Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions

Marina Martínez-Álvaro
SRUC https://orcid.org/0000-0003-2295-5839
Marc Auffret
SRUC
Carol-Anne Duthie
SRUC
Richard Dewhurst
SRUC
Matthew Cleveland
Genus plc
Mick Watson
Roslin Institute https://orcid.org/0000-0003-4211-0358
Rainer Roehe (rainer.roehe@sruc.ac.uk)
SRUC https://orcid.org/0000-0002-4880-3756

Article

Keywords: bovine host genome, rumen, CH4

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-290150/v1

Bovine host genome acts on specific metabolism, communication and genetic processes of rumen microbes host-genomically linked to methane emissions

Marina Martínez-Álvaro[1], Marc D. Auffret[1], Carol-Anne Duthie[1], Richard J. Dewhurst[1], Matthew A. Cleveland[2], Mick Watson[3] and Rainer Roehe[*1]

[1]Scotland's Rural College, Edinburgh, UK
[2]Genus plc, DeForest, WI, USA
[3]The Roslin Institute and the Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

[*]Corresponding author. Email: Rainer.Roehe@sruc.ac.uk

Introductory paragraph

Whereas recent studies in different species showed that the host genome shapes the microbial community profile, our new research strategy revealed substantial host genomic control of comprehensive functional microbial processes in the rumen of bovines by utilising microbial gene profiles from whole metagenomic sequencing. Of 1,107/225/1,141 rumen microbial genera/metagenome assembled uncultured genomes (RUGs)/genes identified, 203/16/352 were significantly (P<2.02 x10^{-5}) heritable (0.13 to 0.61), revealing substantial variation in host genomic control. We found 29/22/115 microbial genera/RUGs/genes host-genomically correlated (-0.93 to 0.92) with emissions of the potent greenhouse gas methane (CH_4), highlighting the strength of host genomic control of specific microbial processes impacting on CH_4. Only one of these microbial genes was directly involved in methanogenesis (*cofG*), whereas others were involved in providing substrates for archaea (e.g. *bcd* and *pccB*), important microbial interspecies communication mechanisms (*ABC.PE.P*), host-microbiome interaction (*TSTA3*) and genetic information processes (*RP-L35*). In our population, selection based on abundances of the 30 most informative microbial genes provided a mitigation potential of 17% of mean CH_4 emissions per generation, which is higher than for selection based on measured CH_4 using respiration chambers (13%), indicating the high potential of microbiome-driven breeding to cumulatively reduce CH_4 emissions and mitigate climate change.

Main text

Ruminant livestock harbour a unique symbiotic gut microbial population that transforms indigestible fibrous feed into high-quality products such as meat and milk for human consumption, which are vital to meet global food security and contribute to poverty[1] reduction in an increasing world population[2]. Yet to be solved is the negative environmental impact, as dairy and beef cattle account for 9.5% of all anthropogenic greenhouse gas (GHG) emissions[3] of which ruminal microbial fermentation represents 40-50%; in particular, due to

the highly potent GHG methane (CH_4)[4]. Additionally, CH_4 emissions imply a significant energy loss to the animal, ranging from 2 to 12% of gross energy intake[5]. Therefore, decreasing CH_4 emissions is expected to contribute significantly to the mitigation of climate change and to optimising the economic efficiency of cattle production.

Ruminal methanogenesis is a complex process dependent on the cooperation of taxonomic communities with different metabolic activities[6–9]. A diverse community of bacteria, ciliate protozoa and anaerobic fungi[10] convert complex diet carbohydrates, proteins and lipids into volatile fatty acids, lactate, microbial proteins and vitamins, whilst releasing CO_2, H_2 and other compounds. Four orders of ruminal methanogenic archaea use electrons derived from H_2, formate or methyl compounds to reduce carbon dioxide into CH_4 to obtain energy for growth[11]. Previous studies in ruminants[6,12–19], monogastric livestock[20,21], and humans[22–27] have shown a host genomic impact on the microbial community profiles. However, these profiles were mostly identified at genus level using sequence polymorphisms of the 16S rRNA gene and therefore did not consider the functional versatility of microbial strains. Furthermore, there is no comprehensive research elucidating how the complex functions of ruminal microbes determined by the abundances of their microbial genes in relation to methane emissions is influenced by the host genome and how this novel information can best be included into breeding of animals to reduce these emissions. In this study we applied a novel strategy in ruminants to identify this host genomic impact, with extensive characterization of ruminal microbiomes using whole metagenome sequencing of rumen microbial DNA samples from a bovine population designed for a powerful host genomic analysis[28–30] with high standardization of diets and other husbandry effects. We characterized the core ruminal microbiome identifying 1,108 cultured microbial genera by mapping our sequences to the Hungate1000 reference genome collection[31] and RefSeq[32] databases (Supplementary Table 1a); 225 Ruminal Uncultured Genomes (RUGs) by *de novo* metagenome-assembly of genomes[33] (Supplementary Table 1b) and 1,142 functional microbial genes (Supplementary Table 1c); all were present in most of our animals (n=359). For each of these 2,475 characteristics of the rumen microbiome the host genomic determination and correlation with methane emissions were analysed. After stringent adjustment for multiple testing, heritabilities (h^2) of microbial profiles significantly deviating from zero were obtained, which shows the effectiveness of this novel strategy.

Our specific hypothesis is that the host genome influences the abundance of not only functional microbial genes involved in metabolism, but also in interspecies communication, host-microbiome interactions and genetic information processing. These play a key integrating role in achieving a ruminal balance where fermentation of feed into essential nutrients utilised by the host is optimized and substrates utilised by methanogenesis e.g. H_2 excess are minimized. We studied the host-genomically influenced correlations between CH_4 emissions and abundances of 34 microbial genes carried by methanogenic archaea directly implicated in CH_4 metabolism; 511 involved in other metabolic pathways of bacteria, archaea, ciliate protozoa or fungi, indirectly influencing methanogenesis by minimizing required substrates through non-methanogenic routes that yield beneficial nutrients for ruminants[34] (e.g. acetogenesis, propionogenesis[35–38]), or generating methanogen-inhibitor

metabolites[39–41]; 207 in microbial communication processes and host-microbiome interaction (e.g. ABC transporters of different metabolites or fucose sensing) carried by fungi, bacteria and archaea[42–44], of importance because the synthesis of CH_4 in cooperation with other main metabolic routes in the rumen[45–47] are syntrophic processes amongst microbial communities[48]; 330 involved in genetic information processes (e.g. ribosomal biosynthesis) related to microbial growth[49]; and 60 at present not functionally characterized. We demonstrate that our hypothesis of a common host genomic control is valid by discovering significant host genomic correlations between specific microbial gene abundances (e.g. *ABC.PE.P* and *ABC.PE.S* in quorum sensing metal ions transport or *argD*, *bcd* and *pccb* in amino acid metabolism) and CH_4 emissions. Our results are obtained in bovines, but also provide an indication of potential host genomic effects on functional microbial genes and their biological processes in other species.

Interventions designed to alter the microbiome for CH_4 mitigation (e.g. protozoa defaunation[36,50], seaweed[51] and 3-NOP[52] additives) have often failed in the long-term due to microbiota adaptation to the new environment[53] or are associated with increasing production costs. In contrast, genomic selection that targets the part of the host genome modulating microbiome composition related to low CH_4-emitting cattle opens up the opportunity to provide a permanent solution based on cumulative responses to selection. Besides providing large insight into the complex host genomic effects on the rumen microbiome function, the novelty of this research goes further by providing the basis for an innovative cost-effective microbiome-driven breeding strategy to mitigate CH_4 emissions from cattle without measuring it directly, which is necessary considering the cost-prohibitive limitations of obtaining individual animal CH_4 emissions.

Results

Bovine host genomics affected CH_4 emissions produced by ruminal archaea. CH_4 emissions[54] were accurately measured from individual beef cattle (n=285) using the gold-standard method of respiration chambers. Animals within the same breed and diet expressed high phenotypic variability in CH_4 emissions with coefficients of variation from 23.2% to 28.5%, (Supplementary Fig 1). Genomic h^2 of CH_4 emissions revealed that 33% (P=3.34x10^{-5}) of this phenotypic variation was explained by host genome variation, which is consistent with other studies[55–57]. The h^2 obtained for CH_4 emissions is at the level of other traits for which substantial gains due to breeding are achieved, such as growth rate[58] and milk yield[59]. In addition, there was large genomic variation for CH_4 emissions as deviation from the mean ranged from -2.67 to 3.51 g/kg of dry matter intake (DMI) with no difference between breeds (P>0.16), which suggests that bovines have most likely not been indirectly selected for CH_4 emissions as a result of a lack of genetic correlation to those traits under selection.

Host genomics shapes the ruminal microbiome composition. We next investigated the proportion of the ruminal microbiome variation at taxonomic and functional levels explained by the host genomic variation among individuals, by estimating h^2 of the ruminal abundances of 1,107 genera, 225 RUGs and 1,141 microbial genes. Our results demonstrate significant h^2 ($P\leq$2.02x10^{-5}) in a range between 0.13 and 0.61 for the abundances of 203 microbial

genera, 16 RUGs, and 352 microbial genes representing cumulatively 58.6%, 5.97% and 28.3%, respectively, of the total relative abundance (RA) (Fig. 1 and Supplementary Table 2a, b, c). Amongst the 203 genera, 20 were highly heritable (h^2>0.40), which belonged exclusively to bacteria (e.g. Firmicutes *Acidaminococcus* (RA=0.3%), h^2=0.54, P=5.61x10^{-11}) and archaea (e.g. hydrogenotrophic methanogen *Methanospirillum*[60] (RA=0.0005%), h^2=0.40, P≤6.2x10^{-7}). Host genome also shaped the abundance of the hydrogenotrophic/methylotrophic[48] methanogen *Candidatus Methanoplasma* (RA =0.002%, h^2=0.32, P≤1.20x10^{-7}), and to a lesser extent the abundance of ubiquitous *Methanobrevibacter* (RA=5.02%, h^2=0.24, P=8.75x10^{-6})[12–14], *Candidatus Methanomethylophilus* (RA=0.05%, h^2=0.26, P=1.74x10^{-5}) and *Methanothermus* (RA=0.002%, h^2=0.25, P=6.87x10^{-6}). Reinforcing the evidence of a host-genomic component in the abundance of methanogenic archaea, 6 RUGs annotated as uncultured *Methanobrevibacter* sp. (RA>0.27%) demonstrated moderate to high h^2 estimates (0.35-0.48, P≤1.69x10^{-5}), indicating that more specific classification using RUGs provides the opportunity to find highly heritable *Methanobrevibacter* sp. The most abundant complex carbohydrates degraders in the rumen - *Eubacterium* (RA=1.02%), *Prevotella* (RA=39.2%), *Butyrivibrio* (RA=2.54%), *Bacteroides* (RA=1.39%) and *Pseudibutyrivibrio* (RA=0.54%) – were highly (h^2=0.51 for *Eubacterium*, P=5.49x10^{-9}) or moderately (h^2=0.23-0.33 for the others, P≤9.67x10^{-6}) heritable; with 8 highly abundant RUGs (RA>0.21%) classified as uncultured *Prevotellaceae bacterium* having h^2 from 0.24 to 0.45 (P≤1.25x10^{-5}). These results support the concepts of a "core heritable microbiome"[15,61] and stability over time of certain microbial genera abundance such as *Prevotella*[62]. None of the fungi and protist genera, which are considered to be non-essential for rumen function and highly variable within different host species[63], were highly heritable.

For the first time we elucidated that specific functional capacity of the ruminal microbiome is heritable by estimating the h^2 of a comprehensive set of microbial genes, of which 31 were highly (h^2 >0.4), 273 moderately (0.2<h^2<0.4) and 48 lowly (h^2 <0.2) heritable. These microbial genes are involved in a wide variety of metabolic functions (Fig. 1b and Supplementary Table 2c), e.g. synthesis of microbial proteins or volatile fatty acids, suggesting that the host genome influences the growth of microbes responsible for the release of nutrients during microbial fermentation[64,65]. Among 34 microbial genes involved in the CH_4 metabolism pathway, 15 showed moderate h^2 of 0.20-0.27 (P≤1.9x10^{-5}), e.g. *mcrA*, *mcrB*, *mcrG, mtrD*, *mtrE,* and *mtrH*. Ribosomal biosynthesis was revealed to be under strong host-genomic control with 51 heritable microbial genes, representing a cumulative RA of 17.7%, including 9 highly heritable genes (h^2=0.40-0.52, P≤7.98x10^{-6}) synthesizing the large ribosomal subunit. Intracellular ribosomal biosynthesis reflects the growth rate of microbial organisms, given that ribosomes can account for up to 40% of their cellular dry mass[49], and cell fitness and optimal growth is tightly coupled to efficient protein synthesis[66]. Demonstrating that differences among animals in complex microbiome functions are partly due to host genomic variation opens up opportunities to consider a new source of genetic variation not only in ruminants but also in humans, where the h^2 of microbial gene abundances was estimated to be even larger (0.65-0.91)[67].

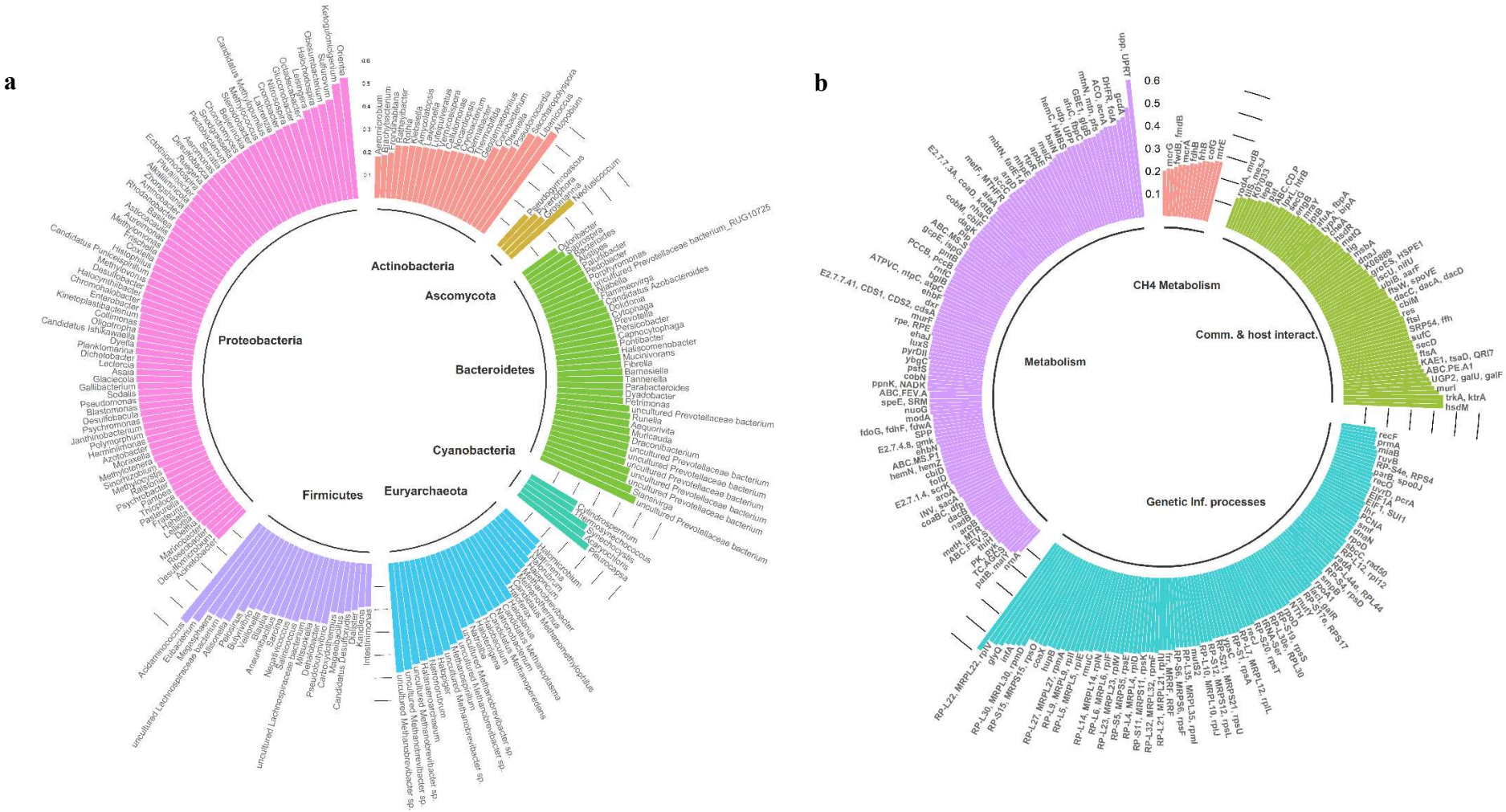


Fig. 1| Genomic heritability (h^2) estimates of log-ratio transformed abundances of microbial taxa (a) and their genes (b) in the rumen of bovines. Bars show the h^2 values of 203/16/352 rumen microbial genera/uncultured genomes (RUGs)/genes tested exhibiting non-zero h^2 estimates ($P < 2.02 \times 10^{-5}$) a. Cultured microbial genera and RUGs classified within phylum. **b.** Microbial genes grouped by microbial biological processes: Microbial communication and host-microbiome interaction (Comm. & host interact.), Genetic information processes (Genetic Inf. processes), metabolism other than methane (Metabolism), and methane metabolism (CH_4 metabolism).

