Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/7/eaav8391/DC1

Supplementary Materials for

A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions

R. John Wallace*, Goor Sasson, Philip C. Garnsworthy, Ilma Tapio, Emma Gregson, Paolo Bani, Pekka Huhtanen, Ali R. Bayat, Francesco Strozzi, Filippo Biscarini, Timothy J. Snelling, Neil Saunders, Sarah L. Potterton, James Craigon, Andrea Minuti, Erminio Trevisi, Maria L. Callegari, Fiorenzo Piccioli Cappelli, Edward H. Cabezas-Garcia, Johanna Vilkki, Cesar Pinares-Patino, Kateřina O. Fliegerová, Jakub Mrázek, Hana Sechovcová, Jan Kopečný, Aurélie Bonin, Frédéric Boyer, Pierre Taberlet, Fotini Kokou, Eran Halperin, John L. Williams, Kevin J. Shingfield, Itzhak Mizrahi*

*Corresponding author. Email: john.wallace@abdn.ac.uk (R.J.W.); imizrahi@bgu.ac.il (I.M.)

Published 3 July 2019, *Sci. Adv.* **5**, eaav8391 (2019)
DOI: 10.1126/sciadv.aav8391

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/7/eaav8391/DC1)

Data S1 (Microsoft Excel format). Animals used in the experiment together with diet, measured phenotypes, and other experimental variables.
Data S2 (Microsoft Word format). Presence of bacterial taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the current study.
Data S3 (Microsoft Word format). Presence of archaeal taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the present study.
Data S4 (Microsoft Word format). Presence of protozoal taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the present study.
Data S5 (Microsoft Excel format). Summary of abundance and occupancy of the core microbial species (prokaryotes, fungi, and protozoa).
Data S6 (Microsoft Excel format). Heritable microbes.
Data S7 (Microsoft Excel format). Microbes associated with phenotypic traits.
Data S8 (Microsoft Excel format). Closest representative sequenced genomes for heritable microbes.
Data S9 (Microsoft Excel format). Season-affected microbes.
Data S10 (.zip format). Animal genotypes (SNP values).

Supplementary Materials and Methods

Supplementary Text. Interactions between main categories of ruminal microbes in abundance, diversity, and animal phenotype

Total microbial DNA in all digesta samples was analysed by qPCR using universal primers for bacteria, protozoa and archaea. The analysis confirmed many previous observations (*4*), in that bacteria and protozoa were most abundant and negatively correlated (fig. S4A) and that protozoal abundance impacts upon ruminal VFA concentrations, negatively for propionate and positively for butyrate concentration (fig. S4C). When qPCR data were combined with diversity analysis, protozoal abundance was positively correlated with bacterial diversity (fig. S4B); furthermore the richness of bacterial species was inversely related to ruminal propionate concentration (fig. S4D). As protozoa and bacteria form a predator-prey relationship, these observations together imply that predominantly propionate-producing bacteria which are mostly Gram-negative and therefore more susceptible to protozoal predation (*67*), are suppressed by protozoa, leaving a variety of other species to fill the niche. The practical usefulness of suppressing protozoa depends on many factors, particularly dietary composition (*68*). Illustrating possible beneficial impacts, correlations occurred between protozoal abundance and milk fat (positive; fig. S4C) and milk fat (negative; fig. S4C), indicating that suppressing protozoa in animals receiving these diets would lower milk fat and increase protein, both contributing to milk with a better health profile for human consumers.

Archaeal abundance was positively associated with bacterial abundance and negatively associated with protozoal abundance (fig. S4A), which might be expected given that archaea would be consumed by protozoa in a similar way to bacteria and fungi (*69*). Otherwise, archaeal abundance and species richness were only weakly correlated to rumen fermentation products and milk production. Even methane emissions did not correlate, despite rumen archaea being almost entirely methanogenic (fig. S4C, D). Although such a relationship has sometimes been found, it seems that factors other than archaeal abundance dictate quantities of methane produced (*70*). Variations in the bacterial community, and thus in amounts of H_2 (the main substrate, with CO_2, for methanogenesis) produced by different bacterial species, probably explain most of the variation in methane emissions (*70*).

Data and software availability

Datasets will be made freely available to readers from the date of publication. 16S rRNA and other microbial marker gene sequences will be available under Short Reads Archive (SRA). Host genotypes will be available under European Variation Archive (EVA).

