1128/msystems.00422-22.
20. Weimer PJ. Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. Front Microbiol. 2015;6https://doi.org/10.3389/fmicb.2015.00296.
21. Li C, Xue Y, Han M, Palmer LC, Rogers JA, Huang Y, Stupp SI. Synergistic photoactuation of bilayered spiropyran hydrogels for predictable origami-like shape change. Matter. 2021;4(4):1377–90. https://doi.org/10.1016/j.matt.2021.01.016.
22. Wu S, Cui Z, Chen X, Zheng L, Ren H, Wang D, Yao J. Diet-ruminal microbiome-host crosstalk contributes to differential effects of calf starter and alfalfa hay on rumen epithelial development and pancreatic α-amylase activity in yak calves. Journal of Dairy Science. 2021;104(4):4326–40. https://doi.org/10.3168/jds.2020-18736.
23. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques. 2004;36(5):808–12. https://doi.org/10.2144/04365st04.
24. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.
25. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
26. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6. https://doi.org/10.1093/bioinformatics/btv033.
27. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1): 119. https://doi.org/10.1186/1471-2105-11-119.
28. Noguchi H, Park JA, Takagi T: MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Research. 2006;345623-5630.
29. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–2. https://doi.org/10.1093/bioinformatics/bts565.
30. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. Bioinformatics. 2008;24(5):713–4. https://doi.org/10.1093/bioinformatics/btn025.
31. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60. https://doi.org/10.1038/nmeth.3176.
32. Ofek-Lalzar M, Sela N, Goldman-Voronov M, Green SJ, Hadar Y, Minz D. Niche and host-associated functional signatures of the root surface microbiome. Nat Commun. 2014;5(1):4950. https://doi.org/10.1038/ncomms5950.
33. Xue M-Y, Sun H-Z, Wu X-H, Liu J-X, Guan LL. Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. Microbiome. 2020;8(1):64. https://doi.org/10.1186/s40168-020-00819-8.
34. Paropkari AD, Leblebicioglu B, Christian LM, Kumar PS. Smoking, pregnancy and the subgingival microbiome. Sci Rep. 2016;6(1): 30388. https://doi.org/10.1038/srep30388.
35. Walker MA, Pedamallu CS, Ojesina AI, Bullman S, Sharpe T, Whelan CW, Meyerson M. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. Bioinformatics. 2018;34(24):4287–9. https://doi.org/10.1093/bioinformatics/bty501.
36. Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE. Cross-species comparisons of host genetic associations with the microbiome. Science. 2016;352(6285):532–5. https://doi.org/10.1126/science.aad9379.
37. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 2011;88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
38. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–4. https://doi.org/10.1038/ng.2310.
39. Wen T, Xie P, Yang S, Niu G, Liu X, Ding Z, Xue C, Liu Y-X, Shen Q, Yuan J. ggClusterNet: an R package for microbiome network analysis and modularity-based multiple network layouts. iMeta. 2022;1(3):e32. https://doi.org/10.1002/imt2.32.
40. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Boss E, Bowler C, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sieracki M, Velayoudon D, coordinators TO. Structure and function of the global ocean microbiome. Science. 2015;348(6237):1261359. https://doi.org/10.1126/science.1261359.
41. Cheung MW-L. metaSEM: an R package for meta-analysis using structural equation modeling. Front Psychol. 2015;5. https://doi.org/10.3389/fpsyg.2014.01521.
42. Wen C, Yan W, Mai C, Duan Z, Zheng J, Sun C, Yang N. Joint contributions of the gut microbiota and host genetics to feed efficiency in chickens. Microbiome. 2021;9(1):126. https://doi.org/10.1186/s40168-021-01040-x.
43. Fu J, Bonder MJ, Cenit MC, Tigchelaar EF, Maatman A, Dekens JAM, Brandsma E, Marczynska J, Imhann F, Weersma RK, Franke L, Poon TW, Xavier RJ, Gevers D, Hofker MH, Wijmenga C, Zhernakova A. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circulation Research. 2015;117(9):817–24. https://doi.org/10.1161/CIRCRESAHA.115.306807.
44. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri WA Jr, Ahmed T, Gordon JI. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature. 2014;510(7505):417–21. https://doi.org/10.1038/nature13421.
45. Wang W, Zhang Y, Zhang X, Li C, Yuan L, Zhang D, Zhao Y, Li X, Cheng J, Lin C, Zhao L, Wang J, Xu D, Yue X, Li W, Wen X, Jiang Z, Ding X, Salekdeh GH, Li F. Heritability and recursive influence of host genetics on the rumen microbiota drive body weight variance in male Hu sheep lambs. Microbiome. 2023;11(1):197. https://doi.org/10.1186/s40168-023-01642-7.