Ruminal microbial mechanisms related to CH_4 emissions are influenced by host genomics. The existence of a common host genomic influence on CH_4 emissions and the rumen microbiome was evaluated by estimating host-genomic correlations between CH_4 emissions and each microbial genus/RUG/gene abundance (r_{gCH4}). Based on the probability of r_{gCH4} being different from 0 (P_0)≥0.95, our study revealed 29 microbial genera, 22 RUGs and 115 functional microbial genes strongly host-genomically correlated with CH_4 emissions (r_{gCH4} from |0.59| to |0.93|, Supplementary Tables 3a, b, c). Among the significant microbial communities, most were bacteria (22 genera/17 RUGs) belonging to Bacteroidetes (5/14), Firmicutes (6/2) and Proteobacteria (9/1) phyla. Most microbial genes with strong r_{gCH4} were not directly involved in CH_4 metabolism pathways, but rather mechanisms indirectly affecting CH_4 production - most likely by limiting substrates for methanogenesis[9,68], inhibiting methanogens, playing a role coordinating actions among microbial communities and the host or leading microbial genetic processes. Only H_2-oxidizing *Methanoregula* (RA=0.003%) with unknown activity in rumen[42] and the microbial gene *cofG* involved in F_{420} coenzyme biosynthesis[69,70] resulted in significant negative r_{gCH4} (-0.82 and -0.71, P_0≥0.95), suggesting that these are abundant under ruminal conditions unfavourable for other high CH_4 producing methanogens. Four uncultured *Methanobrevibacter* sp. showed negative r_{gCH4} (<-0.72, P_0≥0.95) and one was positive (0.91, P_0=0.99), indicating that the relationship amongst the abundance of *Methanobrevibacter* and CH_4 emissions is complex as different species may have functional versatility. We hypothesize that some *Methanobrevibacter* sp. can produce CH_4 even under a challenging ruminal environment (e.g. low pH value), however, at a substantially lower level than those adapted to more favourable conditions. To visualize which microbial genus/gene abundances in the rumen are governed by a common host genomic background, we constructed a co-abundance network based on Pearson correlations among deregressed host-genomic effects for each microbial genus/RUG/gene (Fig. 2, Supplementary Table 4). This approach revealed co-abundance clusters of bacterial and fungal genera[71] with strong r_{gCH4} and methanogenic archaea, e.g. fungal *Metschnikowia* (r_{gCH4}=0.77, P_0=0.96) and archaeal *Methanosarcina* (cluster 9 in Fig. 2); and of microbial genes not directly involved in CH_4 metabolism but with strong r_{gCH4} (e.g. *RP-L6,* r_{gCH4} 0.71, P_0=0.96) and those involved directly in CH_4 metabolism (e.g. *fbaA*, cluster 1 in Fig. 2).

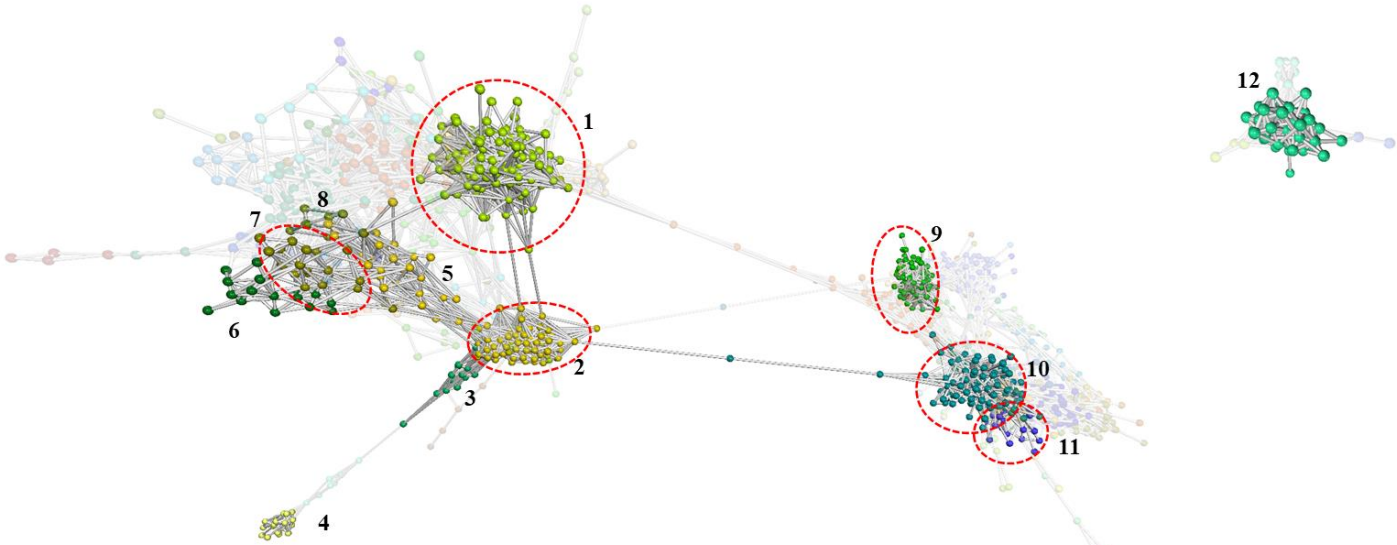


Fig. 2 | Network clusters of commonly host-genomically affected abundances of microbial genera/RUGs/genes identified in the bovine rumen. Nodes represent microbial genus/RUG/genes, and edges represent Pearson correlations among deregressed genomic effects of log-ratio transformed genera/RUGs/gene abundances > 0.70 (n= 359 animals). Clusters including ≥ 3 methanogenic archaea genera, RUGs and microbial genes involved in methane (CH_4) metabolism pathway according to KEGG database or microbial genera/RUGs/genes host-genomically correlated with CH_4 emissions (probability of the host-genomic correlation being higher or lower than 0 (P_0) ≥0.95) are highlighted and numbered from 1 to 12. Red dashed circles indicate the clusters including methanogenic archaea genera or RUGs and microbial genes involved in the CH_4 metabolism pathway and associated with microbial genera, RUGs and genes significantly (P_0>0.95) host-genomically correlated with CH_4 emissions.

The most important host-genomically affected ruminal microbial mechanisms associated with CH_4 production (based on r_{gCH4}) are as follows:

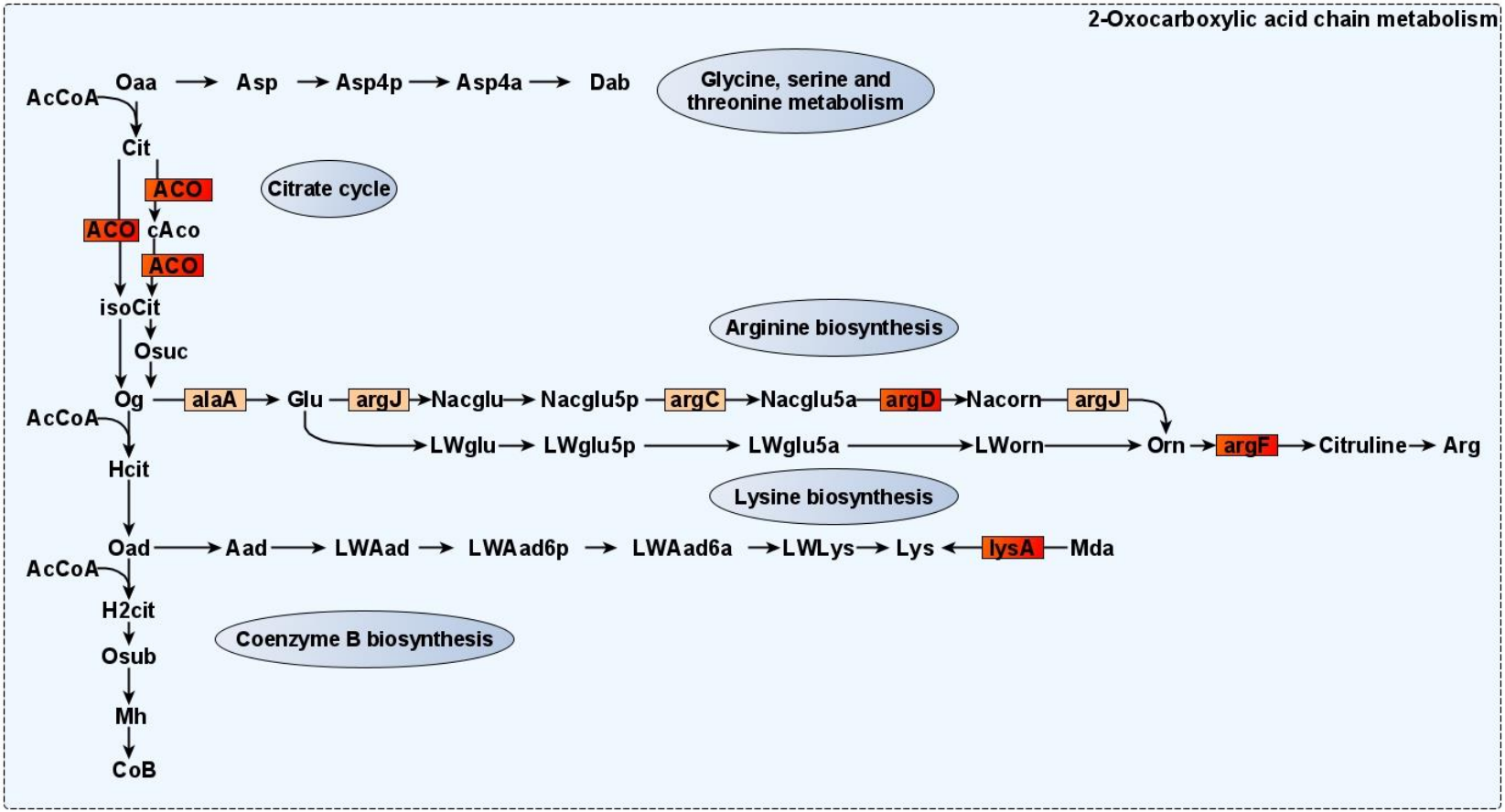
Microbial metabolism. An extensive group of microbial genes involved in amino acid metabolic and transport pathways displayed negative $r_{gCH4.}$ Part of this group of microbial genes was involved in the biosynthesis of arginine[72] and branched-amino acids[73,74] via oxocarboxylic acid metabolism (*argF*, a*rgD, ilvA* with r_{gCH4}=-0.84 to -0.88 $P_0 \geq 0.96$; and *argJ, argC, alaA, ilvH*, *leuB* and *leuD* with r_{gCH4}=-0.55 to -0.77 at lower evidence $P_0 \geq 0.85$, Fig. 3a, b). *Aconitate hydratase* (*ACO)* catalysing the isomerization of citrate to isocitrate in the early stage of the oxocarboxylic chain extension, and *bcd* and *pccB* degrading branched-chain amino acids into branched-chain volatile fatty acids which have an inhibitory effect on methanogens[39], also expressed negative r_{gCH4}=-0.76 to -0.90 ($P_0 \geq 0.95$). We also estimated negative r_{gCH4} for microbial genes coding ABC transporters of polar and branched amino acids (*ABC.PA.A*, *ABC.PA.S, livH*, *livG,* and *livK* r_{gCH4}=-0.83 and -0.93, $P_0 \geq 0.95$). Another group of microbial genes was related to the metabolism of aromatic amino acids tryptophan, tyrosine and phenylalanine (*AROA2*, *trpA*, *trpD, trpE, tyrA2* and *paaH* with r_{gCH4}=-0.74 to -0.87, $P_0 \geq 0.95$ and *aroC, aroA, aroF, trpG* and *trpB,* with $r_{gCH4=}$-0.68 to -0.74 at lower evidence $P_0 \geq 0.85$, Fig 3c). More specifically, *trpE trpD* and *trpA* take part in the metabolism of L-tryptophan (Fig 3c) whose catabolites (e.g. indole) are important signalling molecules in biofilm formation[75], and activation of host immune system[76]. Moreover, of 2-oxocarboxylic acid and tyrosine catabolites are precursors for the biosynthesis of coenzyme B[74,77] and methanofuran[70] methanogenic cofactors, and their diversion into the synthesis of other substrates (e.g. arginine, branched- chain amino acids or tryptophan) could explain their negative r_{gCH4}. Lastly, four microbial genes with negative r_{gCH4} (-0.61 to -0.87, $P_0 \geq 0.95$) were associated with methionine metabolism (*metE*, *DNMT1)* and transport (*metQ* and *metN*). Methionine is associated with minor methylotrophic methanogenesis pathway[78] in the rumen[79,80] and with enhancement of microbial long-chain fatty acid production[81], an extremely H_2 demanding process[37]. Our study highlights that the negative association between microbial amino acid metabolism and CH_4[82,83] has a host genomic component. This could be partly due to host genomic effects[13] on ruminal passage rates, which have opposite effects on microbial protein synthesis efficiency[64] and CH_4 production[84].

We obtained negative r_{gCH4} (from -0.60 to -0.85, $P_0 \geq 0.95$) for the abundance of several microbial genes responsible for sucrose metabolism (*sacA*, *maIZ*, *bgLB*, *SPP,* and *sucrose phosphorylase,* Fig 3d), including the highly abundant sucrose fermenter[85] *Eubacterium* (RA=1.02%)*,* transporters of multiple sugars across the membrane[82] (*ABC.MS.P1*, *ABC.MS.S,* and *ABC.MS.P*), and the microbial gene *PTS-EI* which catalyses the phosphorylation of incoming sugar substrates concomitantly with their translocation across the cell membrane. Microorganisms capable of fast growth on soluble sugars are suggested to be favoured in hosts with low rumen size and high turnover rate[82,86], features also associated with low CH_4-emissions[87]. Degradation of easily fermentable carbohydrates, such as sucrose or starch cause a pH decline which has a strong CH_4 reducing effect as a result of pH sensitivity of methanogens or H_2-producing microbes[88]. Furthermore, previously mentioned microbial genes *aroA* and *trpE* are involved in the shikimate pathway[89] linking sugar metabolism with the synthesis of microbial proteins (aromatic amino acids, tyrosine,

phenylalanine and tryptophan) which are an important source of amino acids for the host. Microbial protein yield from sucrose is suggested to be more persistent over time in comparison to other carbohydrates[90], and partially stored by sucrose utilizers (eg. *Eubacterium*) for the maintenance of the microbial population[90].