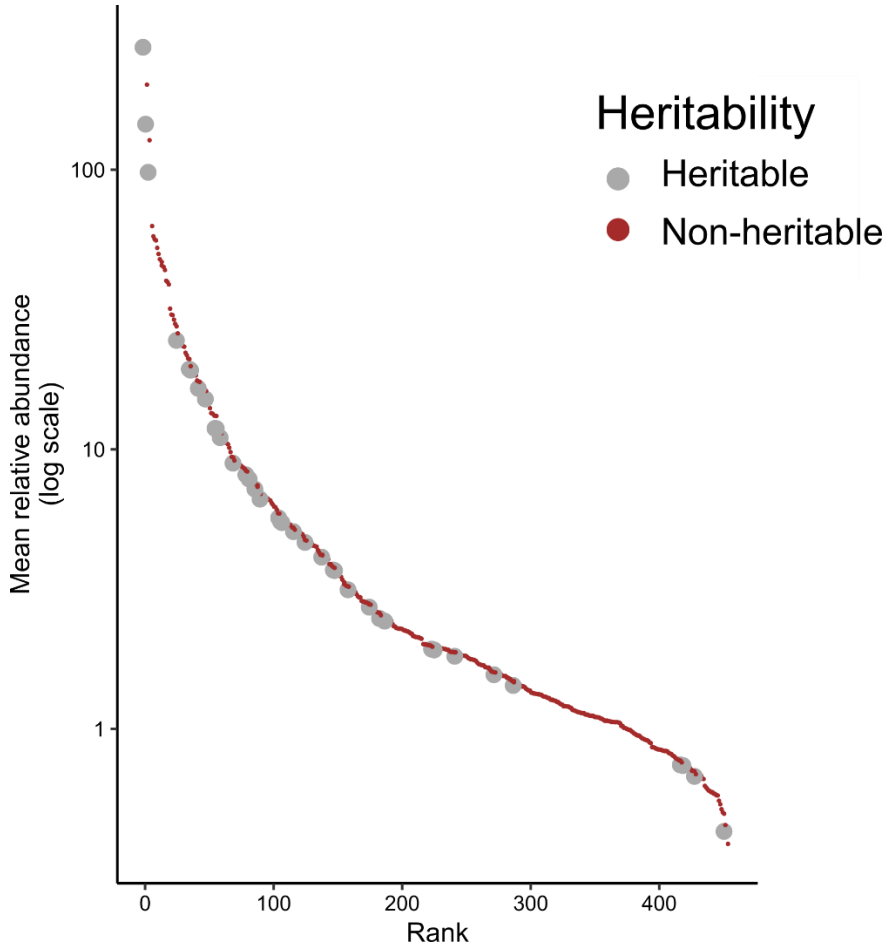

Fig. S1. Rank abundance plot for core microbes along farms. 454 core microbes abundance and abundance-rank (mean along all animals). X-axis: microbe abundance-rank. Y-axis: relative abundance.

