

Stochasticity constrained by deterministic effects of diet and age drive rumen microbiome assembly dynamics

**Ori Furman^{1#}, Liat Shenhav^{2#}, Goor Sasson¹, Fotini Kokou¹, Hen Honig³, Shamay Jacoby³,
Tomer Hertz⁴, Otto.X Cordero⁵, Eran Halperin², Itzhak Mizrahi^{1*}**

¹Department of Life Sciences, Ben-Gurion University of the Negev and the National Institute for Biotechnology in the Negev, Marcus Family Campus, Beer-Sheva, Israel

²Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA

³Institute of Animal Sciences, Agricultural Research Organization, Rishon Letziyon, Israel

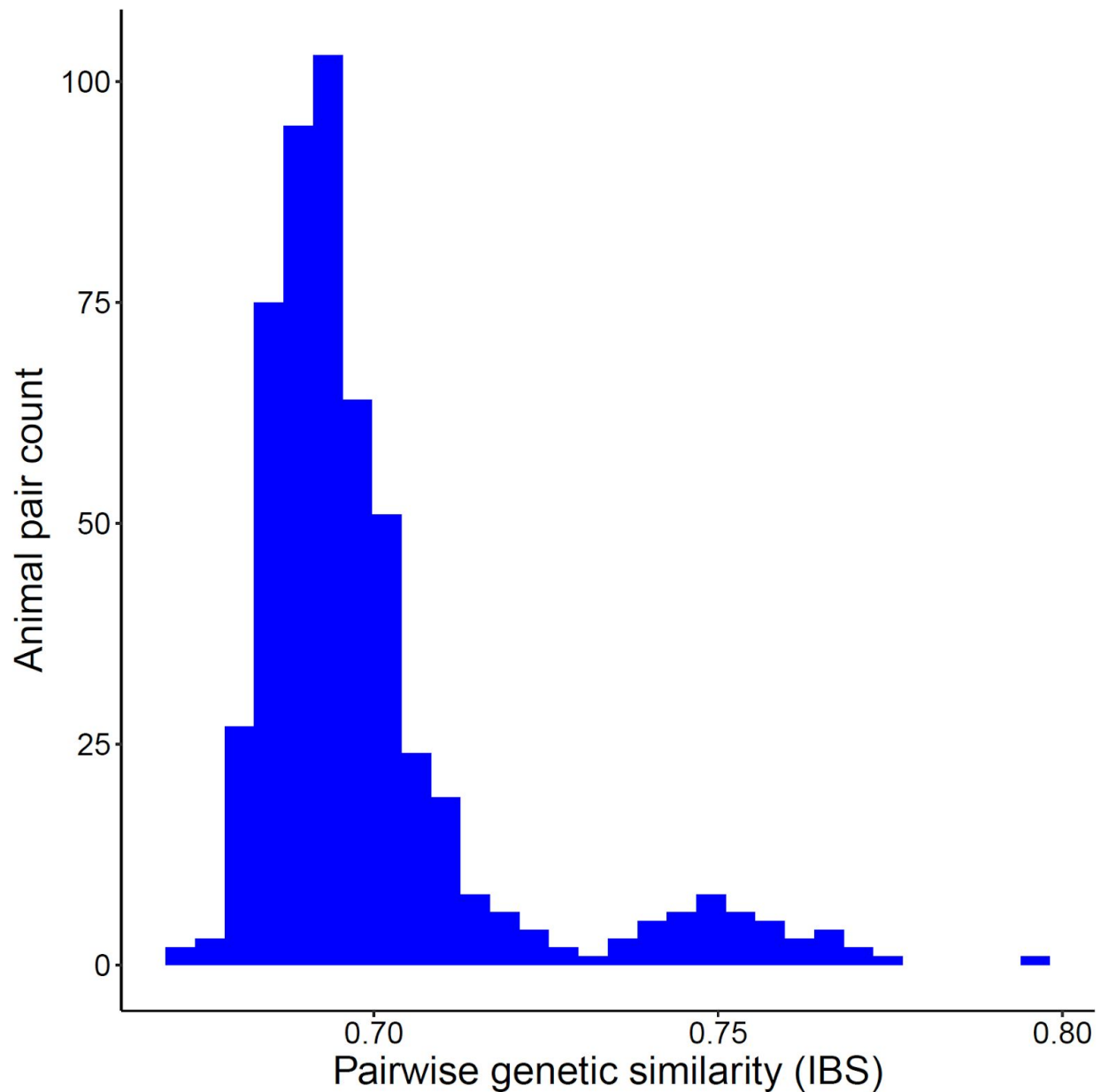
⁴The Shraga Segal Department of Microbiology, Immunology and Genetics, Ben-Gurion University of the Negev and the National Institute for Biotechnology in the Negev, Marcus Family Campus, Beer-Sheva, Israel

⁵Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

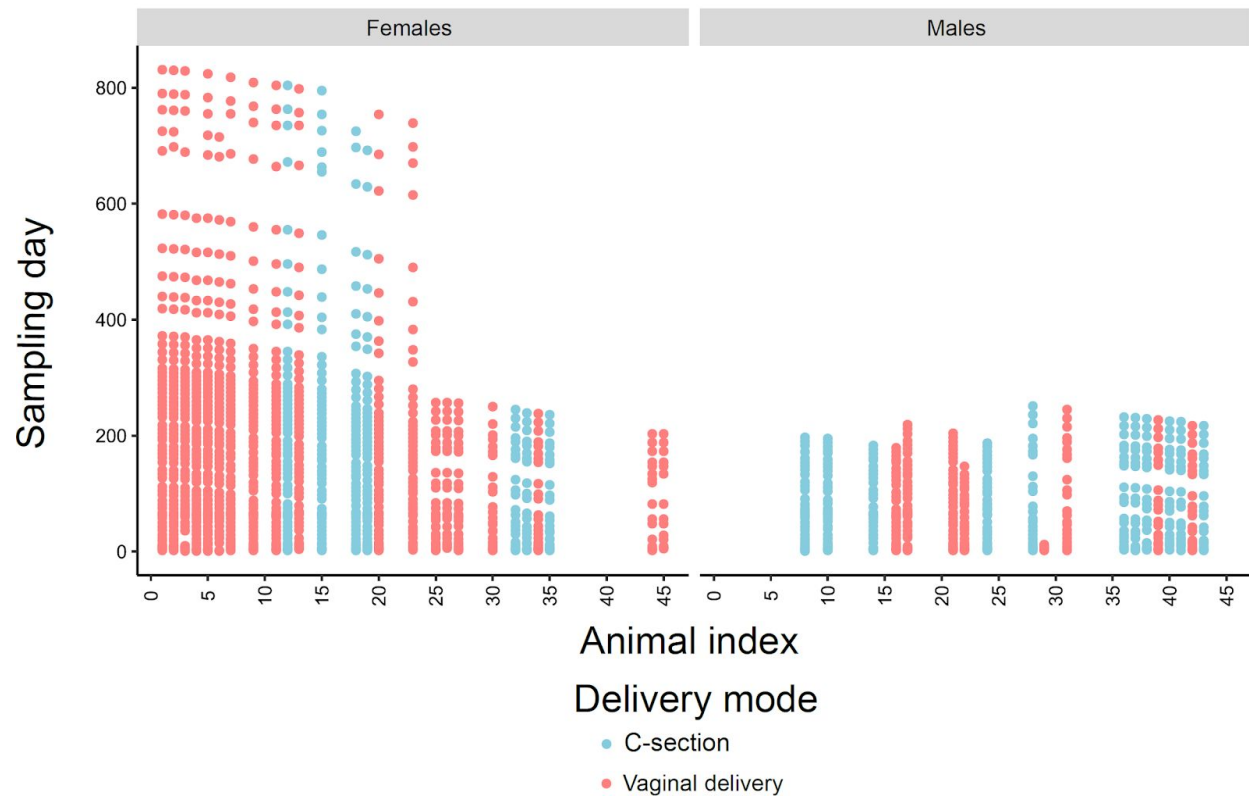
[#]These authors contributed equally

^{*}Corresponding author: Prof. Itzhak Mizrahi, Department of Life Sciences, Ben-Gurion University of the Negev, Marcus Family Campus, Beer-Sheva, Israel. Email: imizrahi@bgu.ac.il; Tel. (+972) 8 647 9836.