46. Li F, Li C, Chen Y, Liu J, Zhang C, Irving B, Fitzsimmons C, Plastow G, Guan LL. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. Microbiome. 2019;7(1):92. https://doi.org/10.1186/s40168-019-0699-1.
47. Stevenson DM, Weimer PJ. Dominance of revotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Appl Microbiol Biotechnol. 2007;75(1):165–74. https://doi.org/10.1007/s00253-006-0802-y.
48. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018;16(9):567–76. https://doi.org/10.1038/s41579-018-0024-1.
49. Ryle MER. Energy Nutrition in Ruminants. Springer Netherlands; 1990. https://doi.org/10.1007/978-94-009-0751-5.
50. He Z, Liu R, Wang M, Wang Q, Zheng J, Ding J, Wen J, Fahey AG, Zhao G. Combined effect of microbially derived cecal SCFA and host genetics on

feed efficiency in broiler chickens. Microbiome. 2023;11(1):198. https://doi.org/10.1186/s40168-023-01627-6.
51. Gao Y, Fang L, Baldwin RL, Connor EE, Cole JB, Van Tassell CP, Ma L, Li C-j, Liu GE. Single-cell transcriptomic analyses of dairy cattle ruminal epithelial cells during weaning. Genomics. 2021;113(4):2045–55. https://doi.org/10.1016/j.ygeno.2021.04.039.
52. Lin S, Fang L, Kang X, Liu S, Liu M, Connor EE, Baldwin RL, Liu G, Li C-J. Establishment and transcriptomic analyses of a cattle rumen epithelial primary cells (REPC) culture by bulk and single-cell RNA sequencing to elucidate interactions of butyrate and rumen development. Heliyon. 2020;6(6): e04112. https://doi.org/10.1016/j.heliyon.2020.e04112.
53. Li QS, Wang R, Ma ZY, Zhang XM, Jiao JZ, Zhang ZG, Ungerfeld EM, Yi KL, Zhang BZ, Long L, Long Y, Tao Y, Huang T, Greening C, Tan ZL, Wang M. Dietary selection of metabolically distinct microorganisms drives hydrogen metabolism in ruminants. ISME J. 2022;16(11):2535–46. https://doi.org/10.1038/s41396-022-01294-9.
54. Gharechahi J, Vahidi MF, Sharifi G, Ariaeenejad S, Ding X-Z, Han J-L, Salekdeh GH. Lignocellulose degradation by rumen bacterial communities: New insights from metagenome analyses. Environ Res. 2023;229115925. https://doi.org/10.1016/j.envres.2023.115925.
55. Kaneko S, Fujimoto Z. α-l-Rhamnosidases: Structures, substrate specificities, and their applications. chapter 16. 2023:349–64. https://doi.org/10.1016/B978-0-323-91805-3.00019-8.
56. Coverdale JPC, Khazaipoul S, Arya S, Stewart AJ, Blindauer CA. Crosstalk between zinc and free fatty acids in plasma. Biochim Biophys Acta (BBA) - Mol Cell Biol Lipids. 2019;1864(4):532–42. https://doi.org/10.1016/j.bbalip.2018.09.007.
57. Wang RL, Liang JG, Lu L, Zhang LY, Li SF, Luo XG. Effect of zinc source on performance, zinc status, immune response, and rumen fermentation of lactating cows. Biol Trace Elem Res. 2013;152(1):16–24. https://doi.org/10.1007/s12011-012-9585-4.
58. Hussain S, Khan M, Sheikh TMM, Mumtaz MZ, Chohan TA, Shamim S, Liu Y: Zinc essentiality, toxicity, and its bacterial bioremediation: a comprehensive insight. Front Microbiol. 2022;13https://doi.org/10.3389/fmicb.2022.900740.
59. Do DN, Bissonnette N, Lacasse P, Miglior F, Sargolzaei M, Zhao X, Ibeagha-Awemu EM. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. Journal of Dairy Science. 2017;100(3):1955–70. https://doi.org/10.3168/jds.2016-11910.
60. Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S, Oh SJ, Lee H-K, Kim H. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mamm Genome. 2017;28(11):528–41. https://doi.org/10.1007/s00335-017-9715-6.
61. Marina H, Pelayo R, Suárez-Vega A, Gutiérrez-Gil B, Esteban-Blanco C, Arranz JJ. Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in Assaf and Churra dairy breeds. Journal of Dairy Science. 2021;104(11):11850–66. https://doi.org/10.3168/jds.2021-20510.