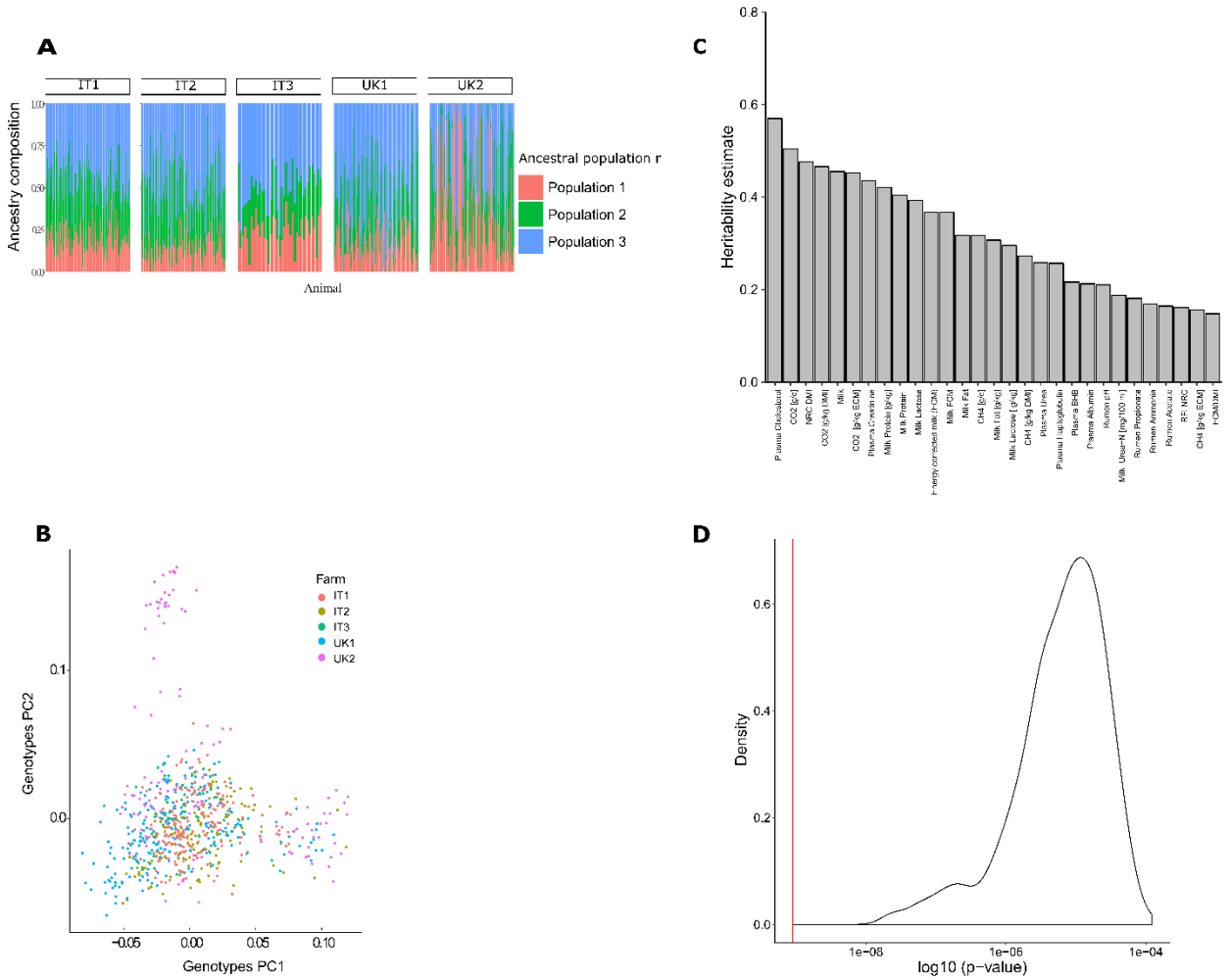

Fig. S2. Overview of genetic components along the Holstein cohort and host SNP-microbe association effort. (**A**) Stacked bar-plot showing the genetic composition within each animal as a mixture of three ancestral populations. X-axis: animals, grouped into farms. Y-axis: the relative portion of ancestral population within the genetics of the animal, where each color represents one ancestral population. Analysis performed using ADMIXTURE (**B**) A PCA scatterplot showing the distribution of animals according to the genotypes two main principal components (PCs). (**C**) Heritability estimates for host traits that showed significant heritability estimates. Estimates were calculated using GCTA (Genetic Complex Trait Analysis) software, followed by a multiple testing correction using the Benjamini-Hochberg procedure. (**D**) Distribution of top SNP association p-values (lowest SNP association p-value for each core microbe). X-axis: p-value. Y-axis: density. Red vertical line designates Bonferroni corrected significance threshold. Association was done using the Holstein subset of the animals excluding UK2 using the program GCTA (genetic complex trait association analysis) in where population structure and farm effects were accounted for by considering the SNP under examination, the farm and top genotype PCs as fixed effects and the GRM as the random effect in a linear mixed model.