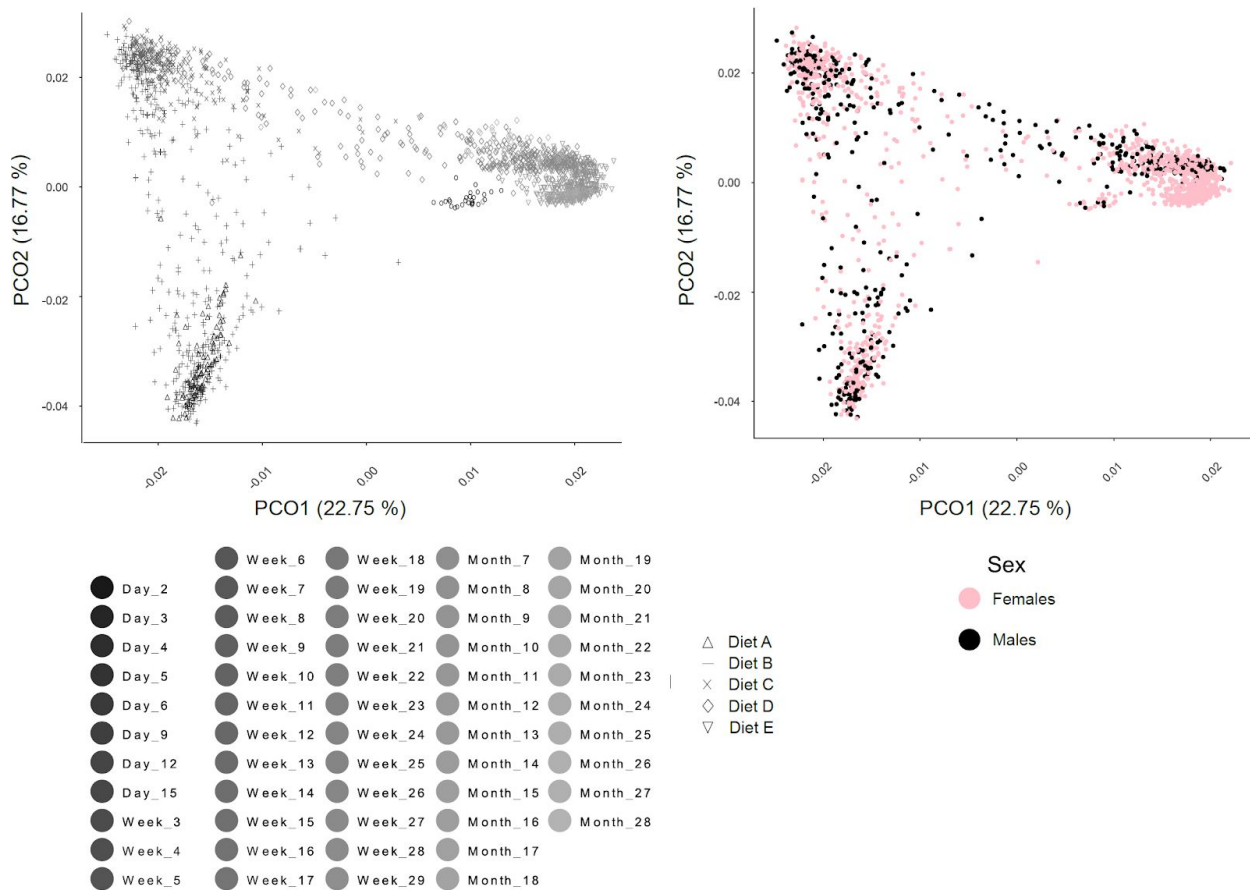
Supplementary Figures



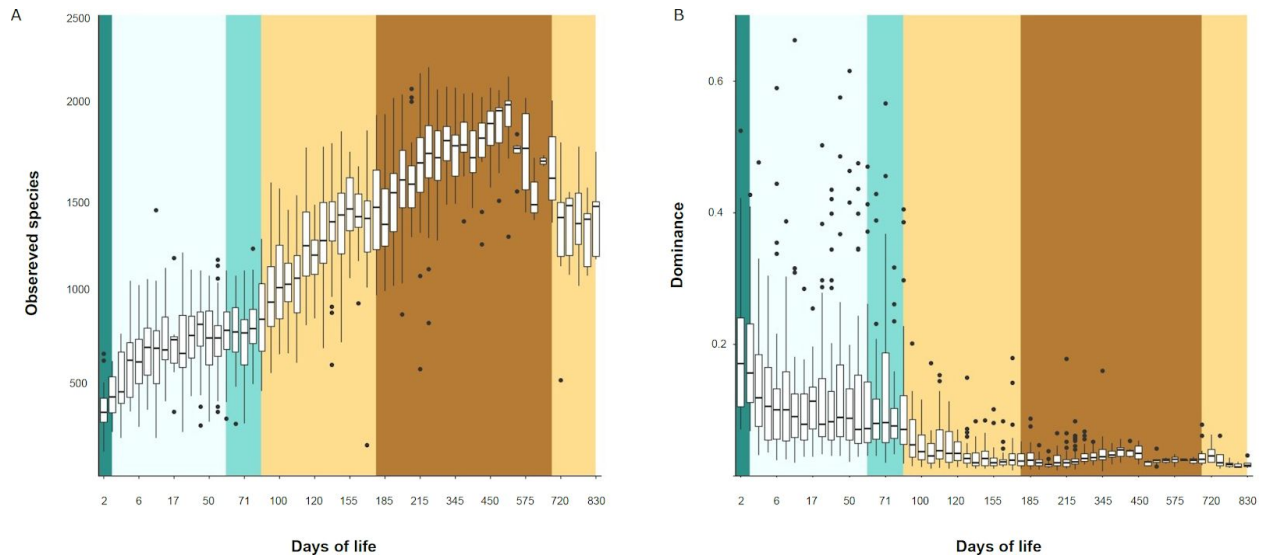
Supplementary Figure 1 - **Pairwise host genetic similarity based on 33 out of 45 animals.** No evidence was found of host genetic similarity within delivery-mode groups (PERMANOVA, $P = 0.316$). In addition, host genetics did not contribute to the similarity in rumen microbial composition and relative abundance (Mantel test, $P = 0.788$). Source data is provided as a Source Data file.



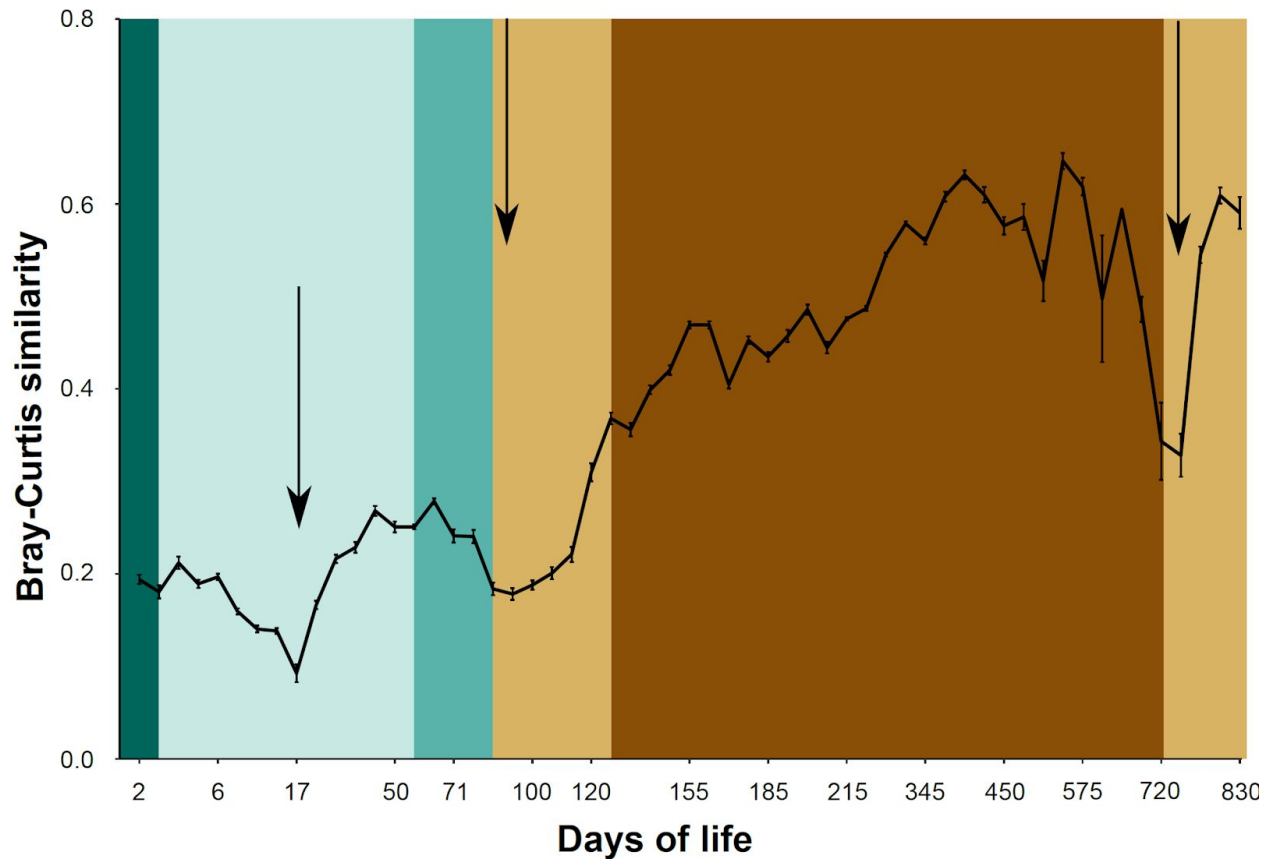
Supplementary Figure 2 - Sampling regime for all animals. Animals were divided between females (left panel) and males (right panel). X-axis represents animal index. Y-axis represents the number of samples (n = 1634 samples across all animals) and sampling days. Blue dots are C-section-born animals (n = 18 animals) and red dots are vaginally delivered animals (n = 27 animals). For further elaboration, see Supplementary Data 6. Source data is provided as a Source Data file.