We also found negative r_{gCH4} for the abundance of hydrogenotrophic acetogenic bacteria *Blautia*[38], together with *Eubacterium*[91](r_{gCH4}=-0.60 and -0.73, $P_0 \geq 0.95$), and the *fhs* microbial gene involved in the reductive Wood-Ljungdahl acetyl-CoA pathway (r_{gCH4}=-0.79, P_0=0.98). Acetogens produce volatile fatty acids (mainly acetate but also propionate and butyrate[92]), which served as host nutrients to improve animal performance[34] and simultaneously compete against methanogens for metabolic H_2[8,35,38]. Despite acetogenesis being thermodynamically less favourable than the reduction of CO_2 into CH_4[93] in rumen, this may vary upon microbial interactions and host-genomically influenced ruminal environmental factors[34,38,65]. Propionogenesis via acrylate[33,82,86,94] and lactaldehyde routes[31] was another microbial mechanism under host genomic influence lowering CH_4 emissions as indicated by negative r_{gCH4} (-0.76 to -0.90, $P_0 \geq 0.95$) for the abundances of microbial genes *bcd* and *pccB* involved in propanoyl-CoA metabolism and *fucO* catalysing the reduction of lactaldehyde into 1,2-propanediol, as well as the highly abundant (RA=0.08%) lactate-producing bacteria *Kandleria* (r_{gCH4}=-0.87, P_0=0.99). Lactate utilization for propionate production not only reduces H_2 availability for methanogenesis[36,95] but also prevents rumen acidosis and results in a more efficient rumen fermentation[96]. The abundance of six microbial genes encoding [4Fe-4S] cluster containing proteins (*bioB*, *cobL*, *cofG*, *nifU, ACO,* and *pflA*) involved in electron transfer mechanisms in redox reactions presented r_{gCH4} from -0.71 to -0.87 ($P_0 \geq 0.96$). The first two proteins are involved in the synthesis of substrates required for methanogenic cofactors; i.e. *bioB* catalyses the conversion of dethiobiotin to biotin[97] which competes with coenzyme B for the synthesis of its alkyl portion[98,99]; and *cobL* together with *hemC* (r_{gCH4}=-0.91, P_0=1.00) take part in porphyrin metabolism, required for different processes including the synthesis of porphyrin-based cofactors vitamin B_{12} and F_{430}[100]. Nitrogen fixation protein *nifU* carries out N_2 reduction into ammonia[101], which can act as an alternative H_2-consuming sink competing with ruminal methanogenesis. Further negative r_{gCH4} were obtained for microbial genes in thiamine metabolism (*iscS, thiD, thiH* and *thiE* with r_{gCH4} from -0.88 to -0.70, $P_0 \geq 0.91$)[102]; hydration of long-chain fatty acid oleate into anti-tumoral hydroxystearic acid[103,104] (*ohyA,* -0.81, P_0=0.95), or import of methanogen-inhibitors long-chain fatty acids[40] (*ABCB-BAC*[105], r_{gCH4}=-0.9, P_0=0.99). Moreover, highly abundant bacteria genera with ruminal fatty acid biohydrogenation activity[106,107], *Eubacterium* and *Butyrivibrio* (RA=2.54%, r_{gCH4}=-0.37, P_0=0.80) were negatively correlated with CH_4.

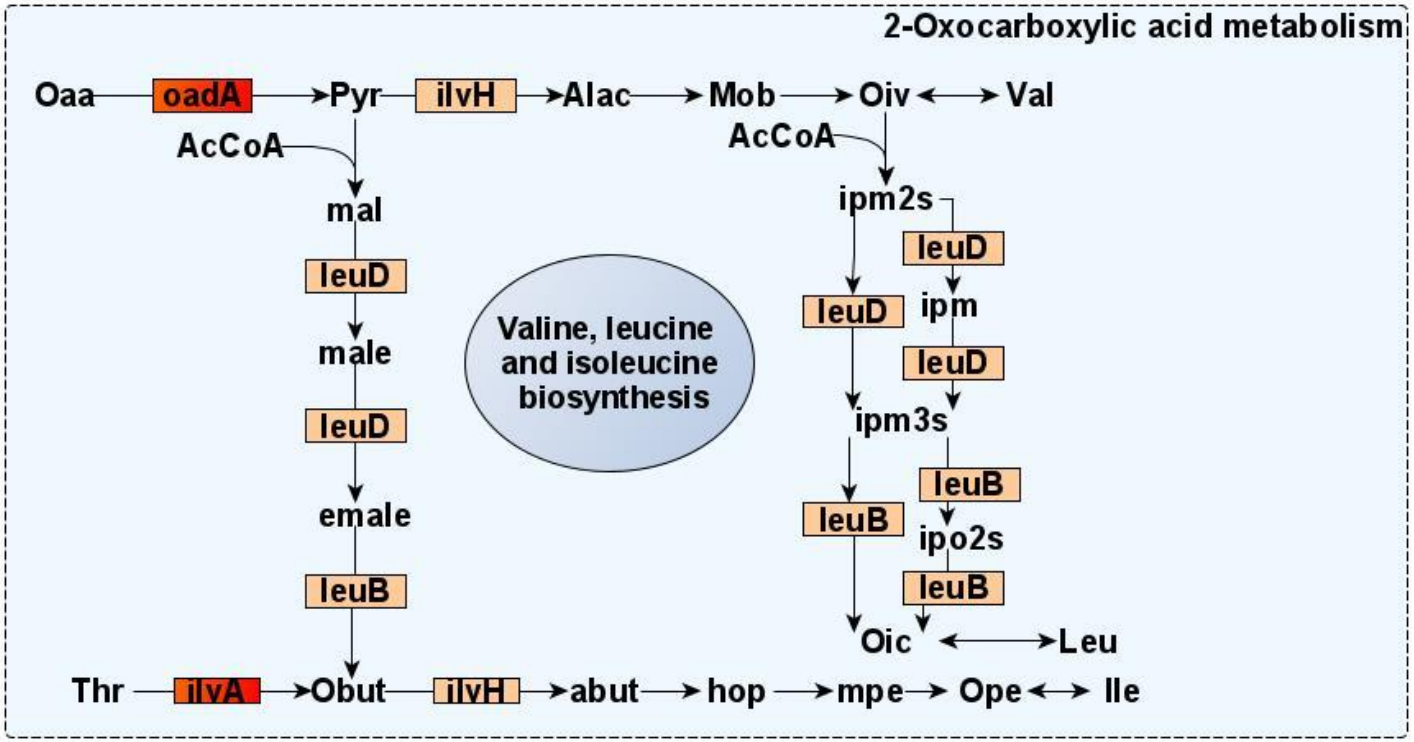
a


2-Oxocarboxylic acid chain metabolism
Glycine, serine and threonine metabolism
Citrate cycle
Arginine biosynthesis
Lysine biosynthesis
Coenzyme B biosynthesis
AcCoA
Oaa
Asp
Asp4p
Asp4a
Dab
Cit
ACO
cAco
isoCit
Osuc
Og
alaA
Glu
argJ
Nacglu
Nacglu5p
argC
Nacglu5a
argD
Nacorn
LWglu
LWglu5p
LWglu5a
LWorn
Orn
argF
Citruline
Arg
Hcit
Oad
Aad
LWAad
LWAad6p
LWAad6a
LWLys
Lys
lysA
Mda
H2cit
Osub
Mh
CoB

311

312

b



313

314

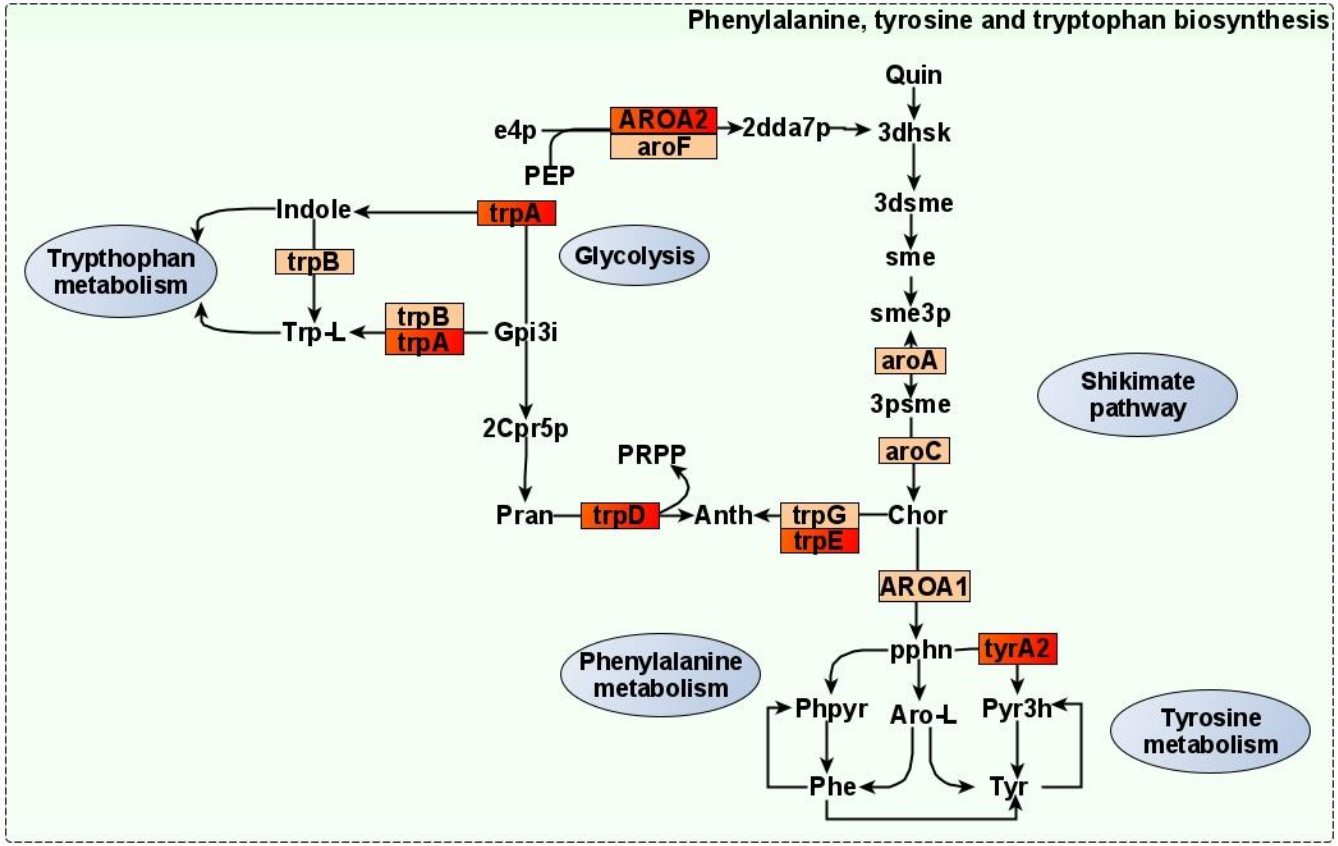
315

316

317

318

c


Phenylalanine, tyrosine and tryptophan biosynthesis
Quin
e4p
AROA2
aroF
2dda7p
3dhsk
PEP
3dsme
Indole
trpA
sme
Trypthophan metabolism
trpB
Glycolysis
sme3p
trpB
trpA
Trp-L
Gpi3i
aroA
Shikimate pathway
3psme
2Cpr5p
aroC
PRPP
Pran
trpD
Anth
trpG
trpE
Chor
AROA1
pphn
tyrA2
Phenylalanine metabolism
Phpyr
Aro-L
Pyr3h
Tyrosine metabolism
Phe
Tyr

d

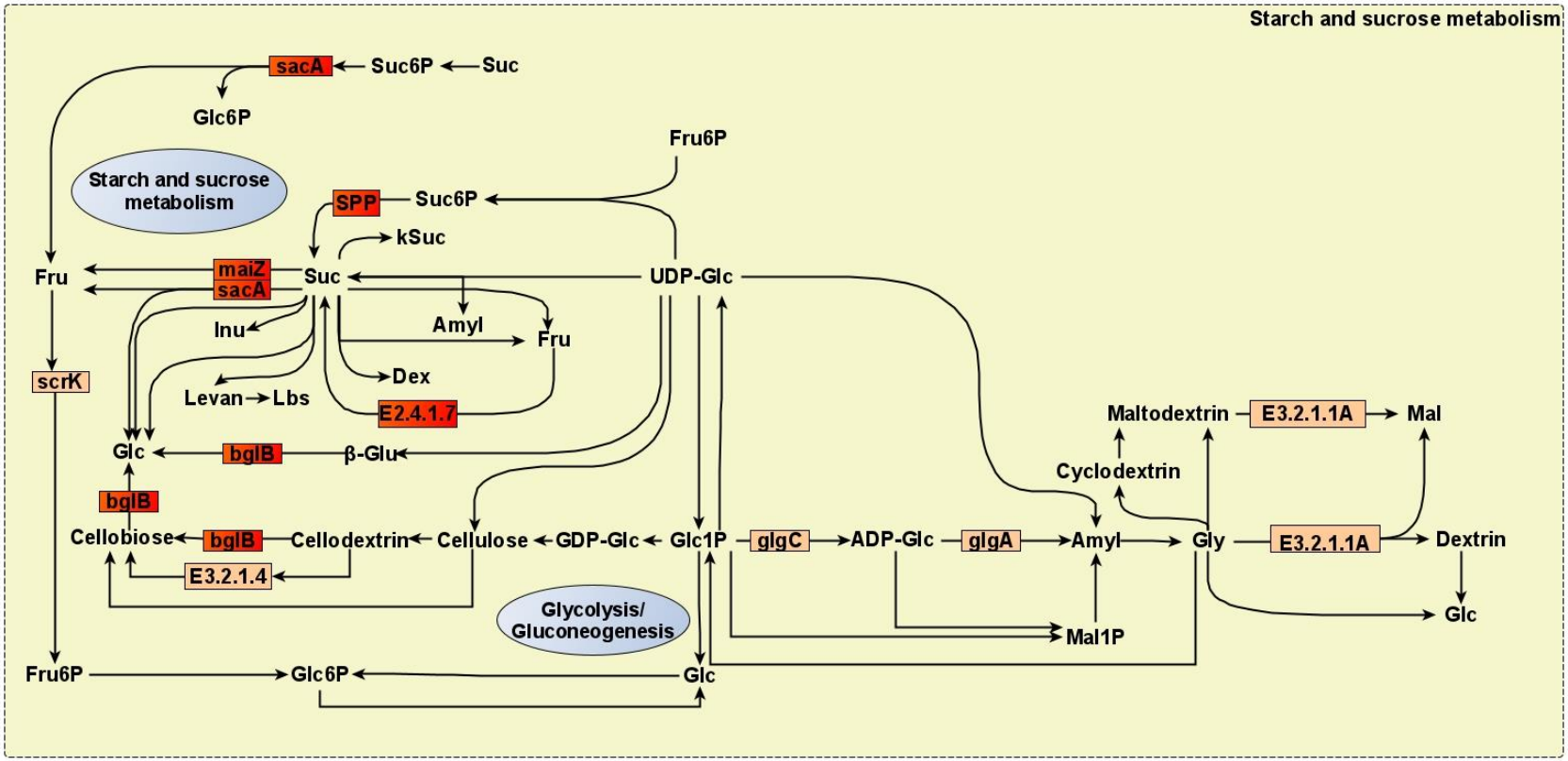


Fig. 3 | Reaction schemes of 2-Oxocarboxylic acid metabolism and (a) glycine, serine, threonine, arginine, lysine and Coenzyme B biosynthesis or (b) branched amino acid biosynthesis, (c) phenylalanine, tyrosine and tryptophan biosynthesis and (d) starch and sucrose metabolism, in which additive log-ratio transformed microbial gene abundances strongly host-genomically correlated with methane emissions (r_{gCH4}) are involved. Small rectangles symbolize proteins encoded by the microbial genes. Microbial genes are highlighted in red when their r_{gCH4} estimates range between -0.74 and -0.93 and shows a probability of being different from 0 (P_0) ≥0.95; and in orange when they range between |0.55| and |0.77| and P_0≥0.85. Compounds are denoted by their short names. Full names of compounds and microbial genes are given in Supplementary Data 1.

Microbial communication and host-microbiome interaction mechanisms. The majority of methanogens in the rumen are integrated into the biofilm on the surface of feed particles where H_2 producing bacteria are active[108–110]. We found strong negative r_{gCH4} (-0.78 to -0.92, $P_0 \geq 0.96$) for abundances of microbial genes mediating microbial interactions, involved in ABC transport of cobalt/nickel (*cbiO,* and *cbiQ*) and quorum sensing-related peptide/nickel ions (*ABC.PE.P*, *ABC.PE.S*, *ABC.PE.A*, *ABC.PE.P1*) - cobalt and nickel being detrimental for hydrogenotrophic and aceticlastic methanogenic activity[41]-, protein export (*secD* and *secF*) and chemotaxis (*cheA* and *mcp*)*;* and positive r_{gCH4} for transcription protein *cbpA* (0.85, P_0=0.97) acting as a microbial response to maintain plasmids replication during amino acid starvation[111]. CH_4 emissions were also genomically correlated with abundances of microbial genes mediating host-microbiome interaction; e.g. *cbh* and *baiN*[112] (r_{gCH4}=-0.80, $P_0 \geq 0.96$) involved in bacterial biosynthesis of secondary bile acids which activate metabolic receptors within gut, host liver and peripheral tissues[112,113] and inhibit CH_4 production in the rumen by transferring metabolic H_2 into propionate production[114]. Another interesting finding is that *TSTA3*, involved in the metabolism of host-microbiome crosstalk mediator fucose[115], displays a positive r_{gCH4} (0.85, P_0=0.98). Fucose is a component of mucins present in saliva[116], which is produced abundantly by ruminants and acts as a pH buffer during ruminal fermentation due to its phosphate and bicarbonate content[117]. Cellulolytic *Fibrobacter*, an indicator of high pH levels in rumen[118] was positively host-genomically correlated to *TSTA3* in our data (0.66, P_0=0.94), whilst lactic acid producer *Kandleria*, generally associated with low pH levels and negative r_{gCH4}, was host-genomically correlated to *TSTA3* negatively (-0.70, P_0=0.90). Thus *TSTA3* could be involved in signalling enhanced saliva production, resulting in increased rumen pH that is known to stimulate the growth of methanogenic archaea and CH_4 emissions[119].