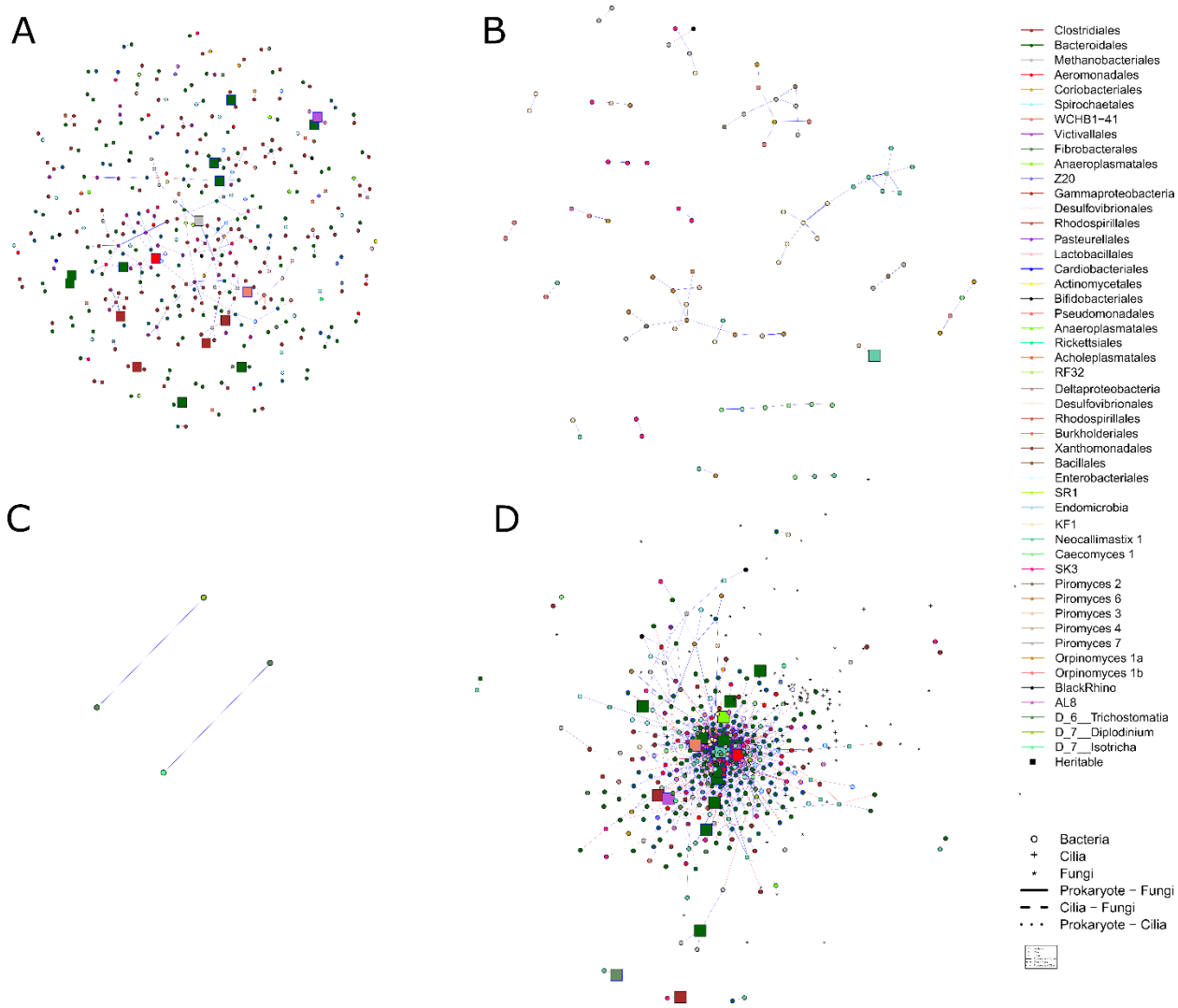

Fig. S3. Microbial species interaction within and between domains. Microbe-microbe interactions within prokaryotes (**A**), fungi (**B**), protozoa (**C**) and inter-domain (**D**). Nodes designate microbes and edges between them represent either positive (blue) or negative (red) interactions. Line thickness is proportional to the number of farms, where the observation has been identified. Within-domain interactions were inferred using the SpiecEASI (SParse InversE Covariance Estimation for Ecological Association Inference) framework. Inter-domain interactions were inferred using a Spearman correlation over the relative abundance profiles, considering correlations with Benjamini-Hochberg FDR $p < 0.05$ and $r > 0.5$ as interactions.