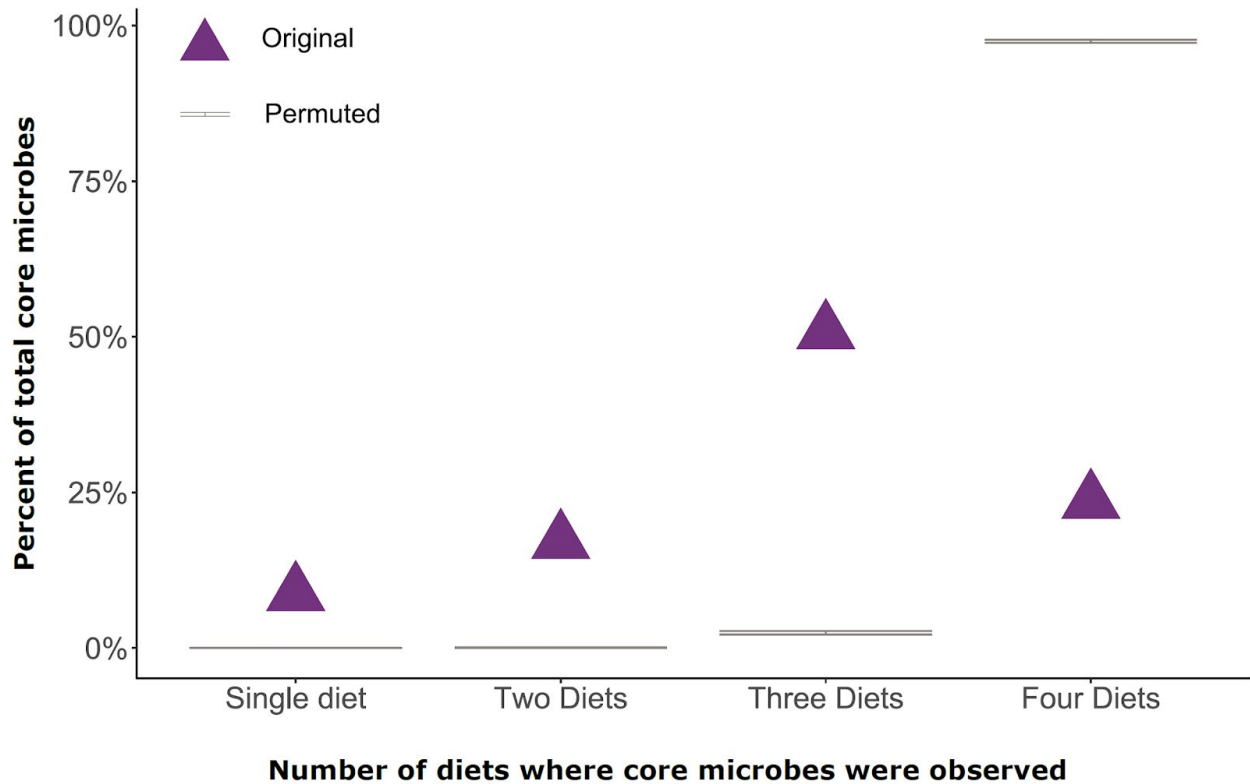
Supplementary Figure 3 - Dynamics of the different microbial families is shaped by age and diet. **A.** Principal coordinate analysis (PCoA), based on Bray–Curtis metrics, showed clustering of operational taxonomic units (OTUs) according to age and diet (PERMANOVA, $P = 0.001$). Animals were fed five different diets during the sampling period. Each dot represents a single sample ($n = 1634$). Samples are colored in grayscale, where darker colors represent older ages. Shapes of the different samples represent the different diets. **B.** Principal coordinate analysis (PCoA), based on Bray–Curtis metrics, showed clustering of operational taxonomic units (OTUs), according to sex. Each dot represents a single sample ($n = 1634$). Samples are colored according to animal sex, which was not found to be a significant factor for clustering of the samples. Females are colored in red and males in black. Source data is provided as a Source Data file.



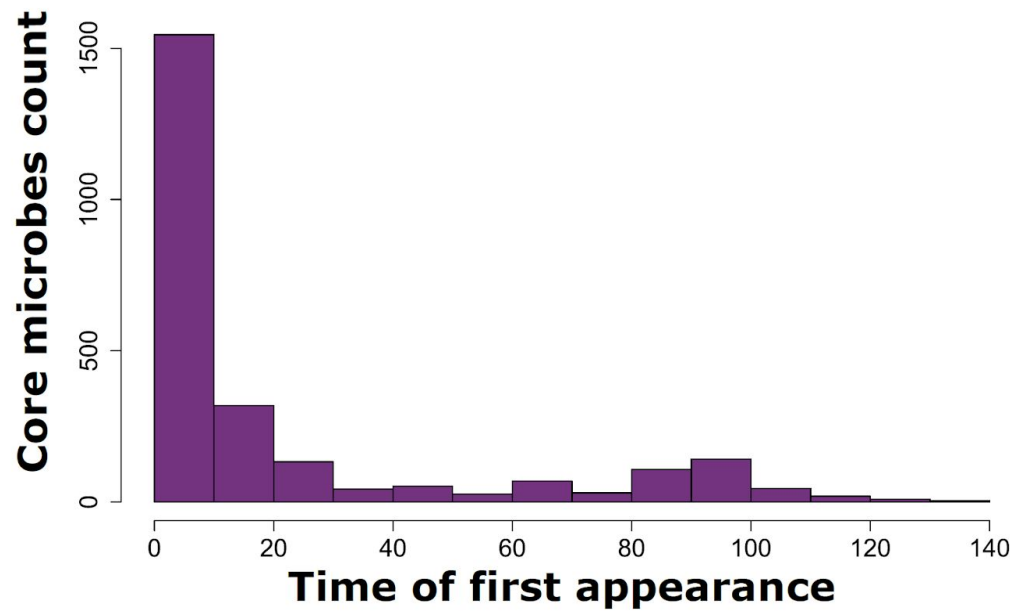
Supplementary Figure 4 - Sample diversity increases over diet and age. A. Number of observed species increased over age and diet. **B.** Dominance of different samples across time, Y-axis represents the number of observed species. X-axis represents age. Samples ($n = 1634$) were binned according to age ($n = 58$ time bins); time bins are defined in Supplementary Data6. Boxes represent the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the horizontal line inside the box defines the median. Whiskers represent the lowest and highest values within 1.5 times the IQR from the first and third quartiles, respectively. Source data is provided as a Source Data file.



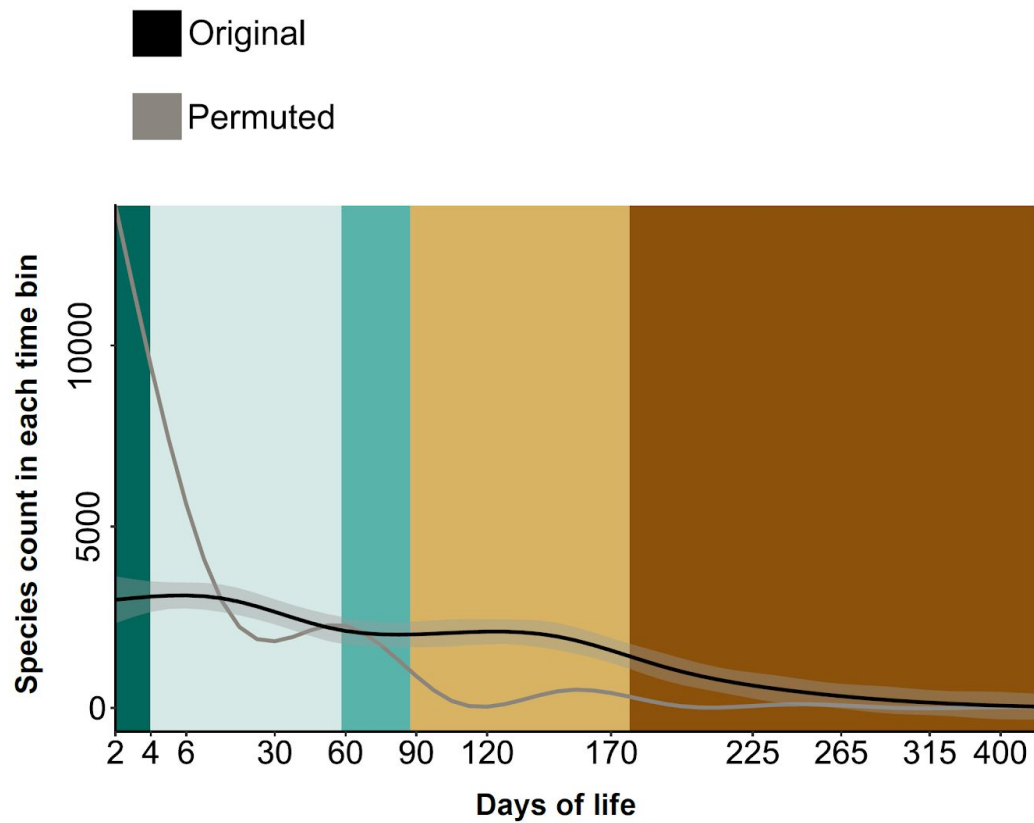
Supplementary Figure 5 - **Similarity (Bray–Curtis) within each time bin increases across time with periodical decreases during dietary changes.** Y-axis represents similarity in microbiome composition using the Bray–Curtis measure of microbial species relative abundance. Error bars represent standard error of mean, samples (n = 1634) were binned according to age (n = 58 time bins); time bins are defined in Supplementary Data 6. Source data is provided as a Source Data file.