Genetic information processes. Ribosomal biogenesis represented by *RP-S10, RP-S12, RP-S17, RP-L2, RP-L3, RP-L6, RP-L23, RP-L28, RP-L34,* and *RP-L35,* was one of the few microbial mechanisms with positive r_{gCH4} from 0.71 to 0.84 ($P_0 \geq 0.95$). All of them are universal ribosomal proteins homologous in bacteria, archaea, and eukarya; except for *RP-L28*, *RP-L34* and *RP-L35* exclusively found in bacteria[120,121]. Given that protein synthesis is highly coupled with cellular growth[66], these results suggest that the rumen environment provided by low CH_4-emitter host genomes are related to lower growth or activities of specific microbes directly or indirectly involved in methanogenesis.

RUGs enriched with microbial genes are strongly host-genomically correlated to CH_4 emissions. The 20 highly-prevalent (present in >200 animals) RUGs containing the highest number of unique proteins from the 115 microbial genes with strong r_{gCH4} were all bacterial RUGs carrying between 114 to 180 unique proteins classified into 60 to 84 microbial genes (Fig. 4 and Supplementary Tables 5 and 6). Of these 20 highly-enriched bacterial RUGs, 18 showed negative r_{gCH4} consistently with the majority of the microbial genes; 6 of them with r_{gCH4}<-0.65 (P_0>0.85) from which 5 RUGs were classified as *uncultured Lachnospiraceae bacterium* (RUG10082, RUG13438, RUG13308, RUG13002, RUG12132) and 1 as *uncultured Clostridiales bacterium* (RUG10940).

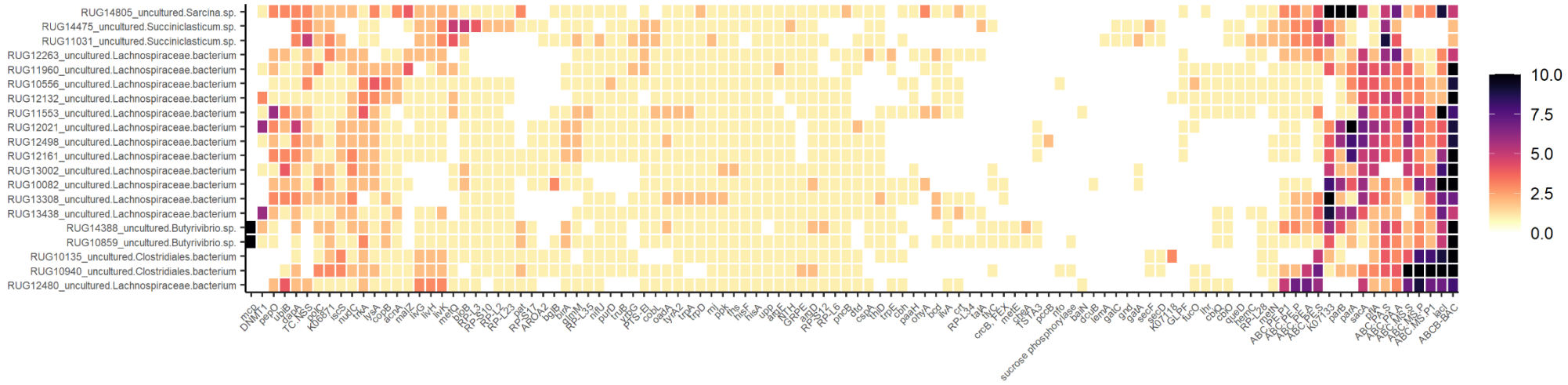


Fig. 4 | Top 20 Rumen Uncultured Genomes (RUGs) highly enriched with the 115 microbial genes host-genomically correlated to methane emissions with a probability of being higher or lower than 0 (P_0) ≥0.95. Colour scale represents the number of unique proteins mapping into each KEGG orthologous group (i.e. microbial gene). Full names of microbial genes are given in Supplementary Data 2.

We also investigated the enrichment of these 115 microbial genes in the 6 RUGs with r_{gCH4} ($P_0 \geq 0.95$) annotated at genus level (Supplementary Table 3c) and in those RUGs annotated as the 29 microbial genera with r_{gCH4} ($P_0 \geq 0.95$). Our findings show that part of the mechanisms identified in this study occur in the 5 uncultured *Methanobrevibacter* sp. RUGs (each carrying at least 45 out of the 115 microbial genes) and also in RUGs annotated as *Eubacterium ruminatum, Eubacterium pyruvativorans, Kandleria vitulina,* and uncultured sp. of *Blautia, Anaerovibrio* and *Succinivibrio* (each carrying at least 49 out of the 115 microbial genes, Supplementary Figure 2). The *uncultured Methanobrevibacter* sp. with positive r_{gCH4} (RUG12982) carried fewer unique proteins (67 vs. 75 to 93) and microbial genes (51 vs. 55 to 62) than the other 4 *uncultured Methanobrevibacter* sp. RUGs with negative r_{gCH4}; lacking, for example *argD* in arginine biosynthesis, *tyrA2* in tyrosine and tryptophan metabolism and *DNMT1* in methionine metabolism, which reinforces the hypothesis of functional versatility amongst different *Methanobrevibacter* species explaining their different effects and estimated r_{gCH4} on CH_4 emissions.

Microbiome-driven breeding of the bovine host for mitigation of CH_4 emissions. The comprehensive findings of the host genomic associations between microbial genus/RUG/gene abundances and CH_4 emissions enabled us to predict its mitigation potential when applying genomic selection targeting each of them individually (Supplementary Table 7), indirectly informing about the impact of each microbial mechanism on methanogenesis. Considering 30% of our cattle population being selected based on the abundances of each microbial gene, *maiZ* in sucrose metabolism, *ABC.PE.P* in quorum sensing peptide/nickel transport, *hemc* in porphyrin or *upp* in pyrimidine metabolism are predicted to result in the highest CH_4 mitigation potential (-5.2, -5.3, -5.8 and -6.54% of CH_4 emissions mean respectively, $P_0 \geq 0.99$). Subsequently, our study aimed to find a group of heritable ($P \leq 2.02 \times 10^{-5}$) ruminal microbial genera/RUGs/genes (RA>0.01%) with strong r_{gCH4} ($P_0 \geq 0.95$) to be used collectively for selecting the host genomes associated with low CH_4 emissions (Supplementary Table 8). We identified 4 microbial genera (*Eubacterium*, *Blautia, Odoribacter* and *Kandleria*), 3 RUGs (two annotated as *uncultured Methanobrevibacter sp.* and one as *uncultured Prevotellaceae bacterium*) and 36 microbial genes meeting these requirements. We selected 30 microbial genes (Fig 5a) covering several microbial mechanisms, e.g. sugar and nickel transport (*ABC.PE.P, ABC.MS.P1* and *ABC.MS.S*), fucose sensing (*TSTA3*), chemotaxis (*mcp*), ribosomal biosynthesis (*RP-L6*, *RP-L23, RP-L28*, *RP-L35, RP-S12,* and *RP-S17*), reductive acetogenesis (*fhs*) and metabolism of amino acids (*argD*), sucrose (*SPP*), CH_4 (*cofG*), biotin (*bioB*), propionate (*pccB*), porphyrin (*hemC*), thiamine (*thiD*) and pyrimidine (*upp*). A deep study of the host-genomic correlations among these 30 selected microbial genes showed a common host genomic background influencing the abundance of *ABC.PE.P, ABC.MS.P1*, *fhs*, *cofG*, *argD*, *hemC, thiD, upp*, *tlyC*, *NTH*, and *copB* with host-genomic correlations among each other ranging from 0.62 (P_0=0.90) to 0.99 (P_0=1.00) (Fig. 5b).

Finally, we evaluated the accuracies and response to selection in CH_4 emission mitigation in our population based on genomic selection using three different selection criteria: (1) CH_4 emissions measured by the "gold-standard" technique of respiration chambers, (2) the 30 microbial gene abundances exhibiting strong r_{gCH4}, and (3) combining both preceding criteria.

Using microbiome-driven breeding based on the abundance of 30 specific microbial genes resulted in a mean estimation accuracy of host-genomic effects for CH_4 emissions to be 34% higher than using measured CH_4 emissions (0.70±0.18 vs. 0.52±0.11) and confirmed that functional microbial genes are an extremely valuable source of information to perform host genomic evaluations for CH_4 emissions. Using the combined selection criteria (3), the accuracy of estimation was 14% larger than using rumen microbial gene information alone (0.80±0.20). Response to selection in CH_4 emissions achieved by selecting animals with low CH_4 emission breeding values predicted exclusively by microbial gene abundance information resulted in a reduction in emissions of -1.43±0.14 to -3.32±0.77 g CH_4/kg DMI per generation, depending on selection intensity (from 1.16 to 2.67 in the analysed population, Fig. 6). These results indicate that in our population, microbiome-driven breeding for CH_4 emissions reduced its magnitude by 7 to 17% of its mean per generation, without the necessity for costly measures of CH_4 emissions.

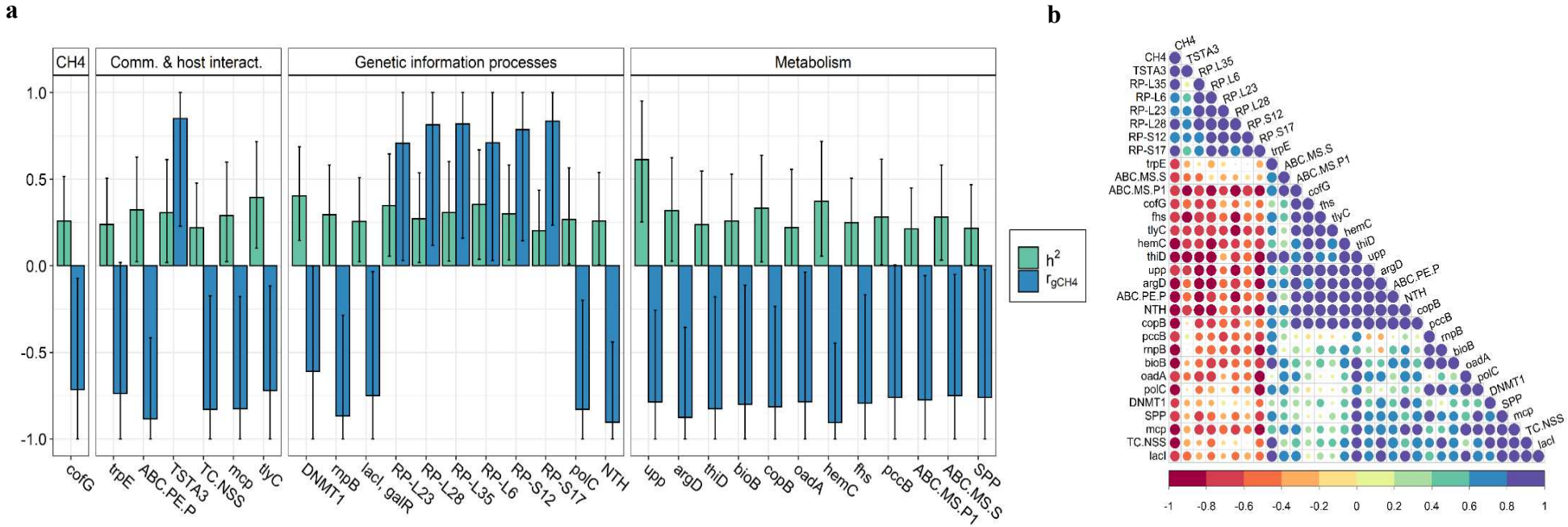


Fig. 5 | Microbial genes selected to be used collectively for selecting the host genomes associated with low CH_4 emissions, meeting 3 criteria: showing significant heritability (h^2) with $P < 2.02 \times 10^{-5}$; a host genomic correlation with CH_4 (r_{gCH4}) with a probability of being higher or lower than 0 (P_0) >0.95, and showing a relative abundance above 0.01%. a. Estimates of h^2 and r_{gCH4} (error bars represent the highest posterior density interval enclosing 95% probability). Microbial genes grouped by microbial biological processes: Methane metabolism (CH_4), Microbial communication and host-microbiome interaction (Comm. & host interact.), Genetic Information processes and metabolism other than CH_4 (Metabolism)**. b.** Correlogram showing the host genomic correlations estimates among the log-ratio transformed microbial gene abundances selected for breeding purposes**.** Full names of microbial genes selected for breeding purposes are given in Supplementary Data 3.

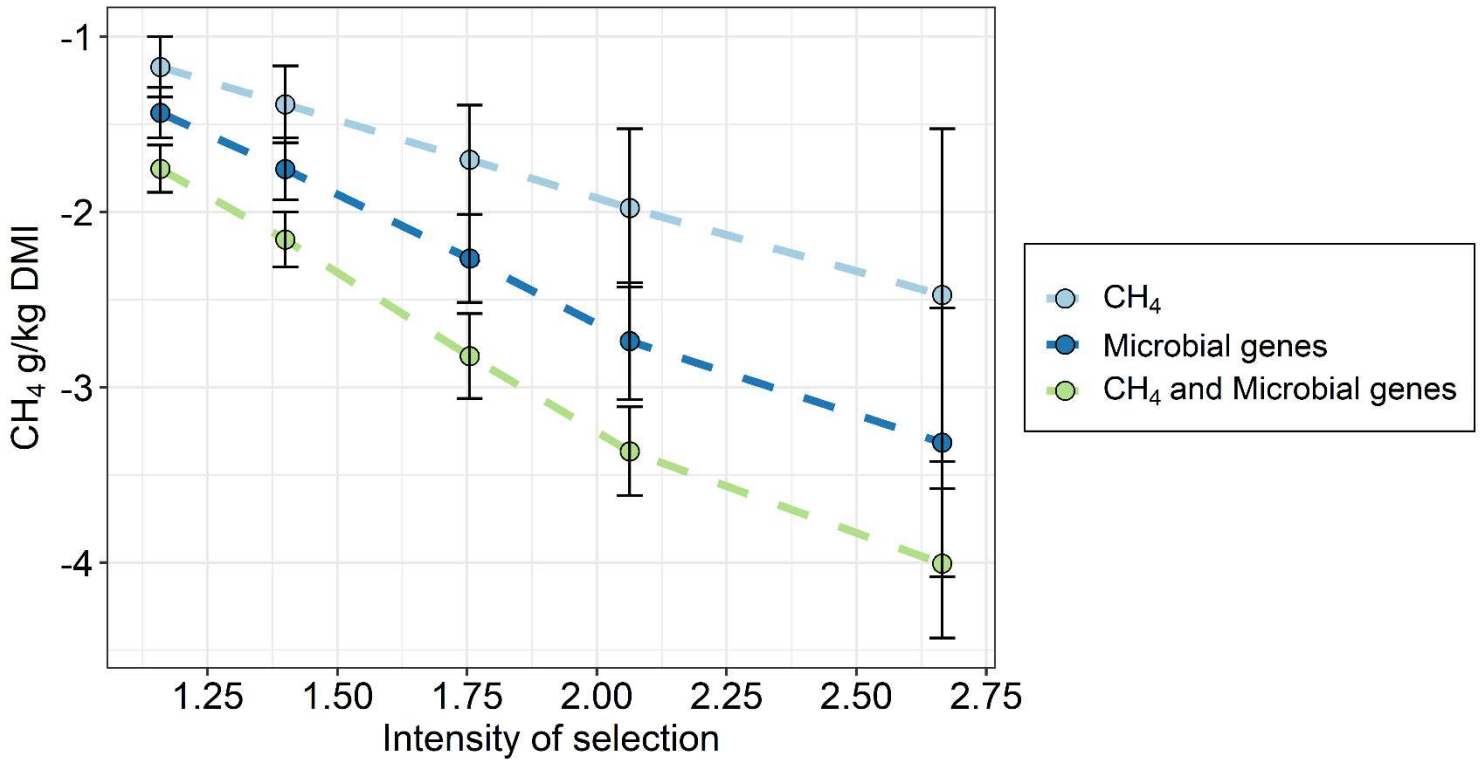


Fig. 6 | Response to selection per generation on methane (CH_4) emissions (medians and standard deviation) estimated using direct genomic selection based on measured CH_4 emissions (light blue), indirect genomic selection based on 30 microbial gene abundances most informative for host genomic selection for methane as described in Supplementary Data 2 (dark blue) or selection on both criteria (green). Intensities of selection 1.1590, 1.400, 1.755, 2.063 or 2.665 are equivalent to selecting 30%, 20%, 10%, 5% or 1%, respectively, of our population based on the above described selection criteria.