62. Dias EAR, Campanholi SP, Rossi GF, Freitas Dell'Aqua CdP, Dell'Aqua JA, Papa FO, Zorzetto MF, de Paz CCP, Oliveira LZ, Mercadante MEZ, Monteiro FM. Evaluation of cooling and freezing systems of bovine semen. Animal Reprod Sci. 2018;195102–111. https://doi.org/10.1016/j.anireprosci.2018.05.012.
63. Ahmad SF, Singh A, Gangwar M, Kumar S, Dutt T, Kumar A: Haplotype-based association study of production and reproduction traits in multi-generational Vrindavani population. Gene. 2023;867147365. https://doi.org/10.1016/j.gene.2023.147365.
64. Sousa Junior LPB, Pinto LFB, Cruz VAR, Junior GAO, Oliveira HR, Chud TS, Pedrosa VB, Miglior F, Schenkel FS, Brito LF. Genome-wide association and functional genomic analyses for various hoof health traits in North American Holstein cattle. J Dairy Sci. 2023. https://doi.org/10.3168/jds.2023-23806.
65. Pelava A, Schneider C, Watkins Nicholas J. The importance of ribosome production, and the 5S RNP–MDM2 pathway, in health and disease. Biochem Soc Trans. 2016;44(4):1086–90. https://doi.org/10.1042/bst20160106.
66. Tripathi MK, Roy U, Jinwal UK, Jain SK, Roy PK. Cloning, sequencing and structural features of a novel Streptococcus lipase. Enzyme and Microbial Technology. 2004;34(5):437–45. https://doi.org/10.1016/j.enzmictec.2003.11.020.
67. Grilli DJ, Mansilla ME, Giménez MC, Sohaefer N, Ruiz MS, Terebiznik MR, Sosa M, Arenas GN: Pseudobutyrivibrio xylanivorans adhesion to epithelial cells. Anaerobe. 2019;561–7. https://doi.org/10.1016/j.anaerobe.2019.01.001.
68. Yau S-y, Yip YSL, Formolo DA, He S, Lee THY, Wen C, Hryciw DH: Chronic consumption of a high linoleic acid diet during pregnancy, lactation and post-weaning period increases depression-like behavior in male, but not female offspring. Behavioural Brain Research. 2022;416113538. https://doi.org/10.1016/j.bbr.2021.113538.
69. Bayat AR, Razzaghi A, Sari M, Kairenius P, Tröscher A, Trevisi E, Vilkki J. The effect of dietary rumen-protected trans-10, cis-12 conjugated linoleic acid or a milk fat-depressing diet on energy metabolism, inflammation, and oxidative stress of dairy cows in early lactation. Journal of Dairy Science. 2022;105(4):3032–48. https://doi.org/10.3168/jds.2021-20543.
70. Denis P, Ferlay A, Nozière P, Gerard C, Schmidely P. Quantitative relationships between ingested and intestinal flows of linoleic and alpha-linolenic acids, body weight and milk performance in mid-lactation dairy cows. Animal. 2022;16(11):100661. https://doi.org/10.1016/j.animal.2022.100661.
71. Biswas AA, Lee SS, Mamuad LL, Kim S-H, Choi Y-J, Bae G-S, Lee K, Sung H-G, Lee S-S. Use of lysozyme as a feed additive on in vitro rumen fermentation and methane emission. Asian-Australas J Anim Sci. 2016;29(11):1601–7. https://doi.org/10.5713/ajas.16.0575.
72. Hardie LC, VandeHaar MJ, Tempelman RJ, Weigel KA, Armentano LE, Wiggans GR, Veerkamp RF, de Haas Y, Coffey MP, Connor EE, Hanigan MD, Staples C, Wang Z, Dekkers JCM, Spurlock DM. The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. Journal of Dairy Science. 2017;100(11):9061–75. https://doi.org/10.3168/jds.2017-12604.
73. Lu Y, Vandehaar MJ, Spurlock DM, Weigel KA, Armentano LE, Connor EE, Coffey M, Veerkamp RF, de Haas Y, Staples CR, Wang Z, Hanigan MD, Tempelman RJ. Genome-wide association analyses based on a multiple-trait approach for modeling feed efficiency. Journal of Dairy Science. 2018;101(4):3140–54. https://doi.org/10.3168/jds.2017-13364.
74. Prinsen RTMM, Rossoni A, Gredler B, Bieber A, Bagnato A, Strillacci MG: A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. Livestock Science. 2017;2027–12. https://doi.org/10.1016/j.livsci.2017.05.011.

Publisher's Note