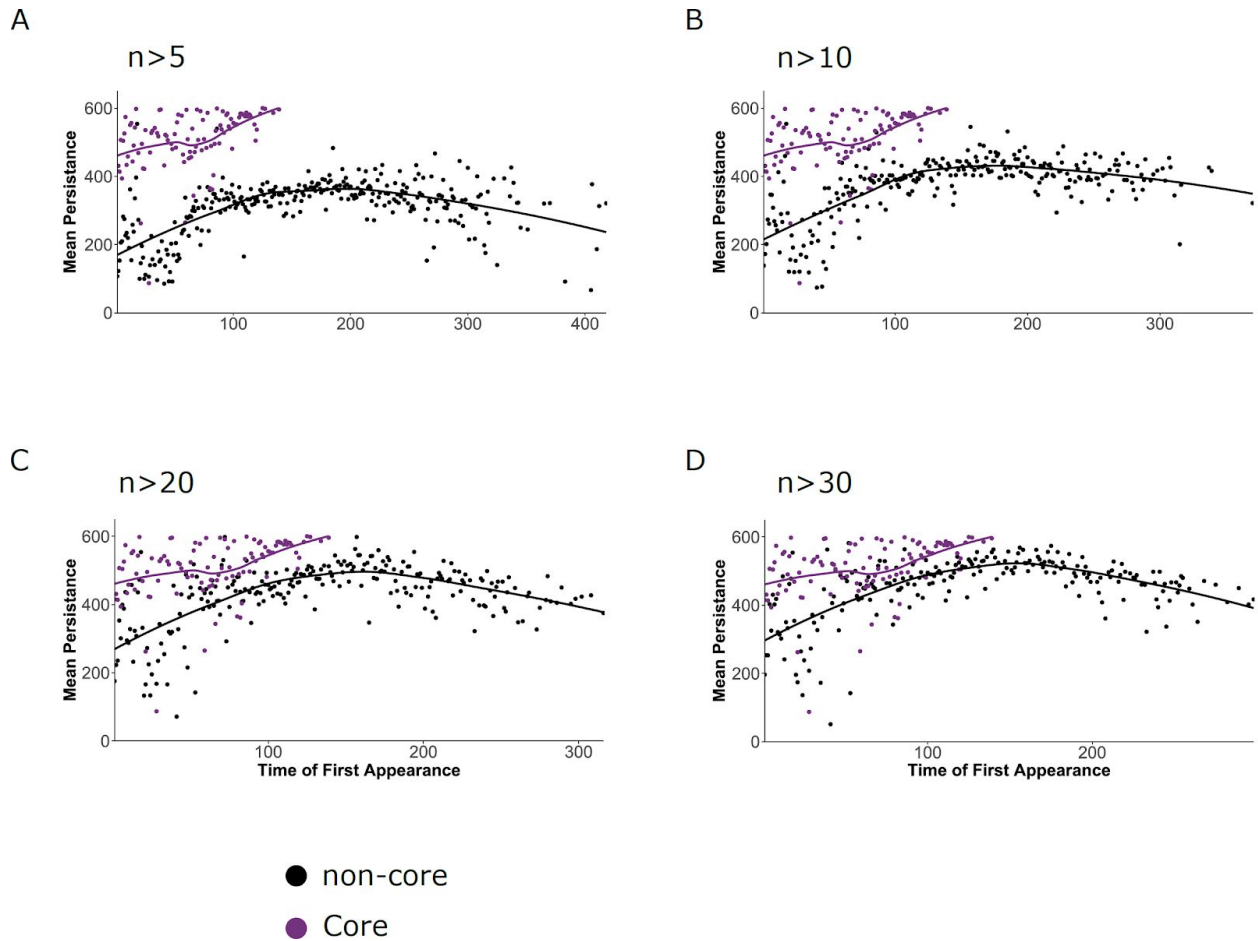
Supplementary Figure 6 - **Core successional microbe appearance under different dietary regimes.** To examine whether bacterial species tend to be diet-specific, we performed a permutation test ($n = 100$), in which each row was shuffled at each iteration. Thus, the labels comprising each row were changed for each iteration, randomizing time and diet. Source data is provided as a Source Data file.



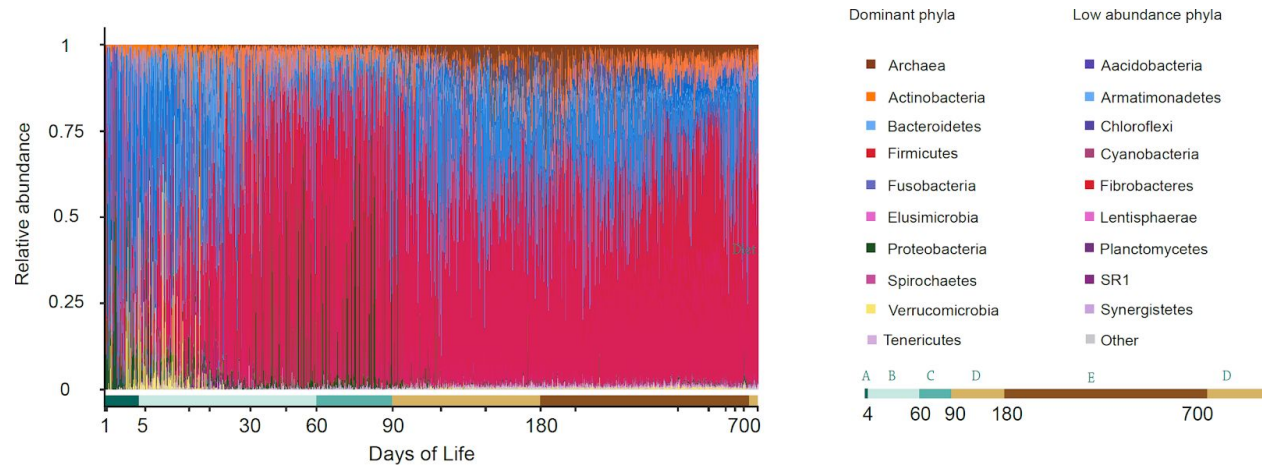
Supplementary Figure 7 - **Most core successional species appear as early as the first week of life.** A histogram describing time of first appearance of core successional microbes (n = 2544 OTUs) in the host. Source data is provided as a Source Data file.



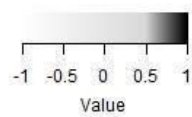
Supplementary Figure 8 - **Time is a prominent factor in the appearance of new species in the host.** Smoothed curve plot describing the number of new microbial species appearing in the rumen as a function of age (the earliest time bin in which it was observed in any given animal (see Methods: Species arrival rate). Time bins are defined in Supplementary Data 6. Source data is provided as a Source Data file.



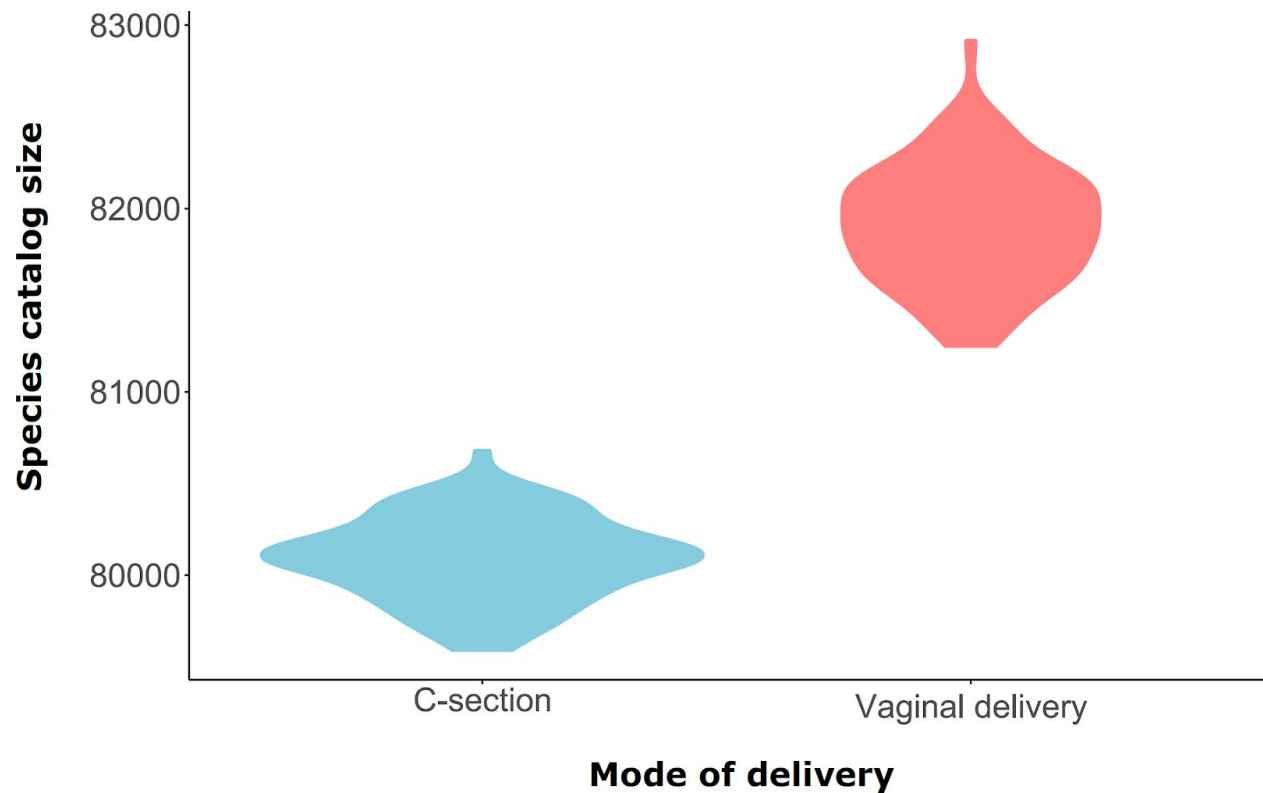
Supplementary Figure 9 - **Non-core taxa are less persistent than the core successional microbiome.** Core successional species persistence (Y-axis) as a function of time of appearance (X-axis). Non-core species were screened by the minimum number of samples in which they were detected: **A.** Minimum of 5 samples ($n = 43423$ species). **B.** Minimum of 10 samples ($n = 24533$ species). **C.** Minimum of 20 samples ($n = 13648$ species). **D.** Minimum of 30 samples ($n = 9683$ species). Each dot represents the average persistence of all microbes that appeared in the ecosystem on the specified day. Species appearance was measured over a 600-day window from first appearance, and persistence was calculated as the mean of the $\Delta(t_{\text{first appearance}} - t_{\text{last appearance}})$. Purple dots represent core successional microbes, grey dots represent non-core microbes. Source data is provided as a Source Data file.