Discussion

Our study confirms that host genomics shapes part of the microbial community profile[6,12–22,24–26,122–124] and provides for the first time a comprehensive understanding into the host genomic control of complex rumen microbial functional mechanisms related to CH_4 emissions and thus gives new insight into bovine and rumen microbiome holobiont. In addition, this research will be of major importance for the mitigation of the highly potent GHG CH_4 through microbiome-driven breeding in bovines. The highlights of our findings are that we identified the host genomically affected microbial gene pathways influencing CH_4 emissions which are creating an environment in the rumen that encourages the growth of reductive acetogenic microbes limiting the excess of metabolic H_2 substrate; promoting a shift in the fermentation towards volatile fatty acids (in particular propionate) and microbial proteins yield, which are expected to lead to animals with an improved efficiency of converting feed into nutrients[34,125]; enhancing the growth of microbes that consume H_2 in alternative pathways (e.g. nitrogen fixation); diverting specific substrates required to produce methanogenic coenzymes or cofactors (coenzyme B and methanofuran) to other pathways; inhibiting methanogenic organisms (e.g. by the presence of branched amino acids or cobalt/nickel) and maintain a low ruminal pH, (sucrose metabolism) preventing gut disorders and enhancing gut health (e.g. lactate-producing bacteria and thiamine metabolism). The latter result supports our hypothesis that hosts genomically resilient to gut disorders produce less CH_4, which agrees with studies demonstrating that blocking methanogenesis has no undesirable effects on cattle health status or feed intake[83], genetically-low CH_4-emitter sheep showed greater parasite resistance[125], and high CH_4 emissions in human breath are associated with intestinal tract delay, chronic constipation[126], and obesity[127]. A further highlight of our study is that the host genome influenced the ABC transport of different metabolites (some of them in quorum sensing processes), interspecies electron transfer, sensitivity of environmental conditions, and host-microbiome interaction mechanisms associated with CH_4 emissions. These results shed light into the complex processes of methanogenesis regulated by different microbial mechanisms where communication between microbial communities and their interactions with the host plays an important role. Genetic information processes in microbiota (e.g. ribosomal biosynthesis) also had a substantial host genomic effect on CH_4 emissions, potentially reflecting different microbial community growth profiles. Our findings are complementary to studies investigating the biological mechanisms underlying host genome influence on the rumen microbial composition, such as host genomic effects on rumen size[87], muscle contraction associated with passage rate[13], ruminal pH[14] or selective absorption of volatile fatty acids[19]. In humans, the mechanisms identified are even more extensive including host-microbiome interaction, secretion of fucosylated mucus glycans in the gastrointestinal physiology and mucosa, salivary amylase, or different gastric enzymes[22,24–26,123]. Studies in bovines have elucidated host candidate genes for CH_4 emissions involved in similar mechanisms (e.g. production of saliva which helps to maintain pH levels in rumen, rate of digesta, and water passage); or with features which may also affect the microbiome (e.g. rumen size or vascular supply to the intestines)[87,128,129]. Combining our findings with those reported on biological mechanisms provides increasing evidence that the host genome shapes the rumen microbiome profile associated with CH_4 emissions.

Specifically, our results provide comprehensive insight into which part of the rumen microbiome associated with CH_4 emissions is host genomically affected by revealing strong host genomic correlations between CH_4 emissions and abundances of some cultured and uncultured microbial taxa but more importantly some functional microbial genes. The higher number of microbial genes than genera/RUGs that were host genomically correlated to CH_4 emissions could be explained by the closely defined function of those genes, e.g. being involved in producing specific substrates or mediating a specific pathway that interferes with CH_4 metabolism whilst each microbial genus expresses many microbial gene functions due to functional versatility within different species or clades classified in the same genus[11,33,86,130,131] as observed within different RUGs annotated as uncultured *Methanobrevibacter* sp.; different niche specificity; or due to horizontal transfer of genes among microbial species[132,133].

Our previous research has shown that the abundance of microbial communities, in particular their genes and interactions, are excellent biomarkers for the phenotypic prediction of CH_4 emissions[6,9,28]. The present study represents a large step further by discovering 36 heritable microbial gene abundances strongly host-genomically correlated with CH_4 emissions. Microbiome-driven (indirect) genomic selection for CH_4 emissions collectively using 30 of these microbial gene abundances resulted in our small population in substantial mitigation of CH_4 (up to 17% of its mean per generation; approximately 8% per year using genomic selection), even larger than direct genomic selection based on the accurately measured CH_4 emissions. This mitigation potential is permanent and can be cumulatively increased over generations. The selection strategy would at least partially avoid the high cost involved in measuring CH_4 emissions; and the cost-effectiveness of indirect selection could be further improved by the development of a microarray to quantify the abundances of the most informative microbial genes[134]. Another advantage of the proposed selection strategy is that it is based on host genomic correlations between microbial gene abundances and CH_4 emissions which as we discussed have biological meanings.

Methods

Animals and CH_4 emissions data. Animal experiments were conducted at the Beef and Sheep Research Centre of Scotland's Rural College (SRUC). The experiment was approved by the Animal Experiment Committee of SRUC and was conducted following the requirements of the UK Animals (Scientific Procedures) Act 1986. The data were obtained from 363 steers used in different experiments[29,30,135–137] conducted over five years. In these experiments, we tested different breeds (rotational cross from Aberdeen Angus and Limousin breeds, Charolais-crosses and pure breed Luing) and two basal diets consisting of 480:520 and 80:920 forage: concentrate ratios (DM basis) and subsequently referred to as forage and concentrate diet. Supplementary Table 9 gives the distribution of the animals across experiments, breeds, and diets. A power analysis indicated that for the given number of animals per experiment, a genetic design of sires with on average 8 progeny per sire showed the highest power to identify genetic differences between sires. CH_4 emissions were individually measured in 285 animals for 48h within six indirect open-circuit respiration chambers[30]. One week before entering the respiration chambers, the animals were housed individually in training pens, identical in size and shape to the pens inside the chambers, to allow them to adapt to being housed individually. At the time of entering the chamber, the average age of the animals was 528±38 days and the average live weight was 659±54 kg. In each experiment, the animals were allocated to the respiration chambers in a randomized design within breed and diet. Animals were fed once daily, and the weight of the feed offered and refused was recorded. CH_4 emissions were expressed as g of CH_4 / kg of dry matter intake.

Hosts genomic and metagenomics samples. For host DNA analysis, 6-10 ml of blood from the 363 steers was collected from the jugular or coccygeal vein in live animals or during slaughter in a commercial abattoir. Additional 7 blood and 23 semen samples from sires of the steers were available. Blood was stored in tubes containing 1.8mg EDTA/ ml blood and immediately frozen to -20ºC. Genomic DNA was isolated from blood samples using Qiagen QIAamp toolkit and from semen samples using Qiagen QIAamp DNA Mini Kit, according to the manufacturer's instructions. The DNA concentration and integrity was estimated with Nanodrop ND-1000 (NanoDrop Technologies). Genotyping was performed by Neogen Genomics (Ayr, Scotland, UK) using GeneSeek Genomic Profiler (GGP) BovineSNP50k Chip (GeneSeek, Lincoln, NE). Genotypes were filtered for quality control purposes using PLINK version 1.09b[138]. SNPs were removed from further analysis if they met any of these criteria: no known chromosomal location according to Illumina's maps[139], non-autosomal locations, call rates less than 95% for SNPs, deviation from Hardy-Weinberg proportions ($\chi 2$ test $P<10^{-4}$), or minor allele frequency (MAF) less than 0.05. Animals showing genotypes with a call rate lower than 90% were also removed. In total, 386 animals and 36,780 autosomal SNPs remained for the analyses.

For microbial DNA analysis, *post-mortem* digesta samples (approximately 50 ml) from 363 steers were taken at slaughter immediately after the rumen was opened to be emptied. Five ml of strained ruminal fluid was mixed with 10 ml of PBS containing glycerol (87 %) and stored at -20ºC. DNA extraction from rumen samples was carried out following the protocol from Yu and Morrison[140] based on repeated bead beating with column filtration and DNA concentrations and integrity was evaluated by the same procedure (Nanodrop ND-1000) as for blood samples. Four animals out of 363 did not yield rumen samples of sufficient quality for metagenomics analysis. DNA Illumina TruSeq libraries were prepared from genomic DNA and sequenced on Illumina HiSeq systems 2500 (samples from 8 animals), HiSeq systems 4000 (samples from 280 animals) [7,33] or NovaSeq (samples from 76 animals) by Edinburgh Genomics (Edinburgh, Scotland, UK). Paired-end reads (2 × 100 bp for Hiseq systems 2500 and 2 x 150 bp for Hiseq systems 400 and NovaSeq) were generated, resulting in between 7.8 and 47.8 GB per sample (between 26 and 159 million paired reads).

Bioinformatics. For phylogenetic annotation of rumen samples, the sequence reads of 359 samples were aligned to a database including cultured genomes from the Hungate 1000 collection[31] and Refseq genomes[32] using Kraken software[141]. From 1,178 cultured microbial genera identified, we used only those present in all the samples and with a RA > 0.001% (1,108 microbial genera) for downstream analysis, equivalent to 99.99% of the total number of counts. We used the 4,941 rumen uncultured genomes (RUGs) generated by Stewart et al[33] with sequences of 282 rumen samples included in this study to identify and quantify the abundance of uncultured species. A detailed description of the metagenomics assembly and binning process and estimation of the depth of each RUG in each sample is described in Stewart *et al.*[33]. For breeding purposes microbial taxa that are present in a large proportion of the animals are required; so we discarded those RUGs present in less than 200 animals (using a cut-off of 1X coverage) and kept 225 RUGs. RUGs coverages <1 were imputed based on a Bayesian-multiplicative replacement by using *cmultrepl* function in zCompositions package[142]. This algorithm imputes zero values from a posterior estimate of the multinomial probability assuming a Dirichlet prior distribution with default parameters for GBM method[143] and performs a multiplicative readjustment of non-zero components to respect original proportions in the composition. The 225 RUGs considered showed a mean RA ≥ 0.15%. Bioinformatic analysis for the identification of rumen microbial genes was carried out as previously described by Wallace *et al.* [144]. Briefly, to measure the abundance of known functional microbial genes whole metagenome sequencing reads were aligned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/ko.html)[145] using Novoalign (www.novocraft.com). Parameters were adjusted such that all hits were reported that were equal in quality to the best hit for each read and allowing up to a 10% mismatch across the fragment. The KEGG orthologous groups (KO) of all hits that were equal to the best hit were examined. If we were unable to resolve the read to a single KO, the read was ignored; otherwise, the read was assigned to the unique KO, the resulting KO grouping corresponding to a highly similar group of sequences. We identified 3,602 KO (also referred to as microbial genes), common in all animals. As for microbial genera, we used only core microbial genes present in all the samples and with a RA > 0.001% (1,142 microbial genes) for downstream analysis, equivalent to 96.25% of the total number of counts. We combined information of KEGG, UniProt, and Clusters of Orthologous Groups of protein databases to classify 1,141 microbial genes into classes depending on the biological processes they are involved in: CH_4 metabolism (34), metabolism other than CH_4 pathway (511), genetic information processes (329), microbial communication and host-microbiome interaction (207) and other unknown or at present poorly characterized (61).

Log-ratio transformation of metagenomic data. To describe the composition of the microbiome at the taxonomic level (cultured microbial genera and RUGs) and functional level (KO or microbial genes) we estimated their RA by dividing each microbial genus/gene (in counts) by the total sum of counts of microbial genera/genes identified in each sample (Supplementary Tables 1a, b, c). To compute host genomic analysis on

the microbial cultured genera and gene abundances, we first applied a log-ratio transformation to attenuate the spurious correlations due to their compositional nature[146]. We used additive log-ratio transformation by using a reference microbial genera/gene because of the linear independence achieved between each variable and all the variables in the composition and because the facility of its interpretation[147,148]. Assuming J denotes the number of variables in each microbial database (J=1,142 for microbial genes and 1,108 for cultured microbial genera), and J-1 all of them excluding the reference microbial genera/gene, the RA of each microbial genus/gene within a sample was transformed as follows [149]:

$$\ln(\frac{x_j}{x_{ref}}) = \ln(x_j) - \ln(x_{ref}), j = 1, \dots, J-1, \quad (1)$$

where x_j is the RA of each microbial genus/gene j and x_{ref} is the RA of a specific microbial genus/gene in the database selected as a reference. We selected the 16S rRNA gene and *Oribacterium* as reference microbial gene and microbial genus, respectively. These reference variables were selected based on the following criteria: (1) present in rumen samples of all animals; (2) highly abundant (mean RA 8.56% and 0.35%, respectively); (3) not mentioned to be associated with CH_4 emissions in previous literature; (4) low log-ratio variance so the variation mainly proceeds to the numerator (0.09 and 0.24, both located in the first quartile when ordering the microbial variables by log-ratio variance in decreasing order) and (5) reproducing the geometry of the full set of log-ratios in the original dataset shown by the estimate of the Procrustes correlation[147,150] between the geometrical space defined by all log-ratios and the one defined by the selected additive log-ratios (Procrustes correlation is 0.95 and 0.92). *Oribacterium* is a strictly anaerobic and non-spore-forming bacterial genus from the order Clostridiales and family of *Lachnospiraceae*; commonly found in the rumen of cattle[13,151] and also in the human oral cavity[152,153]. The abundance of RUGs were centred log ratio-transformed[148] as additive log ratio transformation was here hampered by the difficulty of selecting a reference RUG present in all animals. Assuming J denotes the total number of RUGs (J=225):

$$\ln\left(\frac{x_j}{[\prod_j x_j]^{\frac{1}{J}}}\right) = \ln(x_j) - \frac{1}{J}\sum_j \ln(x_j),\ j = 1, \dots, J \quad (2)$$

where x_j is the depth of each RUG j.

Estimation of host genomic parameters of CH_4 emissions and microbial traits. Genomic heritabilities (h^2) of CH_4 emissions, log-transformed microbial genera (n=1,107), RUGs (n=225) and microbial genes (n=1,141) abundances were estimated by fitting 2,474 GBLUP univariate animal models described as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zg} + \boldsymbol{e}. \quad (3)$$

Data were assumed to be conditionally distributed as:

$$\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{g}, \boldsymbol{R} \sim N(\boldsymbol{Xb} + \boldsymbol{Zg}, \boldsymbol{I}\sigma_e^2), \quad (4)$$

where $\boldsymbol{y}$ is the observed trait, $\boldsymbol{b}$ is the vector of fixed effects including a combination of breed, diet, and experiment effect, $\boldsymbol{g}$ is the random host genomic effect, $\boldsymbol{e}$ is the residual of the model, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known incidence matrices for fixed and random effects. Host genomic effects were normally distributed as:

$$\boldsymbol{g} \mid \boldsymbol{G_{RM}}, \sigma_g^2 \sim N(0, \boldsymbol{G_{RM}}\sigma_g^2). \quad (5)$$

Residuals were independently normally distributed as:

$$\boldsymbol{e} \mid \boldsymbol{I}, \sigma_e^2 \sim N(0, \boldsymbol{I}\sigma_e^2), \quad (6)$$

in which σ_g^2 and σ_e^2 are the host genomic and residual variances, $\boldsymbol{I}$ is an identity matrix of the same order as the number of data, and $\boldsymbol{G_{RM}}$ is the host genomic relationship matrix between the individuals defined as [154]:

$$\boldsymbol{G_{RM}} = \frac{\boldsymbol{W}\boldsymbol{W}'}{2\sum_n^1 p_n(1-p_n)}, \quad (7)$$

where $\boldsymbol{W}$ contains genotypes adjusted for allele frequency, and p_n is the allele frequency for marker n in the whole genotyped population. Host genomic and residual effects were assumed to be uncorrelated between them. Host genomic correlations (r_{gCH4}) among CH_4 emissions and log-transformed abundances of microbial genera, RUGs and microbial genes were estimated by fitting 2,473 GBLUP bivariate animal models including the same effects as (3). Host genomic effects were distributed as:

$$g \mid \boldsymbol{G_{RM}}, \boldsymbol{G_0} \sim N(0, \boldsymbol{G_{RM}} \otimes \boldsymbol{G_0}), \quad (8)$$

and residuals as:

$$\boldsymbol{R} = e \mid \boldsymbol{R_0} \sim N(0, \boldsymbol{I} \otimes \boldsymbol{R_0}), \quad (9)$$

where $\boldsymbol{G_0}$ and $\boldsymbol{R_0}$ are the 2 x 2 host genomic and residual (co)variance matrices between CH_4 emissions and each microbial genus, RUG or microbial gene, $\boldsymbol{I}$ is an identity matrix of the same order as the number of individuals with data. Bayesian statistics were used[155], assuming bounded flat priors for all unknowns. Analyses were computed using the THRGIBBSF90 program[156]. Results were based on Markov chain Monte Carlo chains consisting of 1,000,000 iterations, with a burn-in period of 200,000, and to reduce autocorrelations only 1 of every 100 samples was saved for inferences. In all analyses, convergence was tested using the POSTGIBBSF90[156] program by calculating the Z criterion of Geweke (varying between -0.05 and 0.05 in univariate and -0.09 and 0.1 in bivariate models). Monte Carlo sampling errors were computed using time-series procedures and checked to be at least 10 times lower than the standard deviation of the marginal posterior distribution. As h^2 estimates we used the median of its marginal posterior distribution of CH_4, each microbial genus, RUG or microbial gene and the highest posterior density interval at 95% probability ($HPD_{95\%}$). To test the significance of h^2 estimates, we computed the *P* of host genomic effects by conducting a likelihood-ratio test. For each case, an upper-tailed test was computed assuming the univariate model without host genomic effect as the null hypothesis and the complete model as the tested hypothesis. *P* were obtained by fitting the differences between deviances of the complete and reduced model as a Chi-Square distribution with 1 degree of freedom. We accounted for multiple testing by setting a significance $P \leq 0.05$ threshold corrected by Bonferroni procedures[157] ($0.05/2{,}474 = 2.02 \times 10^{-5}$). Additionally, we considered microbial genes with h^2 estimates < 0.20 being lowly heritable, $0.20 < h^2 < 0.40$ being moderately heritable and h^2 estimates > 0.40 being highly heritable. As estimate for the host genomic correlations, we used the median of its marginal posterior distribution and the $HPD_{95\%}$. To investigate the confidence level of r_{gCH4} we estimated the posterior probability of r_{gCH4} being >0 when the median of the correlation was positive or <0 when the median was negative (P_0). We only considered significant those r_{gCH4} estimates with (P_0) ≥ 0.95.