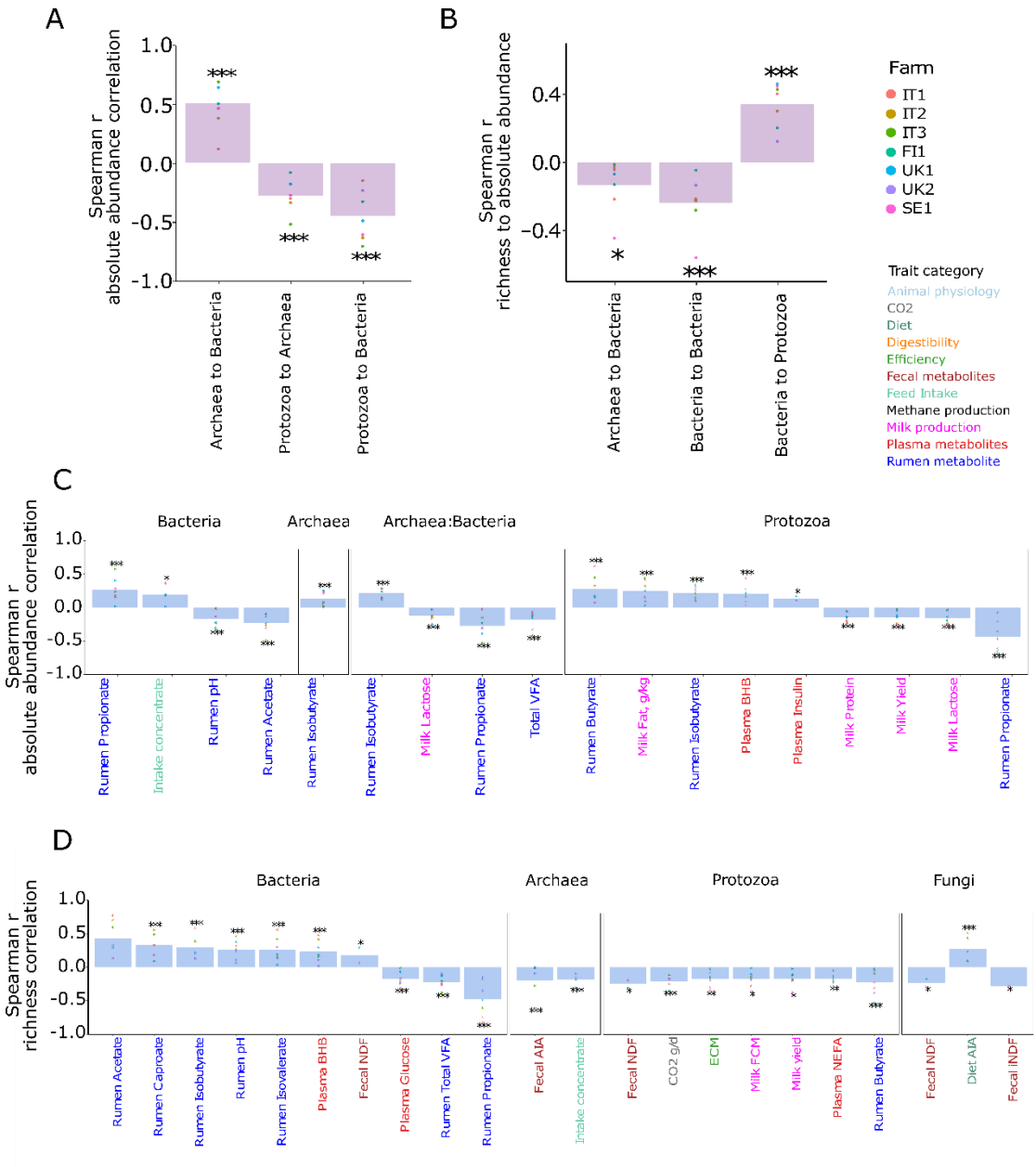

Fig. S4. Species richness and abundance of rumen microbial domains reveal ecological interactions and connection to host traits. Inter-domain interactions (X-axis) were identified within the rumen microbiome with relation to absolute abundance in cell counts (**A**) and richness (**B**). Significant correlations between (**C**) absolute abundance in cell counts and (**D**) richness, to different host trait categories (X-axis) were also revealed. Correlation analysis was performed using Spearman r (Y-axis). The absolute abundance in cell counts was calculated by quantifying the number of domain marker gene copies per ng of extracted DNA using quantitative PCR. Meta-analysis was performed after selection for domain-pairs or experimental variables that were consistently correlating either positively or negatively. Significance was corrected using Benjamini-Hochberg FDR procedure. Indicated *P*-values, $P < 0.05$ with *, $P < 0.005$ with **, $P < 0.0005$ with ***.

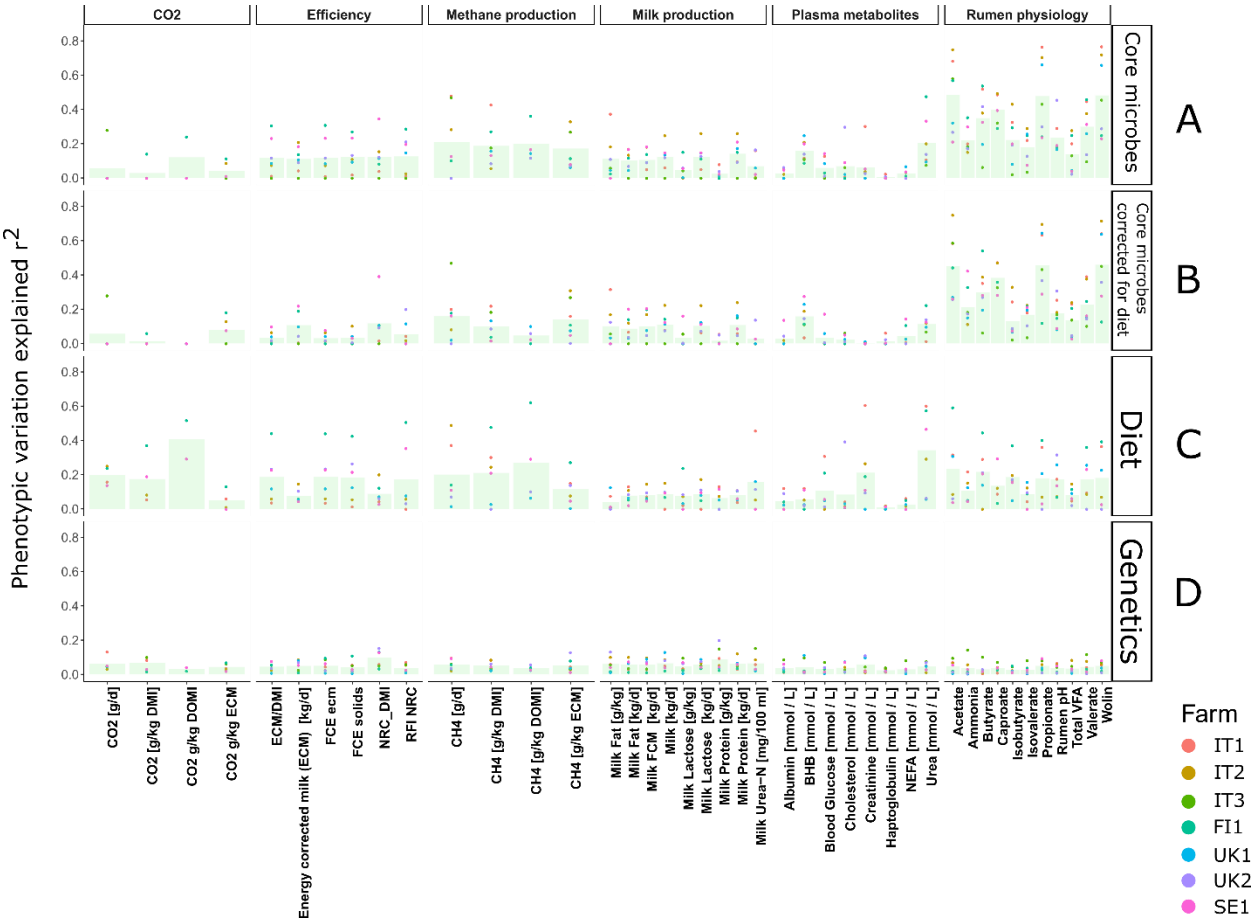

Fig. S5. Host genetics, core microbiome composition, and diet shape the host phenotypic landscape. Regression and genomic prediction analysis was performed to estimate the explained phenotypic variability by A. core microbes, B. core microbes corrected for the dietary components, C. diet, and D. genetics. Each phenotype's explainability (r^2; Y-axis) is a mean of estimates resulting from 10-fold cross-validations for the core composition and diet, and 3-fold cross-validation for the genomic prediction, repeated 100 and 10 times, respectively. Core microbe abundance and diet components were used as predictors in the Ridge regression, where phenotype was the dependent variable. Microbes and phenotypes were corrected for diet in (B) by regressing them out, using a Ridge regression, over the dietary components. For the genomic prediction, Genome Association and Prediction Integrated Tool (GAPIT) was used to estimate r^2, based on genetic kinship matrix between the animals.