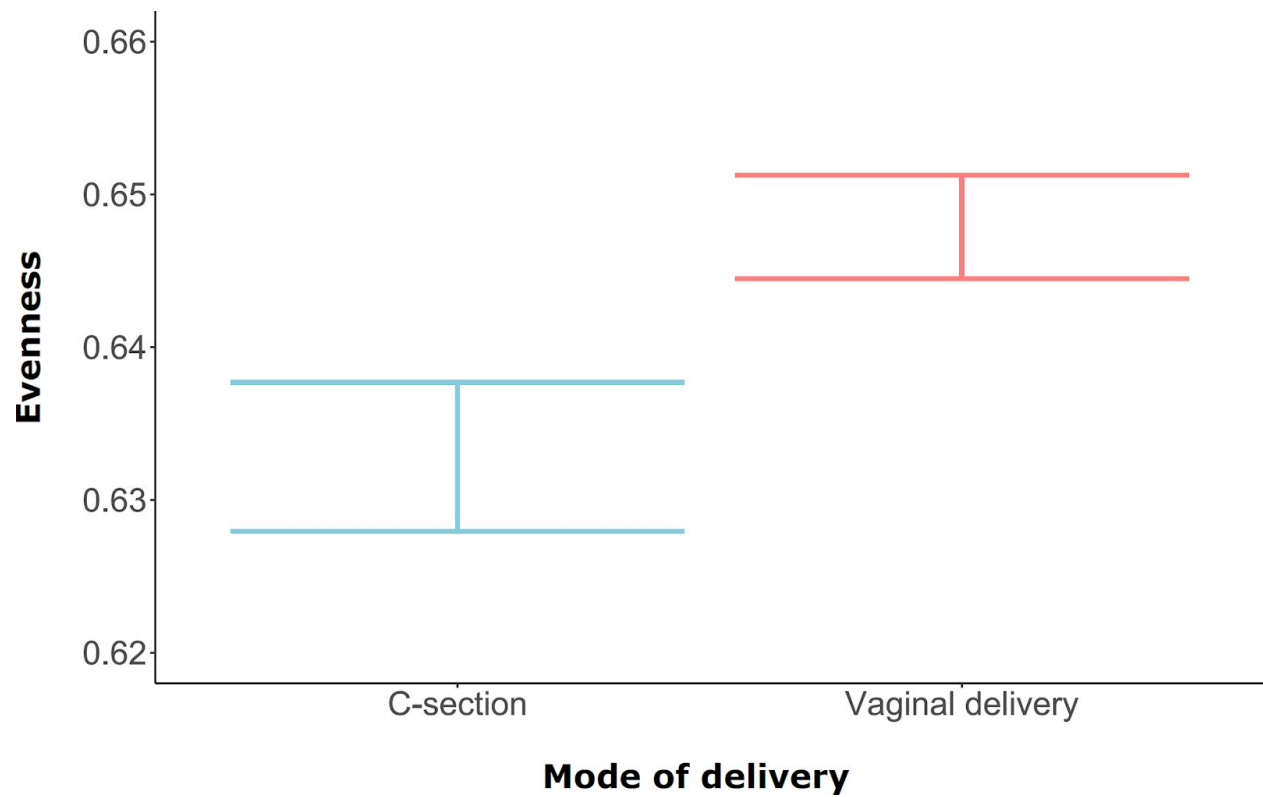
Supplementary Figure 10 - **The core successional microbiome has similar dynamics to the overall microbiome. A.** Relative abundance of the 91 microbial families for the 2544 species found to be present in at least 80% of the animals. Relative abundance values were re-calculated using these species only. Y-axis represents the re-calculated relative abundance, X-axis represents all samples ($n = 1634$), sorted by sampling day. All families belonging to the same phylum are colored by different shades of the same color. The main phyla are described on the right of the figure, see Supplementary Figure 22 for the full family level color code. Source data is provided as a Source Data file.



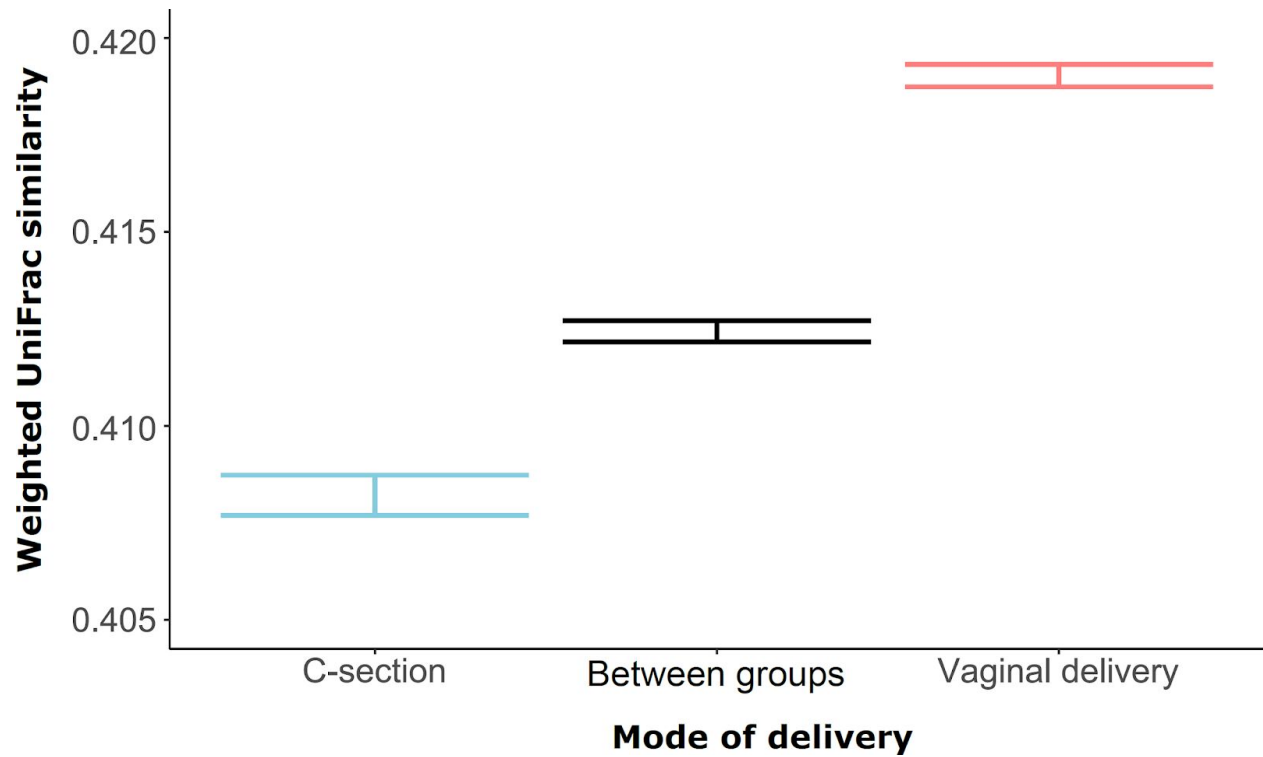
Supplementary Figure 11 - **Heat map showing species unique for each mode-of-delivery group at the first time point.** Each row represents a different OTU, each column represents the first time point for C-section animals (blue) and vaginally delivered animals (red). The Chi square test found significant dependence between mode of delivery and species appearance at time point $t = 1$ ($P < 0.05$). Source data is provided as a Source Data file.



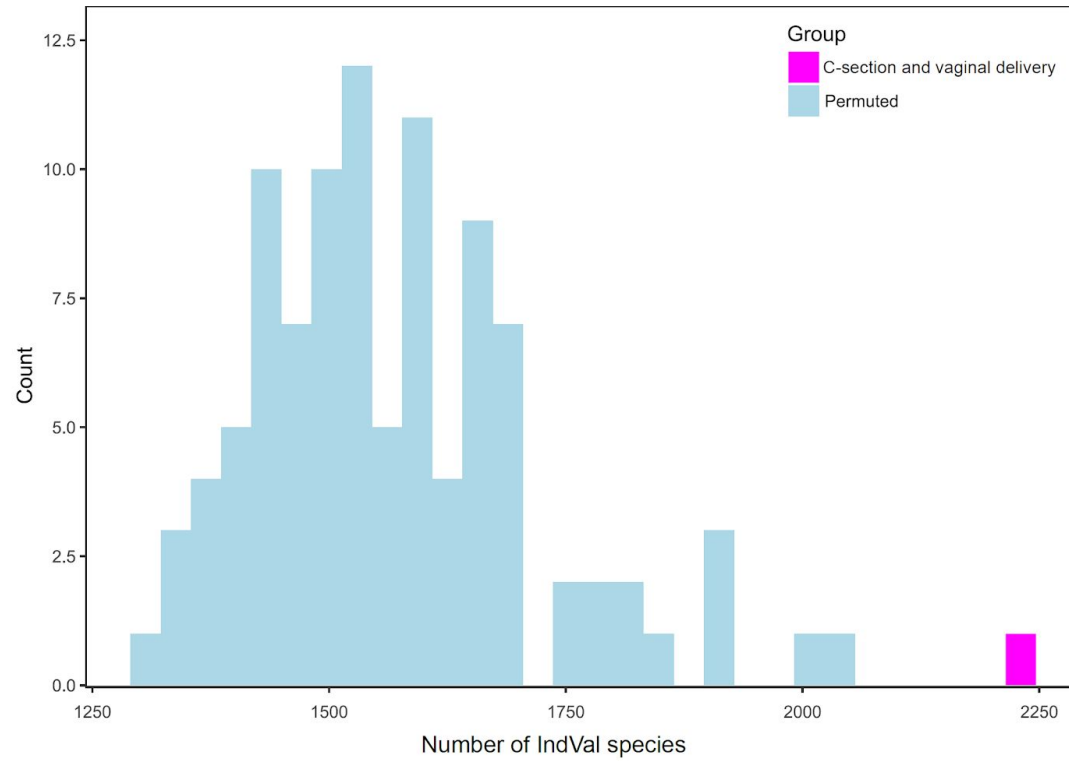
Supplementary Figure 12 - **Vaginally delivered animal sub cohort shows a richer species catalog**. Species catalog per mode of delivery, showing a higher number of species originating from the vaginal vs. C-section delivery mode. A bootstrap approach was taken, where 500 animal microbiome samples from each delivery (total 1000) were randomly picked, and species catalog size was calculated. The procedure was repeated 1000 times, and species catalog sizes were compared between delivery modes (Wilcoxon test, $P = 2.2 \times 10^{-16} < 0.0001$, two sided test). Source data is provided as a Source Data file.