To predict the impact of indirect selection for reduced CH_4 emissions using microbial genera/genes significantly ($P_0 \geq 0.95$) host genomically correlated with CH_4 emissions, we estimated the marginal posterior distribution of the correlated response in CH_4 emissions after host genomic selection for each of these microbial genera/genes, considering only the own performance of each individual[158]:

$$\boldsymbol{R_{CH4j}} = i\, \boldsymbol{h_j r_{gCH4j} \sigma_{aCH4}}\,, \tag{10}$$

where $\boldsymbol{R_{CH4j}}$ presents the selection response in CH_4 emissions after selection for the abundance of each microbial genus/gene *j*, i is the intensity of selection considered to be 1.159 (equivalent to 30% of our cattle population being selected based on the selection criterion), $\boldsymbol{h_j}$ is the marginal posterior distribution of the square root of the h^2 estimate of the microbial genus/gene from univariate analyses , and $\boldsymbol{r_{gCH4j}}$ is the marginal posterior distribution of the host genomic correlation between CH_4 emissions and microbial genus/gene *j* from bivariate models. The median, standard deviation and the probability (P_0) of the correlated response to selection to be higher (lower) than 0 when the correlated response was positive (negative) were computed.

Co-abundance network analysis of host genomic effects on rumen microbiome. To study the correlation structure among host genomic effects of the log-transformed abundances of 1,107 microbial genera, 225 RUGs and 1,141 microbial genes, we built a co-abundance network analysis using deregressed host genomic effects (dGEBVs) for all microbial traits. Deregressed host genomic effects were calculated from previously described univariate GBLUP models by using ACCF90 and DEPROOF90 programs[156]. Co-abundance network (Graphia software[159]) connected or edged microbial traits (nodes) based on a Pearson correlation >0.70 among their dGEBVs. The complexity of the graph was reduced by discarding nodes with a minimum number of incident edges (referred to as node degree) of 2, i.e., only those microbial traits Pearson-correlated (>0.70) with at least other 2 microbial traits were kept. The total number of microbial genera, RUGs and microbial genes included in the network was 2,129 out of the 2,473 tested. The number of edges of each node was reduced by ranking the edges based on k-nearest neighbour algorithm and retaining only the 80% of them. The software applies Markov Clustering algorithm by a flow simulation model[160] to find discrete groups of nodes (clusters) based on their position within the overall topology of the graph. The granularity of the clusters, i.e., the minimum number of nodes that a cluster has to contain, was set to 2 nodes. The network showed 106 clusters, but only those 12 clusters including ≥ 3 methanogenic archaea genera, RUGs and microbial genes involved in CH_4 metabolism pathway according to KEGG[145] database or microbial genera/RUGs/genes host-genomically correlated with CH_4 emissions ($P_0 \geq 0.95$) were studied in-depth.

Enrichment analysis of microbial gene abundances in RUGs. To identify which of the 225 RUGs were carrying the microbial genes (KO) demonstrating a r_{gCH4} with a confidence level $P_0 \geq 0.95$, an enrichment analysis was performed by counting the number of unique proteins clustered in the 115 microbial genes mapped in each of the 225 RUGs.

Identification of most informative microbial traits to predict CH_4 emission breeding values and maximize response to selection. Only microbial variables present in all animals, showing a RA ≥0.01%, with significant h^2 ($P \leq 2.02 \times 10^{-5}$), and host-genomically correlated with CH_4 emissions ($P_0 \geq 0.95$) were considered for breeding purposes. Four microbial genera and 36 microbial genes met these conditions. Due to computation reasons, only 30 microbial gene abundances were carried forward for downstream analysis. To use microbial gene information to select hosts emitting less CH_4, the estimation between their host genomic and residual (co)variance matrices was required. Host genomic and residual (co)variances among the 30 selected microbial gene abundances were estimated using 435 bivariate analyses. Bivariate analyses fitted the same model as previously described for estimation of r_{gCH4} with same assumptions (eq. 8,9). Results were based on Markov chain Monte Carlo chains consisting of 1,000,000 iterations, with a burn-in period of 200,000, and only 1 of every 100 samples was saved for inferences. Convergence was tested with POSTGIBBSF90 program by checking Z criterion of Geweke to be between -0.12 and 0.15. Monte Carlo sampling errors were computed using time-series procedures and checked to be at least 10 times lower than the standard deviation of the posterior marginal distribution[155]. The 31 x 31 host genomic and residual variance-covariance matrices, including CH_4 emissions and the 30 microbial genes were build based on medians of the estimated variance components from the bivariate analyses and mean across all previous bivariate models for host genomic and residual variances of CH_4 emissions. Both matrices needed bending to be positive definite (tolerance for minimum eigenvalues=0.001). The difference between original and bent matrices was never higher than the posterior standard error of the corresponding parameters.

Estimation of the selection response of CH_4 emissions based on different sources of information. We analysed three different scenarios to predict host-genomic effects of CH_4 emissions: (i) by using measured CH_4 emissions only, (ii) by using the 30 microbial gene abundances only, and (iii) by using a combination of both, measured CH_4 emissions and the 30 microbial gene abundances. The three scenarios were computed with data from 285 animals with CH_4 emissions and metagenomics information. All scenarios were calculated by GBLUP analysis assuming as fixed variance components the previously estimated 31 x 31 host genomic and residual variance-covariance matrices of the traits after bending. Scenario (i) was performed using a univariate GBLUP analysis including only measured CH_4 emissions; scenario (ii) was computed by fitting a multivariate GBLUP model including the 30 microbial gene abundances host-genomically correlated to CH_4 emissions (using measured CH_4 emissions as missing value[161]); and scenario (iii) considers besides the abundance of the 30 microbial genes, the measured CH_4 emission values in the GBLUP analysis. In all cases, models included the same effects as in (3). Host genomic values estimates for CH_4 emissions were based on Markov chain Monte Carlo chains consisting of 100,000 iterations, with a burn-in period of 20,000, and to reduce autocorrelation only 1 of every 100 samples was saved for inferences. Response to selection was estimated as the marginal posterior distributions of the difference between the mean of CH_4 emissions host genomic values of all animals with data and the mean of selected animals when alternatively, 1%, 5%, 10%, 20%, 30%, 40%, and 50% of our population were selected. The mean accuracy of the CH_4 emissions genomic values in each scenario was estimated as the average of the individual accuracies:

$$Accuracy_i = \sqrt{1 - \frac{sd_i^2}{g_{RM_{ii}} G_{CH4CH4}}} \quad , \tag{11}$$

where sd_i is the standard deviation of the posterior marginal distribution of the host genomic value for animal i and $g_{RM_{ii}}$ is the $\boldsymbol{G_{RM}}$ diagonal element for animal i.

Data availability. Metagenomic sequence reads for all rumen samples are available under European Nucleotide Archive (ENA) under accession projects PRJEB31266, PRJEB21624, PRJEB10338. The genotpes of the host animals are readily available from the authors.

Code availability. Metagenomic data processing was carried out using Kraken (https://ccb.jhu.edu/software/kraken/) for taxonomic annotation and Novoalign (http://www.novocraft.com/support/download/ available under license) for functional annotation. SNP data filtering was performed PLINK (https://www.cog-genomics.org/plink2). Host genomic analysis were carried out using RENUMF90, THRGIBBSF90, POSTGIBBSF90, ACCF90, DEPROOF90 software which have free access in http://nce.ads.uga.edu/wiki/doku.php?id=application_programs, except for ACCF90 and DEPROOF90 available only under research agreement. Network analysis was carried out by free access to Graphia software whose code source can be found at https://graphia.app/download.html.

References

1. OECD/FAO. *OECD-FAO Agricultural Outlook 2020-2029*. *OECD Publishing Paris, /Food and Agriculture Organization of the United Nations, Rome* (2020). doi:10.1787/1112c23b-en
2. Vollset, S. E. *et al.* Fertility, mortality, migration, and population scenarios for 195 countries and

territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *Lancet* **396**, 1285–1306 (2020).
3. Gerber, P. J. *et al. Tackling climate change through livestock – A global assessment of emissions and mitigation opportunities*. *Food and Agriculture Organization of the United Nations (FAO), Rome* (2013).
4. Myhre, G. *et al.* Anthropogenic and Natural Radiative Forcing: Supplementary Material. *Clim. Chang. 2013 Phys. Sci. Basis. Contrib. Work. Gr. I to Fifth Assess. Rep. Intergov. Panel Clim. Chang. Phys. Sci. Basis. Contrib. Work. Gr. I to Fifth Assess. Rep.* 1–44 (2013).
5. Johnson, K. A. & Johnson, D. E. Methane emissions from cattle. *J. Anim. Sci.* **73**, 2483–2492 (1995).
6. Roehe, R. *et al.* Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance. *PLoS Genet.* **12**, 1–20 (2016).
7. Stewart, R. D. *et al.* Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9**, 1–11 (2018).
8. Wallace, R. J. *et al.* Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci. Rep.* **4**, 5892 (2015).
9. Martínez-Álvaro, M. *et al.* Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. *Front. Microbiol.* **11**, 1–13 (2020).
10. Barrett, K., Jensen, K., Meyer, A. S., Frisvad, J. C. & Lange, L. Fungal secretome profile categorization of CAZymes by function and family corresponds to fungal phylogeny and taxonomy: Example Aspergillus and Penicillium. *Sci. Rep.* **10**, 1–12 (2020).
11. Tapio, I., Snelling, T. J., Strozzi, F. & Wallace, R. J. The ruminal microbiome associated with methane emissions from ruminant livestock. *J. Anim. Sci. Biotechnol.* **8**, 7 (2017).
12. Difford, G. F. *et al.* Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet.* **14**, 1–22 (2018).
13. Zhang, Q. *et al.* Bayesian modeling reveals host genetics associated with rumen microbiota jointly influence methane emission in dairy cows. *ISME J.* **14**, 2019–2033 (2020).
14. Li, F. *et al.* Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* **7**, 1–17 (2019).
15. Wallace, J. R. *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* **5**, eaav8391 (2019).
16. Saborío-Montero, A. *et al.* Structural equation models to disentangle the biological relationship between microbiota and complex traits: Methane production in dairy cattle as a case of study. *J. Anim. Breed. Genet.* **137**, 36–48 (2020).
17. Sasson, G. *et al.* Heritable bovine rumen bacteria are phylogenetically related and correlated with the cow's capacity to harvest energy from its feed. *MBio* **8**, 1–12 (2017).
18. Weimer, P. J., Stevenson, D. M., Mantovani, H. C. & Man, S. L. C. Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. *J. Dairy Sci.* **93**, 5902–5912 (2010).
19. Abbas, W. *et al.* Influence of host genetics in shaping the rumen bacterial community in beef cattle. *Sci. Rep.* **10**, 15101 (2020).
20. Bergamaschi, M. *et al.* Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci. Rep.* **10**, 1–12 (2020).
21. Chen, C. *et al.* Contribution of Host Genetics to the Variation of Microbial Composition of Cecum Lumen and Feces in Pigs. *Front. Microbiol.* **9**, 1–13 (2018).
22. Poole, A. C. *et al.* Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes. *Cell Host Microbe* **25**, 553-564.e7 (2019).
23. Kurilshikov, A. *et al.* Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165 (2021).
24. Turpin, W. *et al.* Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
25. Qin, Y. *et al.* Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population. *medRxiv* (2020). doi:https://doi.org/10.1101/2020.09.12.20193045
26. Hughes, D. A. *et al.* Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. *Nat. Microbiol.* **5**, 1079–1087 (2020).
27. Goodrich, J. K. *et al.* Human Genetics Shape the Gut Microbiome. *Cell* **159**, 789–799 (2014).
28. Auffret, M. D. *et al.* Identification, comparison, and validation of robust rumen microbial biomarkers for methane emissions using diverse Bos Taurus breeds and basal diets. *Front. Microbiol.* **8**, 1–15 (2018).

29. Duthie, C. A. *et al.* The impact of divergent breed types and diets on methane emissions, rumen characteristics and performance of finishing beef cattle. *Animal* **11**, 1762–1771 (2017).
30. Rooke, J. A. *et al.* Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. *Br. J. Nutr.* **112**, 398–407 (2014).
31. Seshadri, R. *et al.* Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* **36**, 359–367 (2018).
32. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a cuarted non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
33. Stewart, R. D. *et al.* Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019).
34. Joblin, K. N. Ruminal acetogens and their potential to lower ruminant methane emissions. *Aust. J. Agric. Res.* **50**, 629–50 (1999).
35. McAllister, T. A. & Newbold, C. J. Redirecting rumen fermentation to reduce methanogenesis. *Aust. J. Exp. Agric.* **48**, 7–13 (2008).
36. Cottle, D. J., Nolan, J. V. & Wiedemann, S. G. Ruminant enteric methane mitigation: A review. *Anim. Prod. Sci.* **51**, 491–514 (2011).
37. Hegarty, R. S. Mechanisms for competitively reducing ruminal methanogenesis. *Aust. J. Agric. Res.* **50**, 629–50 (1999).
38. Greening, C. *et al.* Diverse hydrogen production and consumption pathways influence methane production in ruminants. *ISME J.* **13**, 2617–2632 (2019).
39. Hajarnis, S. R. & Ranade, D. R. Inhibition of methanogens by n- and iso-volatile fatty acids. *World J. Microbiol. Biotechnol.* **10**, 350–351 (1994).
40. Henderson, C. The effects of fatty acids on pure cultures of rumen bacteria. *J. Agric. Sci.* **81**, 107–112 (1973).
41. Paulo, L. M., Ramiro-Garcia, J., van Mourik, S., Stams, A. J. M. & Sousa, D. Z. Effect of nickel and cobalt on methanogenic enrichment cultures and role of biogenic sulfide in metal toxicity attenuation. *Front. Microbiol.* **8**, 1–12 (2017).
42. Zhou, M., Chen, Y. & Guan, L. L. Rumen Bacteria. in *Rumen Microbiology: From Evolution to Revolution* (eds. Puniya, A. K., Singh, R. & Kamra, D. N.) 79–95 (Springer, 2015). doi:10.1007/978-81-322-2401-3_6
43. van Wolferen, M., Orell, A. & Albers, S. V. Archaeal biofilm formation. *Nat. Rev. Microbiol.* **16**, 699–713 (2018).
44. Chen, H. & Fink, G. R. Feedback control of morphogenesis in fungi by aromatic alcohols. *Genes Dev.* **20**, 1150–1161 (2006).
45. Thauer, R. K. Anaerobic oxidation of methane with sulfate: On the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO2. *Curr. Opin. Microbiol.* **14**, 292–299 (2011).
46. McInerney, M. J., Sieber, J. R. & Gunsalus, R. P. Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol.* **20**, 623–632 (2009).
47. McInerney, M. J. *et al.* Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Ann. N. Y. Acad. Sci.* **1125**, 58–72 (2008).
48. Evans, P. N. *et al.* An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* **17**, 219–232 (2019).
49. Nomura, M., Gourse, R. & Baughman, G. Regulation of the synthesis of ribosomes and ribosomal components. *Ann. Rev. Biochem.* **53**, 75–117 (1984).
50. Martin, C., Morgavi, D. P. & Doreau, M. Methane mitigation in ruminants: from microbe to the farm scale. *Animal* **4**, 351–365 (2010).
51. Roque, B. M. *et al.* Red seaweed (Asparagopsis taxiformis) supplementation reduces enteric methane by over 80 percent in beef steers. *bioRxiv* (2020). doi:10.1101/2020.07.15.204958
52. Dijkstra, J., Bannink, A., France, J., Kebreab, E. & van Gastelen, S. Short communication: Antimethanogenic effects of 3-nitrooxypropanol depend on supplementation dose, dietary fiber content, and cattle type. *J. Dairy Sci.* **101**, 9041–9047 (2018).
53. Hristov, A. N. *et al.* Special topics-Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. *J. Anim. Sci.* **91**, 5045–5069 (2013).
54. Garnsworthy, P. C. *et al.* Comparison of methods to measure methane for use in genetic evaluation of dairy cattle. *Animals* **9**, 1–12 (2019).
55. Manzanilla-Pech, C. I. V. *et al.* Genomewide association study of methane emissions in angus beef cattle with validation in dairy cattle. *J. Anim. Sci.* **94**, 4151–4166 (2016).
56. Hayes, B. J. *et al.* Genomic heritabilities and genomic estimated breeding values for methane traits in