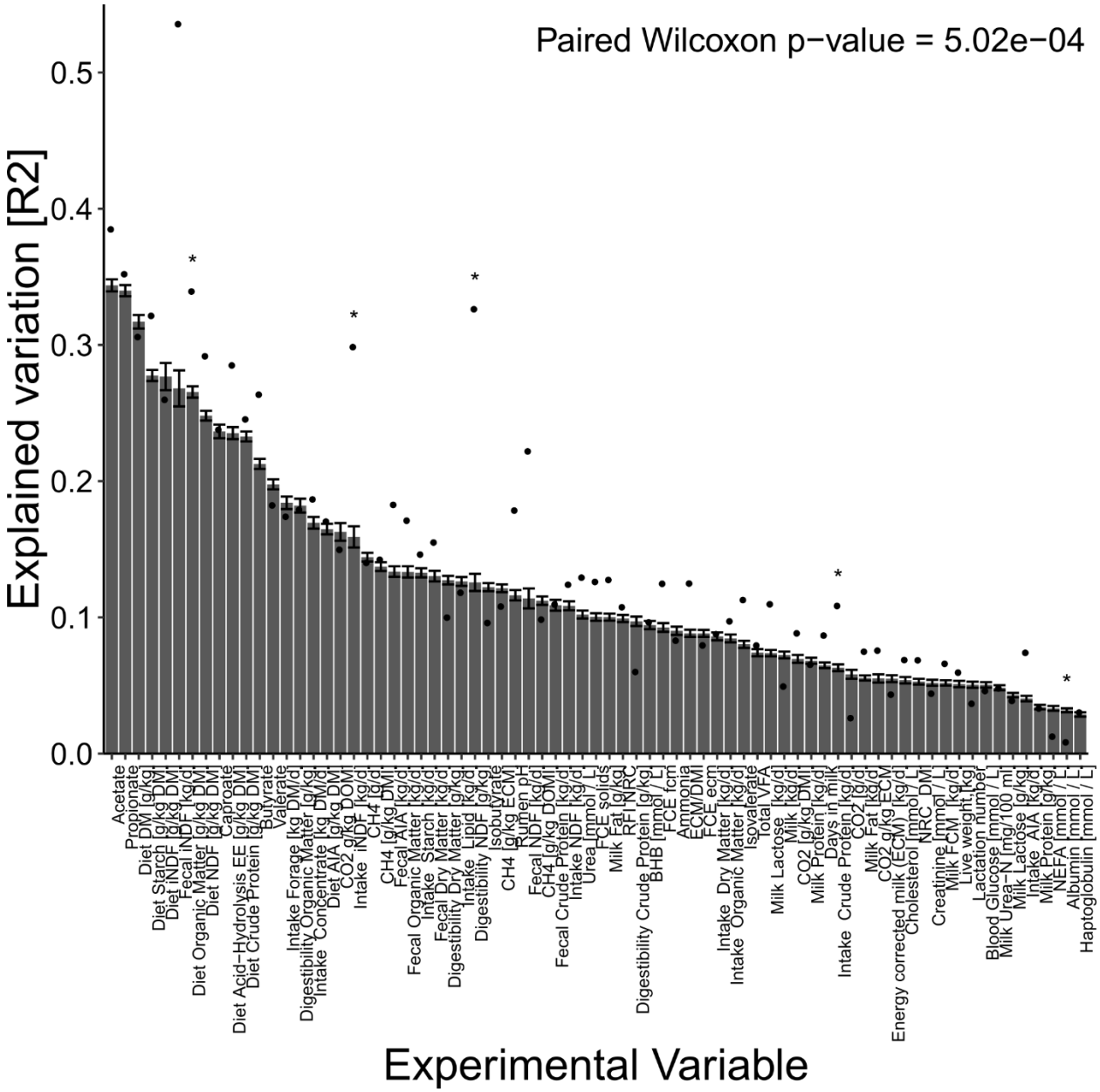

Fig. S6. Heritable microbes tend to explain experimental variables better in comparison to nonheritable core microbes. X-axis: experimental variable. Y-axis: Ridge regression R^2 value for explaining the phenotype. Point: R^2 when heritable microbes used as independent variables. Bar-lot and whiskers relate to mean and standard error of R^2 values obtained from 1,00 random samples of non-heritable core microbes that were used as independent variables. Wilcoxon paired rank-sums test was used to compare heritable microbes' R^2 values for explaining the different experimental variables to that of non-heritable core microbes (mean R^2).