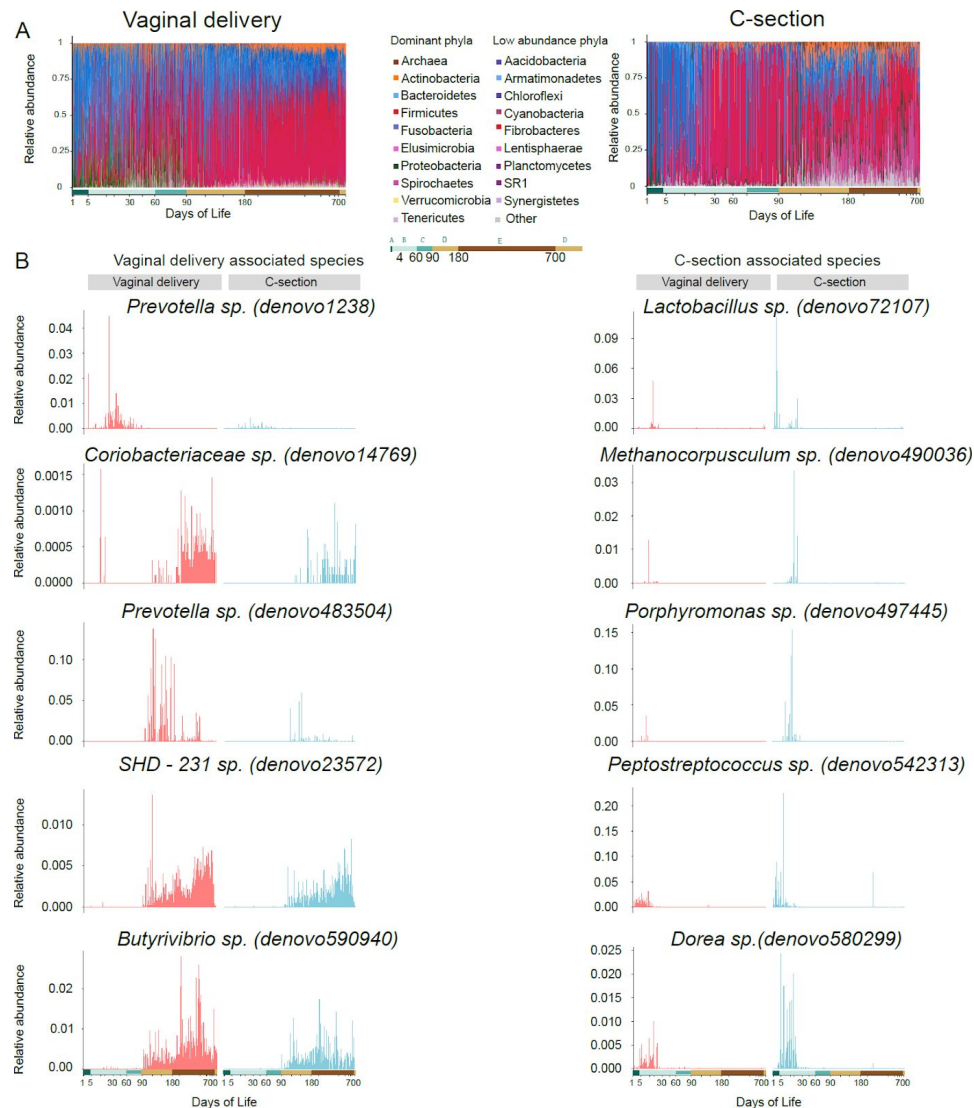
Supplementary Figure 13 - **Evenness per mode of delivery, showing higher values in vaginally delivered cows compared to C-section-delivered cows** (Wilcoxon test, $P = 0.0135$, two sided test). Source data is provided as a Source Data file.



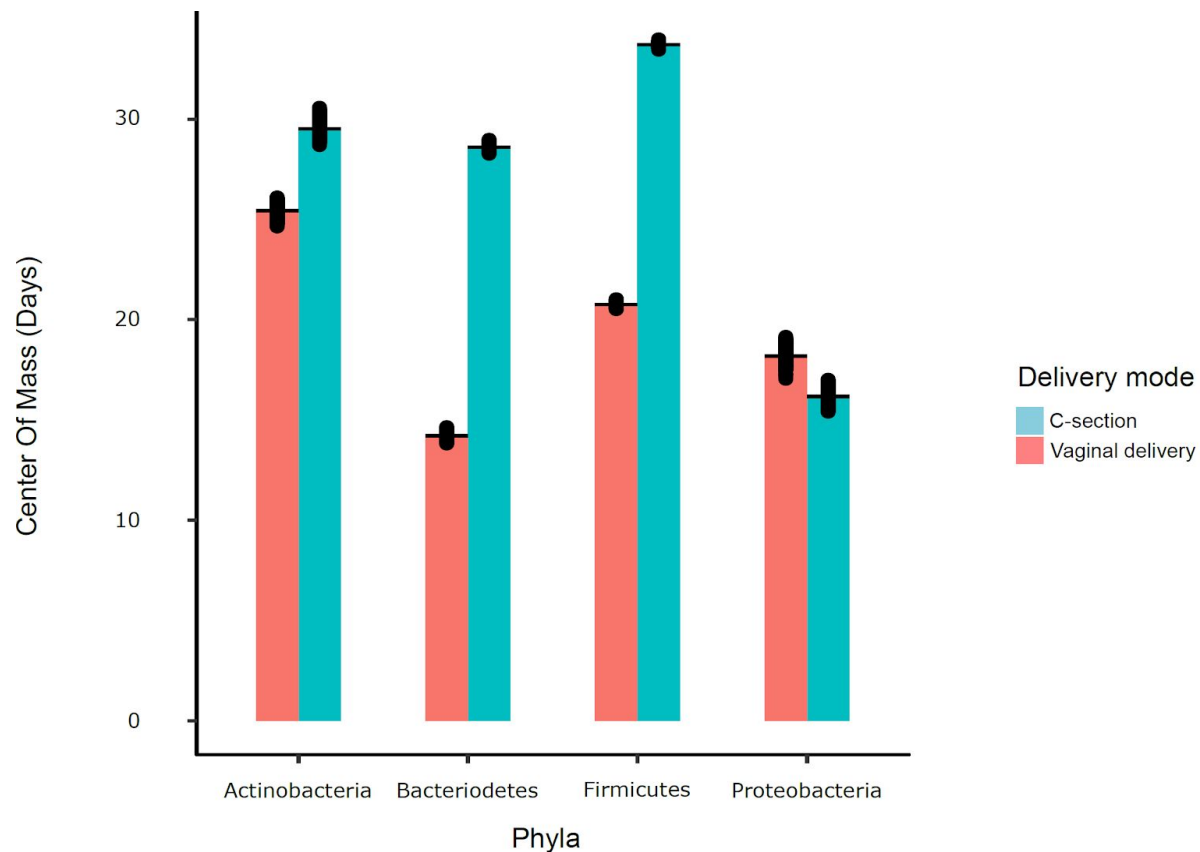
Supplementary Figure 14 - **Weighted UniFrac similarity, calculated within and between delivery-mode groups.** Higher similarity was found in the vaginally delivered cows (non-parametric Bonferroni-corrected, using 1000 Monte Carlo simulations, $P = 0.01$). Source data is provided as a Source Data file.



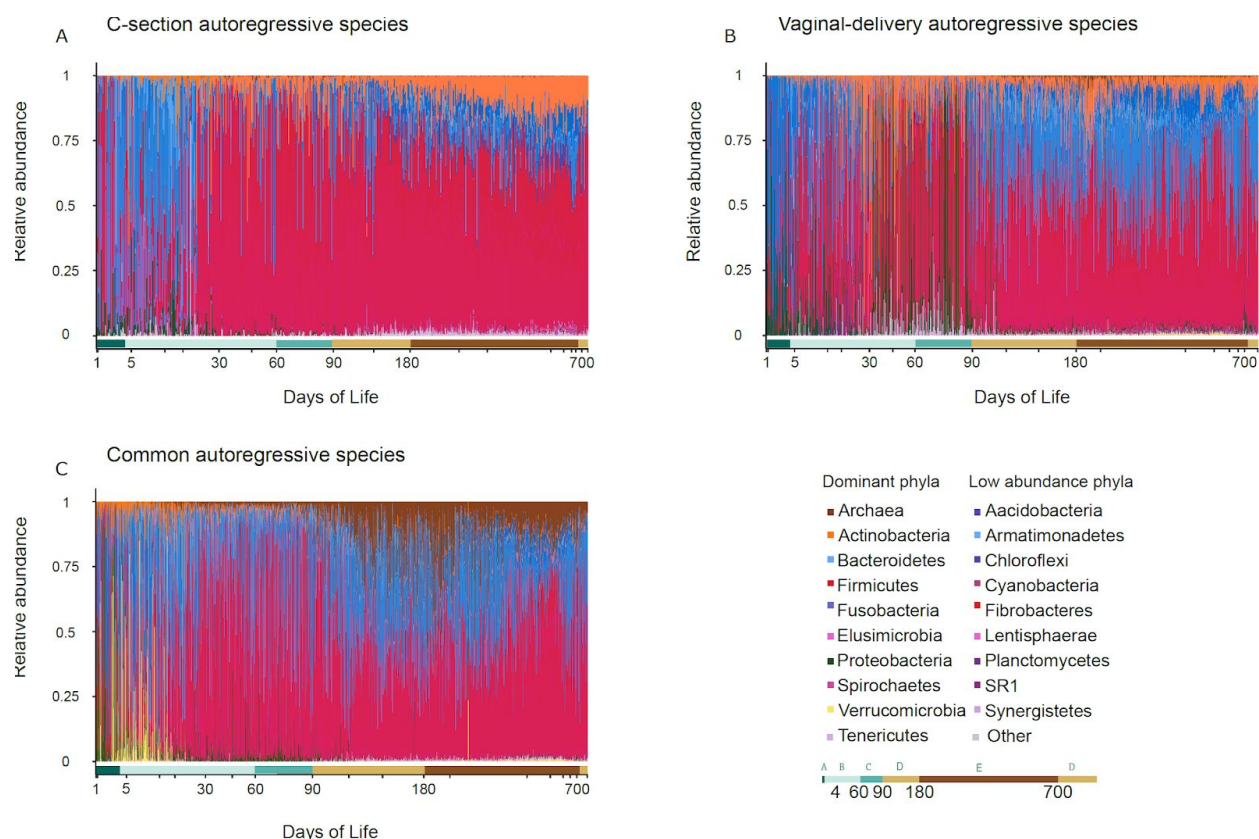
Supplementary Figure 15 - **Statistical validation for IndVal-selected species.** Number of habitat-associated species (OTUs) for C-section and vaginal delivery as the differentiating ecological identity, compared to 100 permutations where the delivery-mode labels were shuffled. The test was performed on 18 animals in each group, where nine vaginally delivered cows were omitted. Source data is provided as a Source Data file.