Angus cattle. *J. Anim. Sci.* **94**, 902–908 (2016).
57. Donoghue, K. A., Bird-Gardiner, T., Arthur, P. F., Herd, R. M. & Hegarty, R. F. Genetic and phenotypic variance and covariance components for methane emission and postweaning traits in Angus cattle. *J. Anim. Sci.* **94**, 1438–1445 (2016).
58. Cesarani, A. *et al.* Beef trait genetic parameters based on old and recent data and its implications for genomic predictions in Italian Simmental cattle. *J. Anim. Sci.* **98**, 1–8 (2020).
59. Gengler, N., Wiggans, G. R. & Gillon, A. Adjustment for heterogeneous covariance due to herd milk yield by transformation of test-day random regressions. *J. Dairy Sci.* **88**, 2981–2990 (2005).
60. Gunsalus, R. P. *et al.* Complete genome sequence of Methanospirillum hungatei type strain JF1. *Stand. Genomic Sci.* **11**, 1–10 (2016).
61. Henderson, G. *et al.* Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* **5**, 14567 (2015).
62. Snelling, T. J. *et al.* Temporal stability of the rumen microbiota in beef cattle, and response to diet and supplements. *Anim. Microbiome* **1**, 1–14 (2019).
63. Morgavi, D. P., Forano, E., Martin, C. & Newbold, C. J. Microbial ecosystem and methanogenesis in ruminants. *Animal* **4**, 1024–1036 (2010).
64. Uddin MJ, K. Dynamics of microbial protein synthesis in the rumen - A Review. *Ann. Vet. Anim. Sci.* **2**, 116–131 (2015).
65. Lovendahll, P. *et al.* Review: Selecting for improved feed efficiency and reduced methane emissions in dairy cattle. *Animal* **12**, S336–S349 (2018).
66. Tobin, C. Removal and replacement of ribosomal proteins. (Uppsala University, 2011).
67. Liu, X. *et al.* M-GWAS for the Gut Microbiome in Chinese Adults Illuminates on Complex Diseases. *bioRxiv* (2019). doi:https://doi.org/10.1101/736413
68. Vanwonterghem, I. *et al.* Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
69. Graham, D. E., Xu, H. & White, R. H. Identification of the 7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase required for coenzyme F420 biosynthesis. *Arch. Microbiol.* **180**, 455–464 (2003).
70. Grochowski, L. L. & White, R. H. Biosynthesis of the methanogenic coenzymes. *Compr. Nat. Prod. II Chem. Biol.* **7**, 711–748 (2010).
71. Peng, X. *et al.* Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat. Microbiol.* (2021). doi:10.1038/s41564-020-00861-0
72. Miyazaki, J., Kobashi, N., Nishiyama, M. & Yamane, H. Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through α-aminoadipate. *J. Bacteriol.* **183**, 5067–5073 (2001).
73. Andries, J. I., Buysse, F. X., De Brabander, D. L. & Cottyn, B. G. Isoacids in ruminant nutrition: Their role in ruminal and intermediary metabolism and possible influences on performances - A review. *Anim. Feed Sci. Technol.* **18**, 169–180 (1987).
74. Drevland, R. M., Waheed, A. & Graham, D. E. Enzymology and evolution of the pyruvate pathway to 2-oxobutyrate in Methanocaldococcus jannaschii. *J. Bacteriol.* **189**, 4391–4400 (2007).
75. Lee, J. H. & Lee, J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* **34**, 426–444 (2010).
76. Roager, H. M. & Licht, T. R. Microbial tryptophan catabolites in health and disease. *Nat. Commun.* **9**, 1–10 (2018).
77. Drevland, R. M., Jia, Y., Palmer, D. R. J. & Graham, D. E. Methanogen homoaconitase catalyzes both hydrolyase reactions in coenzyme B biosynthesis. *J. Biol. Chem.* **283**, 28888–28896 (2008).
78. Neill, a R., Grime, D. W. & Dawson, R. M. Conversion of choline methyl groups through trimethylamine into methane in the rumen. *Biochem. J.* **170**, 529–535 (1978).
79. Janssen, P. H. & Kirs, M. Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* **74**, 3619–3625 (2008).
80. Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* **1125**, 171–189 (2008).
81. Chamberlain, D. G. & Thomas, P. C. The effect of supplemental methionine and inorganic sulphate on the ruminal digestion of grass silage in sheep. *J. Sci. Food Agric.* **34**, 440–446 (1983).
82. Kamke, J. *et al.* Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* **4**, 1–16 (2016).
83. Yanibada, B. *et al.* Inhibition of enteric methanogenesis in dairy cows induces changes in plasma metabolome highlighting metabolic shifts and potential markers of emission. *Sci. Rep.* **10**, 1–14 (2020).
84. Pinares-Patiño, C. S. *et al.* Heritability estimates of methane emissions from sheep. *Animal* **7**, 316–321

(2013).
85. Stewart, C. S., Flint, H. J. & Bryant, M. P. The rumen bacteria. in *The Rumen Microbial Ecosystem* (eds. Hobson, P. N. & Stewart, C. S.) 10–72 (Blackie academic and professional, 1997). doi:10.1007/978-94-009-1453-7_2
86. Kittelmann, S. *et al.* Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS One* **9**, 1–9 (2014).
87. Goopy, J. P. *et al.* Low-methane yield sheep have smaller rumens and shorter rumen retention time. *Br. J. Nutr.* **111**, 578–585 (2014).
88. Strobel, H. J. & Russell, J. B. Effect of pH and Energy Spilling on Bacterial Protein Synthesis by Carbohydrate-Limited Cultures of Mixed Rumen Bacteria. *J. Dairy Sci.* **69**, 2941–2947 (1986).
89. Herrmann, K. M. & Weaver, L. M. The shikimate pathway. *Annu. Rev. Plant Biol.* **50**, 473–503 (1999).
90. Hall, M. B. & Herejk, C. Differences in yields of microbial crude protein from in vitro fermentation of carbohydrates. *J. Dairy Sci.* **84**, 2486–2493 (2001).
91. Nollet, L. & Verstraete, W. Gastro-enteric methane versus sulphate and volatile fatty acid production. *Environ. Monit. Assess.* **42**, 113–131 (1996).
92. Demeyer, D., De Graave, K., Durand, M. & Stevani, J. Acetate: a hydrogen sink in hindgut fermentation as opposed to rumen fermentation. *Acta Vet Scabd Suppl.* **86**, 68–75 (1989).
93. Lopez, S., Mcintosh, F. M., Wallace, R. J. & Newbold, C. J. Effect of adding acetogenic bacteria on methane production by mixed rumen microorganisms. **78**, 1–9 (1999).
94. Baldwin, R. L., Wood, W. A. & Emery, R. S. Conversion of lactate-c'4 to propionate by the rumen microflora. *J. Bacteriol.* **83**, 907–913 (1961).
95. Janssen, P. H. Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Anim. Feed Sci. Technol.* **160**, 1–22 (2010).
96. Doyle, N. *et al.* Use of Lactic Acid Bacteria to Reduce Methane Production in Ruminants, a Critical Review. *Front. Microbiol.* **10**, 2207 (2019).
97. Ugulava, N. B., Sacanell, C. J. & Jarrett, J. T. Spectroscopic changes during a single turnover of biotin synthase: Destruction of a [2Fe-2S] cluster accompanies sulfur insertion. *Biochemistry* **40**, 8352–8358 (2001).
98. Howell, D. M., Harich, K., Xu, H. & White, R. H. α-keto acid chain elongation reactions involved in the biosynthesis of coenzyme B (7-mercaptoheptanoyl threonine phosphate) in methanogenic archaea. *Biochemistry* **37**, 10108–10117 (1998).
99. Widdel, F. Growth of methanogenic bacteria in pure culture with 2-propanol and other alcohols as hydrogen donors. *Appl. Environ. Microbiol.* **51**, 1056–1062 (1986).
100. Moore, S. J. *et al.* Elucidation of the biosynthesis of the methane catalyst coenzyme F430. *Nature* **543**, 78–82 (2017).
101. Bulen, W. A. & LeComte, J. R. The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **56**, 979–986 (1966).
102. Wang, M., Wang, H., Zheng, H., Dewhurst, R. J. & Roehe, R. A heat diffusion multilayer network approach for the identification of functional biomarkers in rumen methane emissions. *Methods* (2020). doi:10.1016/j.ymeth.2020.09.014
103. Jenkins, T. C., Abughazaleh, A. A., Freeman, S. & Thies, E. J. The production of 10-hydroxystearic and 10-ketostearic acids is an alternative route of oleic acid transformation by the ruminal microbiota in cattle. *J. Nutr.* **136**, 926–931 (2006).
104. Abe, A. & Sugiyama, K. Growth inhibition and apoptosis induction of human melanoma cells by omega-hydroxy fatty acids. *Anticancer. Drugs* **16**, 543–549 (2005).
105. Martin, A. & Daniel, J. The ABC transporter Rv1272c of Mycobacterium tuberculosis enhances the import of long-chain fatty acids in Escherichia coli. *Biochem. Biophys. Res. Commun.* **496**, 667–672 (2018).
106. Jenkins, B., West, J. A. & Koulman, A. A review of odd-chain fatty acid metabolism and the role of pentadecanoic acid (C15:0) and heptadecanoic acid (C17:0) in health and disease. *Molecules* **20**, 2425–2444 (2015).
107. Jenkins, T. C. Lipid Metabolism in the Rumen. *J. Dairy Sci.* **76**, 3851–3863 (1993).
108. Leng, R. A. Interactions between microbial consortia in biofilms: A paradigm shift in rumen microbial ecology and enteric methane mitigation. *Anim. Prod. Sci.* **54**, 519–543 (2014).
109. Won, M. Y., Oyama, L. B., Courtney, S. J., Creevey, C. J. & Huws, S. A. Can rumen bacteria communicate to each other? *Microbiome* **8**, 1–8 (2020).
110. Patra, A., Park, T., Kim, M. & Yu, Z. Rumen methanogens and mitigation of methane emission by anti-methanogenic compounds and substances. *J. Anim. Sci. Biotechnol.* **8**, 1–18 (2017).

111. Węgrzyn, A., Taylor, K. & Węgrzyn, G. The cbpA chaperone gene function compensates for dnaJ in λ plasmid replication during amino acid starvation of Escherichia coli. *J. Bacteriol.* **178**, 5847–5849 (1996).
112. Wahlström, A., Sayin, S. I., Marschall, H. U. & Bäckhed, F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab.* **24**, 41–50 (2016).
113. Ramírez-Pérez, O., Cruz-Ramón, V., Chinchilla-López, P. & Méndez-Sánchez, N. The Role of the Gut Microbiota in Bile Acid Metabolism. *Ann. Hepatol.* **16**, S21–S26 (2017).
114. Immig, I. The effect of porcine bile acids on methane production by rumen contents in vitro. *Arch. Anim. Nutr.* **51**, 21–26 (1998).
115. Hooper, L. V. & Gordon, J. I. Glycans as legislators of host-microbial interactions: Spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* **11**, 1–10 (2001).
116. Hoorens, P. R. *et al.* Genome wide analysis of the bovine mucin genes and their gastrointestinal transcription profile. *BMC Genomics* **12**, 140 (2011).
117. Aschenbach, J. R., Penner, G. B., Stumpff, F. & Gäbel, G. Ruminant nutrition symposium: Role of fermentation acid absorption in the regulation of ruminal pH. *J. Anim. Sci.* **89**, 1092–1107 (2011).
118. Lee, M., Jeong, S., Seo, J. & Seo, S. Changes in the ruminal fermentation and bacterial community structure by a sudden change to a high-concentrate diet in Korean domestic ruminants. *Asian-Australasian J. Anim. Sci.* **32**, 92–102 (2019).
119. Van Kessel, J. A. S. & Russell, J. B. The effect of pH on ruminal methanogenesis. *FEMS Microbiol. Ecol.* **20**, 205–210 (1996).
120. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D. & Poch, O. Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. *Nucleic Acids Res.* **30**, 5382–5390 (2002).
121. Smith, T. F., Lee, J. C., Gutell, R. R. & Hartman, H. The origin and evolution of the ribosome. *Biol. Direct* **3**, 1–13 (2008).
122. Snijders, A. M. *et al.* Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nat. Microbiol.* **2**, 1–8 (2016).
123. Kurilshikov, A. *et al.* Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165 (2021).
124. Zhu, W., Lin, Y., Liao, H. & Wang, Y. Selection of reference genes for gene expression studies related to intramuscular fat deposition in Capra hircus skeletal muscle. *PLoS One* **10**, e0121280 (2015).
125. Rowe, S. J. *et al.* Selection for divergent methane yield in New Zealand sheep - A ten year perspective. *Proc. Assoc. Adv. Anim. Breed. Genet.* 306–309 (2019).
126. Triantafyllou, K., Chang, C. & Pimentel, M. Methanogens, methane and gastrointestinal motility. *J. Neurogastroenterol. Motil.* **20**, 31–40 (2014).
127. Mathur, R. *et al.* Methane and hydrogen positivity on breath test associated to obesity. *J Clin Endocrinol Metab.* **98**, E698–E702 (2013).
128. Pszczola, M., Strabel, T., Mucha, S. & Sell-Kubiak, E. Genome-wide association identifies methane production level relation to genetic control of digestive tract development in dairy cows. *Sci. Rep.* **8**, 1–11 (2018).
129. Maekawa, M., Beauchemin, K. A. & Christensen, D. A. Effect of concentrate level and feeding management on chewing activities, saliva production, and ruminal pH of lactating dairy cows. *J. Dairy Sci.* **85**, 1165–1175 (2002).
130. Danielsson, R. Methane Production in Dairy Cows Impact of Feed and Rumen Microbiota. (Acta universitatis agriculturae sueciae, 2016).
131. Poehlein, A., Schneider, D., Soh, M., Daniel, R. & Seedorf, H. Comparative genomic analysis of members of the genera methanosphaera and methanobrevibacter reveals distinct clades with specific potential metabolic functions. *Archaea* **2018**, 1–9 (2018).
132. Ricard, G. *et al.* Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* **7**, 1–13 (2006).
133. Klieve, A. V. *et al.* Naturally occurring DNA transfer system associated with membrane vesicles in cellulolytic Ruminococcus spp. of ruminal origin. *Appl. Environ. Microbiol.* **71**, 4248–4253 (2005).
134. Hess, M. K. *et al.* A restriction enzyme reduced representation sequencing approach for low-cost, high-throughput metagenome profiling. *PLoS One* **15**, 1–18 (2020).
135. Duthie, C.-A. *et al.* Impact of adding nitrate or increasing the lipid content of two contrasting diets on blood methaemoglobin and performance of two breeds of finishing beef steers. *Animal* **10**, 786–795 (2016).
136. Duthie, C. A. *et al.* The effect of dietary addition of nitrate or increase in lipid concentrations, alone or in combination, on performance and methane emissions of beef cattle. *Animal* **12**, 280–287 (2018).
137. Somarriba, M. *et al.* The effects of a composite chronic stress treatment on fear responses and attention

bias in beef cattle. in *ISAE 2019. Proceedings of the 53rd Congress of the ISAE* **53**, 333 (2019).
138. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
139. Matukumalli, L. K. *et al.* Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS One* **4**, e5350 (2009).
140. Yu, Z. & Morrison, M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, 808–812 (2004).
141. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, 1–12 (2014).
142. Palarea-Albaladejo, J. & Martín-Fernández, J. A. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* **143**, 85–96 (2015).
143. Martín-Fernández, J. A., Hron, K., Templ, M., Filzmoser, P. & Palarea-Albaladejo, J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Modelling* **15**, 134–158 (2015).
144. Wallace, R. J. *et al.* The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* **16**, 839 (2015).
145. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
146. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
147. Greenacre, M. Variable Selection in Compositional Data Analysis Using Pairwise Logratios. *Math. Geosci.* **51**, 649–682 (2018).
148. Greenacre, M. *Compositional data analysis in practise*. *CRC Press* (2019).
149. Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B(methodological)* **44**, 139–177 (1982).
150. Greenacre, M. Compositional data analysis. *Annu. Rev. Stat. Its Appl.* **8**, 271–299 (2021).
151. Zeng, H., Guo, C., Sun, D., Seddik, H. E. & Mao, S. The ruminal microbiome and metabolome alterations associated with diet-induced milk fat depression in dairy cows. *Metabolites* **9**, 154 (2019).
152. Kang, S., Denman, S. & McSweeney, C. Draft Genome Sequence and Annotation of Oribacterium sp. Strain C9, Isolated from a Cattle Rumen. *Microbiol. Resour. Announc.* **8**, e01562-18 (2019).
153. Iwasawa, K. *et al.* Dysbiosis of the salivary microbiota in pediatric-onset primary sclerosing cholangitis and its potential as a biomarker. *Sci. Rep.* **8**, 1–10 (2018).
154. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
155. Blasco, A. *Bayesian Data Analysis for Animal Scientists: The Basics*. (2017). doi:10.1007/978-3-319-54274-4
156. Misztal, I. *et al.* Manual for BLUPF90 family of programs. *Univ. Georg. Athens, USA* 125 (2018).
157. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
158. D.S.Falconer & T.F.C.Mackay. *Introduction to Quantitative Genetics*. (Pearson, 1981).
159. Freeman, T. C. *et al.* Graphia: A platform for the graph-based visualisation and analysis of complex data. *bioRxiv* 2020.09.02.279349 (2020).
160. Freeman, T. C. *et al.* Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Comput. Biol.* **3**, 2032–2042 (2007).
161. Schneeberger, M., Barwick, S. A., Crow, G. H. & Hammond, K. Economic indices using breeding values predicted by BLUP. *J. Anim. Breed. Genet.* **109**, 180–187 (1992).