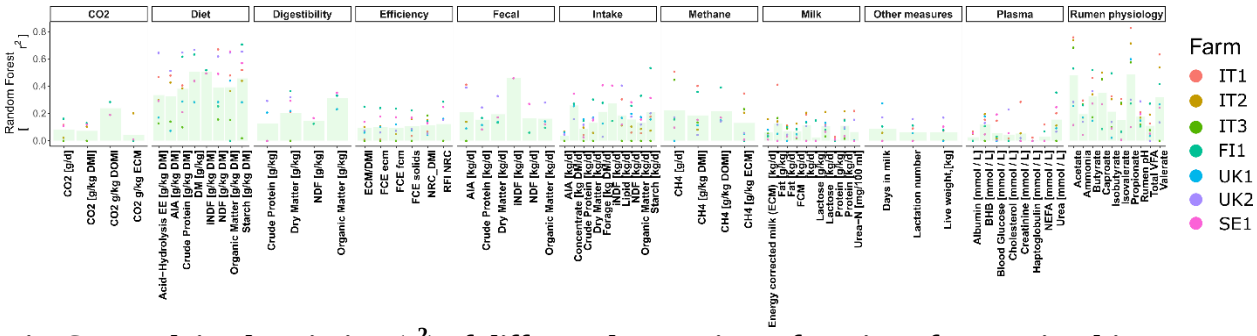

Fig. S7. Explained variation (r^2) of different host traits as function of core microbiome composition, according to RF prediction model. r^2 estimates were derived from a machine-learning approach where a trait-value was predicted for a given animal using a Random-Forest model that was constructed from all other animals in farm (leave-one-out regression). Thereafter, prediction r^2 value was calculated between the vectors of observed and predicted trait values. Indicated host traits were significantly explained (via prediction) by core microbe (OTU) abundance profiles. Dots stand for individual farms' prediction r^2 while bar heights represent mean of individual farms' r^2.