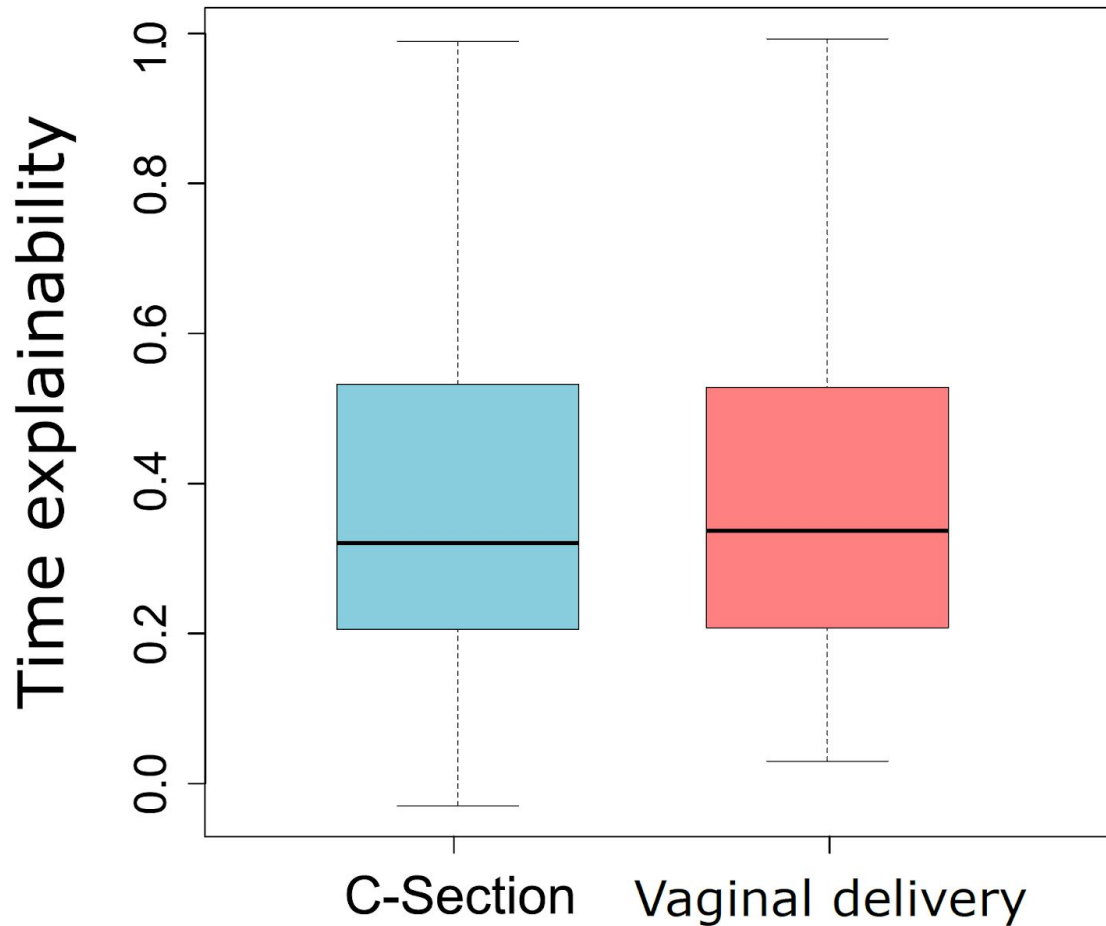
Supplementary Figure 16 - **Microbial dynamics are altered by mode of delivery.** **A.** Relative abundance of delivery mode-associated species (values were re-calculated using these species only). Left panel, relative abundance of the 81 microbial families of the 809 species found to be associated with vaginally delivered animals. Right panel, relative abundance of the 59 microbial families for the 1041 species found to be associated with C-section animals. Y-axis represents the recalculated relative abundance, X-axis represents all of the respective samples, sorted by sampling day. All families belonging to the same phylum are colored by different shades of the same color, the main phyla are described at the bottom of the figure. See Supplementary Figure 22 for the full family level color code. **B.** Relative abundance of 10 representative microbial species, generally associated with either vaginal delivery (Left panels) or C-section (Right panels). Each panel is divided into the relative abundance of the specific species in the relevant delivery mode. Y-axis, values were calculated based on the relative abundance of all microbial species. Source data is provided as a Source Data file.



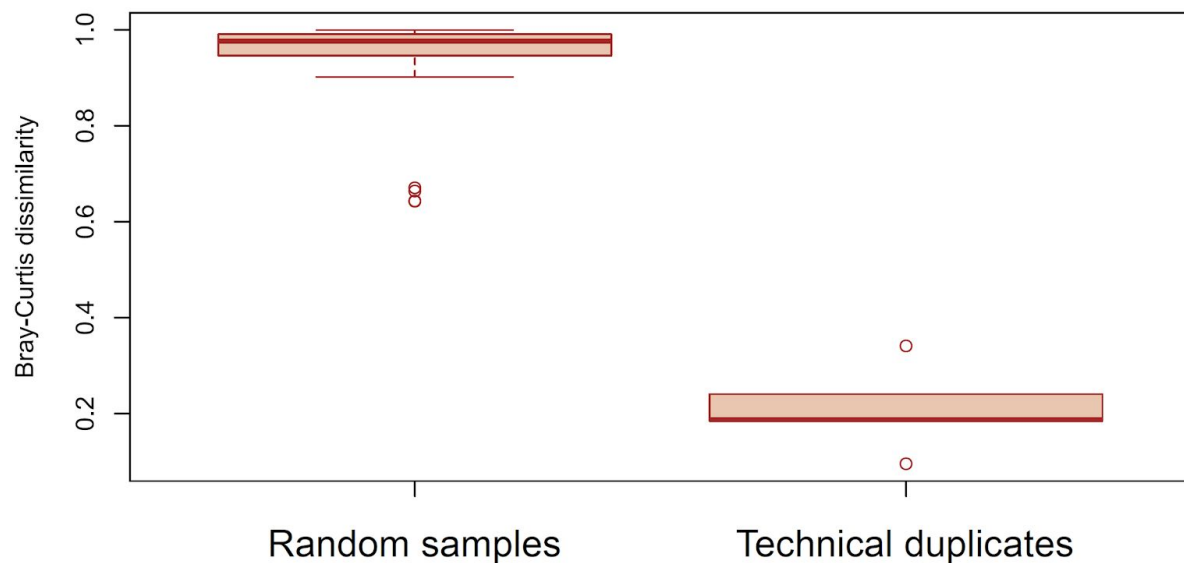
Supplementary Figure 17 - **Statistical validation for phylum-distribution analysis.** The distribution of the four main phyla (Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria) was calculated by counting the number of OTUs in each of these phyla within each time bin. The difference between the distribution of each phylum between the two modes of delivery was determined by comparing the Center Of Mass COM (see Methods: COM) between the two modes of delivery by random sampling of 70% of values, independently for C-section and vaginally delivered cows, and calculating COM (n = 1000 times). We then compared the two vectors for each phylum (1,000 COM values for C and 1,000 COM values for N) using Wilcoxon test. Error bars (horizontal lines on the top of each bar) represent the calculated standard error of mean. Source data is provided as a Source Data file.