Acknowledgments. The authors thank Prof. Ignacy Misztal and Dr. Shogo Tsuruta for making software available to us; Prof. Agustín Blasco and Prof. Chris Haley for their statistic advice in multiple testing approach; and Prof. Michael Greenacre for his advice in compositional data analysis We also thank Bin Zhao for his contribution to the identification and biological description of metagenomics data and Dr. Larissa Zetouni for her comments on the manuscript.

Author contributions

M.M-A. R.R. and M.W. conceived and designed the overall study, and M.M-A., M.W. and R.R. conceived, designed and executed the bioinformatics analysis. M.D.A., C-A.D., R.J.D, and M.A.C. provided essential insight into microbiology, rumen metabolism, nutrition, methane emissions and animal breeding. M.M-A. and R.R. wrote the initial draft, and subsequently all authors contributed intellectually to the interpretation and presentation of the results in the manuscript, which was edited and approved by all authors.

Competing interests

The authors declare no competing interests.

Figure legends

Fig. 1| Genomic heritability (h^2) estimates of log-ratio transformed abundances of microbial taxa (a) and their genes (b) in the rumen of bovines. Bars show the h^2 values of 203/16/352 rumen microbial genera/uncultured genomes (RUGs)/genes tested exhibiting non-zero h^2 estimates ($P < 2.02 \times 10^{-5}$) a. Cultured microbial genera and RUGs classified within phylum. **b.** Microbial genes grouped by microbial biological processes: Microbial communication and host-microbiome interaction (Comm. & host interact.), Genetic information processes (Genetic Inf. processes), metabolism other than methane (Metabolism), and methane metabolism (CH_4 metabolism).

Fig. 2 | Network clusters of commonly host-genomically affected abundances of microbial genera/RUGs/genes identified in the bovine rumen. Nodes represent microbial genus/RUG/genes, and edges represent Pearson correlations among deregressed genomic effects of log-ratio transformed genera/RUGs/gene abundances > 0.70 (n= 359 animals). Clusters including ≥ 3 methanogenic archaea genera, RUGs and microbial genes involved in methane (CH_4) metabolism pathway according to KEGG database or microbial genera/RUGs/genes host-genomically correlated with CH_4 emissions (probability of the host-genomic correlation being higher or lower than 0 (P_0) ≥0.95) are highlighted and numbered from 1 to 12. Red dashed circles indicate the clusters including methanogenic archaea genera or RUGs and microbial genes involved in the CH_4 metabolism pathway and associated with microbial genera, RUGs and genes significantly (P_0≥0.95) host-genomically correlated with CH_4 emissions.1) Cluster 01 containing 272 microbial genes from which 9 are involved in CH_4 metabolism (*serA, serC, glyA, pta, ackA*, *fbaA, gpmA, acs*); and 11 show positive r_{gCH4} from 0.71 to 0.85 (P_0≥0.95) mainly representing the ribosomal small and large protein biosynthesis (*RP-L3, RP-L6, RP-L23,RP-L28, RP-L34, RP-S10, RP-S12*); 2) Cluster 02 containing *Methanobrevibacter* genera and 194 microbial genes; from which 43 encode proteins required for methanogenesis (e.g. *methyl-conezyme M (mcrA, mcrB, mcrC, mcrG)*, *coenzyme F_{420}* (*frhA*, *frhB*, *frhG*), *tetrahydomethanopterin S-methyltransferase (mtrA, mtrD, mtrE, mtrG, mtrH)*) and 2 which display strong negative r_{gCH4} of -0.71 and -0.73 with P_0≥0.95 (*cofG* in CH_4 metabolism and *queD* in folate biosynthesis); 3) Cluster 21 with 13 out of 19 microbial genes with negative r_{gCH4} from -0.71 to -0.88 (P_0≥0.96) representing arginine (*argD*) and phenylalanine metabolism (*paaH*), pyrimidine metabolism (*upp*), peptide/nickel quorum sensing transport (*ABC.PE.P*), protein export (*secD* and *secF*), nitrogen fixation (*nifU*), copper transport (*copB*), and bacterial conversion of bile acids (*choloylglycine hydrolase*); 4) Cluster 19 made from 20 microbial genes from which 11 are negatively host-genomically correlated to CH_4 emissions with r_{gCH4} from -0.82 to -0.93, P_0≥0.96 involved in e.g., ABC transport (*livH*,

livK and *livG*) and biosynthesis (*ilvA*) of branched amino acids, propionogenesis by lactaldehyde route (*fucO*) and sucrose metabolism (*sucrose phosphorylase*); 5) Cluster 14 with 38 microbial genes including 17 with negative r_{gCH4} (-0.69 to -0.91, $P_0 \geq 0.95$) associated to cobalt/nickel transport (*cbiQ*, *cbiO*) amino acid biosynthesis (*trpA*, *trpE, lysA*), porphyrin metabolism (*hemC*) and histidine metabolism (*hisA*, *hisF*); 6) Cluster 22 built by 18 microbial genes from which 9 show negative r_{gCH4} (-0.78 to -0.92, $P_0 \geq 0.96$, e.g. associated to peptide/nickel transport (*ABC.PE.S*, *ABC.PE.P1*, *ABC.PE.A*), polar amino acid transport (ABC.PA.A), or aminoacyl-tRNA biosynthesis (*gatC, gatA*); 7) Cluster 18 composed of 20 microbial genes including 6 with negative r_{gCH4} from -0.76 to -0.87 ($P_0 \geq 0.95$) involved in methionine transport (*metE, metN* and *metQ*), oxocarboxylic chain extension (*ACO*), propioniogenesis (*pccB*) and arginine biosynthesis (*argF*) and one microbial gene encoding for *enolase* in CH_4 metabolism; 8) Cluster 42 composed by 7 microbial genes including 4 with r_{gCH4} from -0.80 to -0.85 ($P_0 \geq 0.95$), e.g., *cobL* in porphyrin metabolism, *baiN* in secondary bile acids biosynthesis and neurotransmitter:Na^+ symporter (*TC.NSS*); 9) Cluster 04 with 163 microbial genera from which 117 are fungi including *Moniliophthora*, *Histoplasma* and *Metschnikowia* (r_{gCH4}=0.74-0.83, $P_0 \geq 0.95$) and 5 are methanogenic archaea (*Methanocaldococcus, Methanococcus, Methanosarcina, Methanothermococcus* and *Methanotorris*); 10) Cluster 03 composed by 175 microbial genera containing *Methanocella* and *Candidatus Methanoplasma* methanogenic genera together with Proteobacteria *Ottowia* which showed a r_{gCH4} of 0.85 (P_0=0.95); 11) Cluster 16 composed of 24 microbial genera mainly from Proteobacteria phyla but also including including *Syntrophobotulus* Firmicutes (r_{gCH4}=-0.79, P_0=0.95) and methanogens *Methanomassiliicoccus* and *Methanosaeta;*12) Cluster 11 built with 62 RUGs, 9 annotated as uncultured *Methanobrevibacter sp.* from which 5 are host-genomically correlated to CH_4 emissions, positively (r_{gCH4}=0.91, P_0=0.99) and negatively (r_{gCH4}=-0.72 to -0.86, $P_0 \geq 0.95$); and 22 annotated as uncultured *Prevotellaceae bacterium* from which 5 are positively host-genomically correlated with CH_4 emissions (r_{gCH4}=0.83 to 0.92, $P_0 \geq 0.97$).

Fig. 3 | Reaction schemes of 2-Oxocarboxylic acid metabolism and (a) glycine, serine, threonine, arginine, lysine and Coenzyme B biosynthesis or (b) branched amino acid biosynthesis, (c) phenylalanine, tyrosine and tryptophan biosynthesis and (d) starch and sucrose metabolism, in which additive log-ratio transformed microbial gene abundances strongly host-genomically correlated with methane emissions (r_{gCH4}) are involved. Small rectangles symbolize proteins encoded by the microbial genes. Microbial genes are highlighted in red when their r_{gCH4} estimates range between -0.74 and -0.93 and shows a probability of being different from 0 (P_0) ≥0.95; and in orange when they range between |0.55| and |0.77| and $P_0 \geq 0.85$. Compounds are denoted by their short names. Full names of compounds and microbial genes are given in Supplementary Data 1.

Fig. 4 | Top 20 Rumen Uncultured Genomes (RUGs) highly enriched with the 115 microbial genes host-genomically correlated to methane emissions with a probability of being higher or lower than 0 (P_0) ≥0.95. Colour scale represents the number of unique proteins mapping into each KEGG orthologous group (i.e. microbial gene). Full names of microbial genes are given in Supplementary Data 2.

Fig. 5 | Microbial genes selected to be used collectively for selecting the host genomes associated with low CH_4 emissions, meeting 3 criteria: showing significant heritability (h^2) with $P < 2.02 \times 10^{-5}$; a host genomic correlation with CH_4 (r_{gCH4}) with a probability of being higher or lower than 0 (P_0) >0.95, and showing a relative abundance above 0.01%. a. Estimates of h^2 and r_{gCH4} (error bars represent the highest posterior density interval enclosing 95% probability). Microbial genes grouped by microbial biological processes: Methane metabolism (CH_4), Microbial communication and host-microbiome interaction (Comm. & host interact.), Genetic Information processes and metabolism other than CH_4 (Metabolism)**. b.** Correlogram showing the host genomic correlations estimates among the log-ratio transformed microbial gene abundances selected for breeding purposes**.** Full names of microbial genes selected for breeding purposes are given in Supplementary Data 3.

Fig. 6 | Response to selection per generation on methane (CH_4) emissions (medians and standard deviation) estimated using direct genomic selection based on measured CH_4 emissions (light blue), indirect genomic selection based on 30 microbial gene abundances most informative for host genomic selection for methane as described in Supplementary Data 2 (dark blue) or selection on both criteria (green). Intensities of selection 1.1590, 1.400, 1.755, 2.063 or 2.665 are equivalent to selecting 30%, 20%, 10%, 5% or 1%, respectively, of our population based on the above described selection criteria.

Supplementary Figures

Supplementary Figure 1 | Phenotypic variability observed in methane emissions (CH_4) within animals from (a) the same breed (Aberdeen Angus (AAx), Charolais (Ch), Limousin (LIMx) or Luing), (b) experiment (2011, 2012, 2013, 2014 and 2017) or (c) offered the same diet (concentrate (Con) or forage (For) based). Coefficient of variation within animals from the same breeds are AAx 23.9%, Ch 28.5%, LIMx 23.2%, Luing 28.3%; within animals offered the same diet are 16.3% for forage-based and 22.6% for concentrate-based; and within animals belonging to the same experiment: 31.1% 2011, 25.5% 2012, 27.2% 2013, 13.36% 2014 and 18.8% 2017.

Supplementary Figure 2 | Presence of the 115 microbial genes host-genomically correlated to methane emissions (r_{gCH4}) with a probability of being higher or lower than 0 (P_0) $\geq$0.95 in the Rumen Uncultured Genomes (RUGs) with r_{gCH4} (P_0) $\geq$0.95 classified at genera level, and in other RUGs annotated within the r_{gCH4} (P_0) $\geq$0.95 microbial genera (*Eubacterium, Kandleria, Blautia, Anaerovibrio* and *Succinivibrio*). Colour scale represents the number of unique proteins mapping into each KEGG orthologous group (i.e. microbial gene). Full names of microbial genes are given in Supplementary Data 2.

Supplementary Tables

Supplementary Table 1a. Descriptive statistics of microbial genera (mean relative abundances (RA) and their coefficients of variation (CV)) in ruminal microbiota.

Supplementary Table 1b. Descriptive statistics of Rumen Uncultured Genomes (RUGs) (mean relative abundances (RA) and their coefficients of variation (CV)) in ruminal microbiota.

Supplementary Table 1c. Descriptive statistics of microbial genes (mean relative abundances (RA) and their coefficients of variation (CV)) and their involvement in microbial biological processes in ruminal microbiome.

Supplementary Table 2a. Microbial genera abundances with significant ($P<2.02x10^{-5}$) heritability (h^2) in rumen microbiota.

Supplementary Table 2b. Rumen Uncultured Genomes (RUGs) with significant ($P<2.02x10^{-5}$) heritability (h^2) in rumen microbiota.

Supplementary Table 2c. Microbial gene abundances with significant ($P<2.02x10^{-5}$) heritability (h^2) in rumen microbiome.

Supplementary Table 3a. Microbial genera abundances with a probability (P_0) ≥0.95 of being host-genomically correlated with methane emissions (g/kg DMI).

Supplementary Table 3b. Rumen Uncultured Genome (RUG) abundances with a probability (P_0) ≥0.95 of being host-genomically correlated with methane emissions (g/kg DMI).

Supplementary Table 3c. Microbial gene abundances with a probability (P_0) ≥0.95 of being host-genomically correlated with methane emissions (g/kg DMI).

Supplementary Table 4. Composition of clusters from a co-abundance network analysis among deregressed host genomic effects (dGEBVs) of microbial genus/RUG/gene abundances in rumen microbiome.

Supplementary Table 5. Proteins clustered in KEGG orthologous groups (KO) host-genomically correlated with methane emissions identified in rumen uncultured genomes (RUG).

Supplementary Table 6. Enrichment analysis of the microbial genes host-genomically correlated to methane emissions in each Rumen Uncultured Genome (RUG).

Supplementary Table 7. Correlated responses to selection in methane (CH_4) emissions after selection for each microbial genus/RUG/gene abundance with a probability (P_0) ≥0.95 of being host-genomically correlated with CH_4.

Supplementary Table 8. Microbial genus/RUG/gene abundances recommended for microbiome-driven breeding to mitigate methane (CH_4) emissions showing a relative abundance (RA) > 0.01%, being significantly heritable ($P < 2.02\ x10^{-5}$) and with a probability (P_0) ≥0.95 of being host-genomically correlated with CH_4 emissions.

Supplementary Table 9. Experimental design displaying the number of animals within each breed, diet and experiment.

Supplementary Data

Supplementary Data 1. Full names of compounds and microbial genes are given in Figure 3.

Supplementary Data 2. Full names of compounds and microbial genes are given in Figure 4.

Supplementary Data 3. Full names of compounds and microbial genes are given in Figure 5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Combinedsupplementarymaterial.pdf
- Combinedsupplementarymaterial.pdf