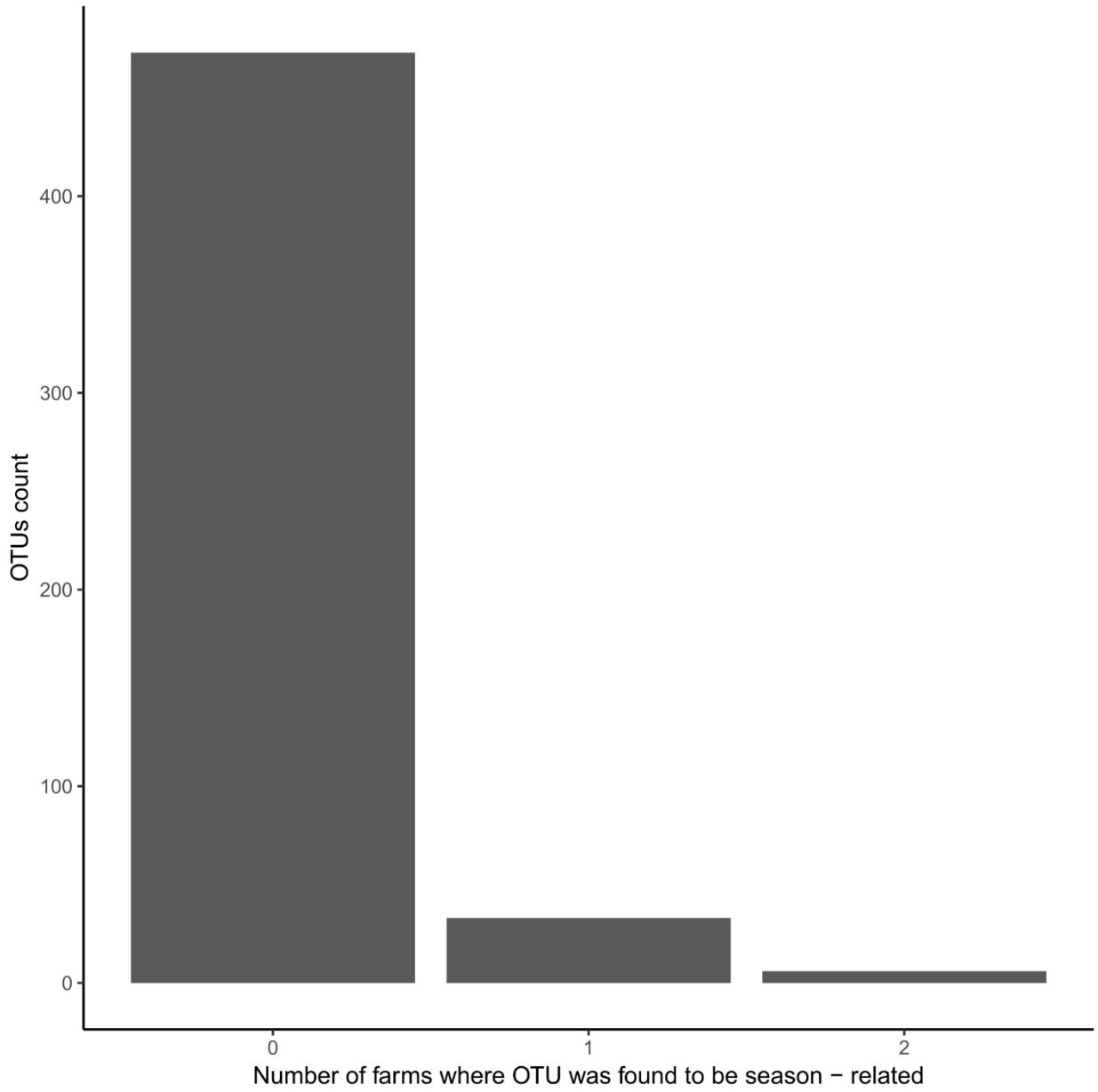

Fig. S8. The vast majority of core microbes do not show a seasonal association, and evidence for seasonality usually does not repeat in more than one farm. Analysis began by correcting core microbes in a farm for diet. Thereafter, the samples in the farm were partitioned into two groups, winter (fall equinox to spring equinox) and summer (spring equinox to fall equinox). Then, microbial OTU abundances were compared using Wilcoxon rank-sums test that was used to test for difference between the abundance of the given OTU between the two seasons, followed by a multiple comparison correction using the Bonferroni method. Core microbial OTU with corrected $P < 0.05$ in at least one farm were considered as showing a seasonal association. X-axis: The number of farms in which a microbe was found to be seasonality-associated. Y-axis: count of core microbes.

Table S1. Average diet formulations and composition on each farm.

	Farm						
kg/t (DM basis)	FI1	IT1	IT2	IT3	SE1	UK1	UK2
Grasssilage	550				553	236	442
Maize silage		482	428	466		250	267
Grass hay		112	99	100			
Alfalfa hay			45	83			
Wheat silage						230	
Clover Silage					6		
Fodder beet or potatoes							76
Barley/wheat straw					5		
Barley grain (rolled)	210				273		
Barley meal	55						
Maize meal		57	117	139			
Molassed sugar beet pulp	55						
Soya meal		94	106			91	40
Rapeseed meal	115				101	91	100
Cotton seed		90	62				
Molasses						40	

Protected fat						33	15
Mineral and vitamin premix	15					18	8
Commercial concentrate[1]		165	143	212	62	11	52
	1000	1000	1000	1000	1000	1000	1000
Feeding Method[2]	TMR	TMR	TMR	TMR	TMR	PMR	PMR
Composition (g/kg DM unless stated)							
Dry matter (g/kg)	539	557	580	562	417	533	441
Organic matter	927	928	932	934	931	938	927
Crude protein	165	158	157	149	168	168	159
Neutral-detergent fibre	389	307	324	310	400	338	383
Ether extract, acid hydrolysis		38	39	39	34	65	52
Starch	155	266	269	263	162	149	135
Metabolisable Energy (MJ/kg DM)	11.0	12.3	12.1	11.2	11.2	12.0	11.5

[1] Commercial concentrates: typical ingredients Brewers'Grains, Distillers Syrup, Wheatfeed, Potato Mash, Moist Citrus Pulp, Biscuit Meal, Malt Residual Meal, Potato Peel, Rapeseed Meal, Pressed pulp; except Farm UK1 Distillers grains (33.3%), sugar beet pulp (33.3%), soya hulls (33.3%).

[2] TMR = Total Mixed Ration fed ad libitum; PMR = Partial Mixed Ration fed ad libitum, plus concentrates fed during milking at a rate of 3.5kg per cow per day, and an additional 0.45 kg/kg milk yield above 35kg/d.

Table S2. Conditions for quantitative PCR.

Target group	Primer pair	Forward and reverse primers (5'-3')	Region of amplification	Tm	Number of PCR cycles	$MgCl_2$ concentration
Bacteria	Bact	BactF - GGATTAGATACCCTGGTAGT BactR - CACGACACGAGCTGACG	16S rRNA V5-6	57°C	30	2.0 mM
Archaea	Arch	ArchF - CCTGCTCCTTGCACACAC ArchR - CCTACGGCTACCTTGTTAC	16S rRNA V9	58°C	40	2.5 mM
Fungi	Neoc	NeocF - TACCCTTTGTGAATTTGTT NeocR - ATCCATTGTCAAAAGTTGT	ITS1	49°C	35	2.5 mM
Ciliate protozoa	Cili	CiliF - CGATGGTAGTGTATTGGAC CiliR - GGAGCTGGAATTACCGC	18S rRNA	55°C	35	2.5 mM
Bacteria and archaea	Prok	ProkF - GCCAGCMGCCGCGGTAA ProkR - GGACTACCMGGGTATCTAATC	16S rRNA	59°C	30	2.0 mM

Data S1. Animals used in the experiment together with diet, measured phenotypes, and other experimental variables.

Data S2. Presence of bacterial taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the current study.

Data S3. Presence of archaeal taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the present study.

Data S4. Presence of protozoal taxonomic groups that were found to be most abundant in Henderson *et al.* (*6*) and appear also in the present study.

Data S5. Summary of abundance and occupancy of the core microbial species (prokaryotes, fungi, and protozoa).

Data S6. Heritable microbes.

Data S7. Microbes associated with phenotypic traits.

Data S8. Closest representative sequenced genomes for heritable microbes.

Data S9. Season-affected microbes.

Data S10. Animal genotypes (SNP values).

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, John Wallace (john.wallace@abdn.ac.uk).