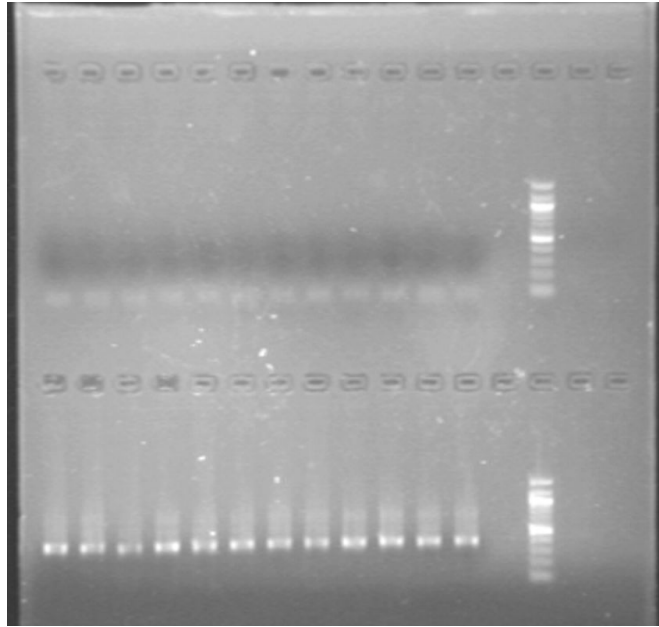
Supplementary Figure 18 - **Microbial dynamics of autoregressive species is delivery-mode dependent.** **A.** Relative abundance of the 50 microbial families for the 494 species found to be autoregressive only in C-section animals. Relative abundance values were re-calculated using these species only. Y-axis represents the re-calculated relative abundance, X-axis represents all C-section samples ($n = 572$), sorted by sampling day. All families belonging to the same phylum are colored by different shades of the same color. The main phyla are described at the bottom-right of the figure. **B.** Relative abundance of the 58 microbial families for the 494 species found to be autoregressive only in vaginally delivered animals. Relative abundance values were re-calculated using these species only. Y-axis represents the re-calculated relative abundance, X-axis represents all vaginally delivered samples ($n = 1062$), sorted by sampling day. **C.** Relative abundance of the 52 microbial families for the 401 species found to be autoregressive in both C-section and vaginally delivered animals. Relative abundance values were re-calculated using these species only. Y-axis represents the re-calculated relative abundance, X-axis represents all samples ($n = 1634$), sorted by sampling day. See Supplementary Figure 22 for the full family level color code. Source data is provided as a Source Data file.



Supplementary Figure 19 - **Time explainability is similar for the two delivery-mode groups.** Time explainability is defined as the dependence of OTU ($n = 494$ OTUs for C-Section, $n = 494$ OTUs for Vaginal delivery) relative abundance on the previous time point's community composition ($t-1$) (see Methods). In the data shown as box plots, the horizontal line in the box represents the median and whiskers indicate the lowest and highest point within 1.5 interquartile ranges of the lower or upper quartile, respectively. Source data is provided as a Source Data file.



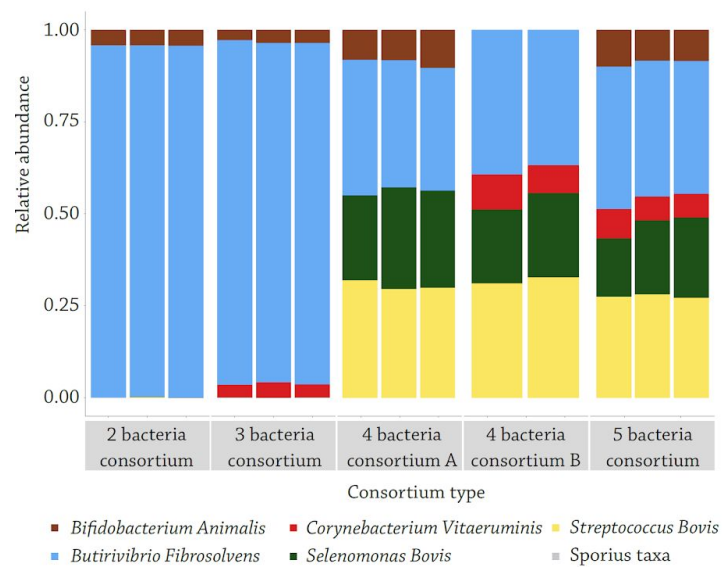
Supplementary Figure 20 - **Technical duplicates were more similar to each other than to other biological samples.** Bray-Curtis dissimilarity showed that technical duplicates ($n = 5$) were more similar to each other than to other random biological samples (t-test, $P = 2.7 \times 10^{-19} < 0.001$, two sided test). Boxes represent the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the horizontal line inside the box defines the median. Whiskers represent the lowest and highest values within 1.5 times the IQR from the first and third quartiles, respectively. Source data is provided as a Source Data file.



Supplementary Figure 21 - **Example for contamination check performed during product amplification.** DNA gel performed on PCR products from the extracted DNA of rumen samples. Top row in the gel represents control wells, and the bottom row represents PCR products of rumen samples.

f Nitrososphaeraceae	f Gaiellaceae	f Sporolactobacillaceae	c Alphaproteobacteria	f Shewanellaceae
f Halobacteriaceae	o Solirubrobacterales	f Staphylococcaceae	o BD7-3	f Cardiobacteriaceae
f Methanobacteriaceae	f Conexibacteraceae	f Thermoactinomyctaceae	f Caulobacteraceae	o Chromatiales
f Methanocorpusculaceae	f Patulibacteraceae	o Bacillales	o RF32	f Chromatiaceae
f Methanosaetaceae	f Solirubrobacteraceae	o Gemellales	o Rhizobiales	f Enterobacteriaceae
f Methanosarcinaceae	f [Fimbrimonadaceae]	o Lactobacillales	f Aurantimonadaceae	o Legionellales
f [Methanomassiliococcaceae]	o RB046	f Aerococcaceae	f Beijerinckiaceae	f Coxiellaceae
f Deinococcaceae	f Rhodothermaceae	f Carnobacteriaceae	f Bradyrhizobiaceae	f Legionellaceae
f Trueperaceae	o [Saprospirales]	f Enterococcaceae	f Brucellaceae	f Methylococcaceae
o DS-100	f Chitinophagaceae	f Lactobacillaceae	f Hyphomicrobiaceae	f Halomonadaceae
f Ellin6075	f Saprospiraceae	f Leuconostocaceae	f Methylobacteriaceae	f Pasteurellaceae
o iii1-15	o Bacteroidales	f Streptococcaceae	f Methylocystaceae	f Moraxellaceae
f RB40	f [Barnesiellaceae]	o Lactobacillales	f Phyllobacteriaceae	f Pseudomonadaceae
o Acidimicrobiales	f [Paraprevotellaceae]	f Turicibacteraceae	f Rhizobiaceae	f Pseudoalteromonadaceae
f AKIW874	f Bacteroidaceae	c Bacilli	f Xanthobacteraceae	f Vibrionaceae
f C111	f BS11	c Clostridia	o Rhizobiales	f Sinobacteraceae
f Iamiaceae	f Marinilabiaceae	o Clostridiales	f Rhodobacteraceae	f Xanthomonadaceae
f Microthrixaceae	f p-2534-18B5	f [Acidaminobacteraceae]	f Acetobacteraceae	c Gammaproteobacteria
o Actinomycetales	f Porphyromonadaceae	f [Mogibacteriaceae]	f Rhodospirillaceae	o PHOS-HD29
f Actinomycetaceae	f Prevotellaceae	f [Tissierellaceae]	o Rickettsiales	o PL-11B10
f Actinopolysporaceae	f RF16	f Caldicoprobacteraceae	f mitochondria	c Spirochaetes
f Actinosynnemataceae	f Rikenellaceae	f Christensenellaceae	o Sphingomonadales	o M2PT2-76
f Beutenbergiaceae	f S24-7	f Clostridiaceae	f Erythrobacteraceae	f Spirochaetaceae
f Bogoriellaceae	o Bacteroidales	f Dehalobacteriaceae	o Sphingomonadales	p SR1
f Brevibacteriaceae	f Cyclobacteriaceae	f ET0H8	c Alphaproteobacteria	f Dethiosulfovibrionaceae
f Cellulomonadaceae	f Cytophagaceae	f Eubacteriaceae	o Burkholderiales	f Synergistaceae
f Corynebacteriaceae	f Flammeovirgaceae	f Gracilbacteraceae	f Alcaligenaceae	c CK-1C4-19
f Dermabacteraceae	f [Weeksellaceae]	f JTB215	f Burkholderiaceae	c Mollicutes
f Dermacoccaceae	f Cryomorphaceae	f Lachnospiraceae	f Comamonadaceae	o Acholeplasmatales
f Dermatophilaceae	f Flavobacteriaceae	f Peptococcaceae	f Oxalobacteraceae	f Acholeplasmataceae
f Dietziaceae	o Sphingobacteriales	f Peptostreptococcaceae	o Burkholderiales	f Anaeroplasmataceae
f Geodermatophilaceae	f Sphingobacteriaceae	f Symbiobacteriaceae	o Ellin6067	f Mycoplasmataceae
f Glycomycetaceae	c VC2_1_Bac22	f Syntrophomonadaceae	o Methylophilales	o RF39
f Gordoniaceae	c PRR-11	f Veillonellaceae	f Methylophilaceae	o ML615J-28
f Intrasporangiaceae	f Rhabdochlamydiaceae	o Clostridiales	f Neisseriaceae	c SJA-4
f Jonesiaceae	f Anaerolinaceae	f Halanaerobiaceae	f Rhodocyclaceae	c TM7-3
f Microbacteriaceae	f A4b	o MBA08	c Betaproteobacteria	f F16
f Micrococcaceae	f SHA-31	o Natranaerobiales	f Bacteriovoracaceae	o [Pedosphaerales]
f Micromonosporaceae	c Ellin6529	f Anaerobranchaceae	f Bdellovibrionaceae	f R4-41B
f Mycobacteriaceae	c Gitt-GS-136	o OPB54	f Desulfobulbaceae	f [Chthoniobacteraceae]
f Nocardioideae	o AKYG1722	o Thermoanaerobacterales	f Desulfovibrionaceae	c Opitutae
f Nocardioseae	o JG30-KF-CM45	f Erysipelotrichaceae	f Pelobacteraceae	c [Cerasiococcaceae]
f Promicromonosporaceae	o Sphaerobacterales	p Firmicutes	o GMD14H09	c Opitutae
f Pseudonocardioideae	o B07_WMSP1	o Fusobacteriaceae	o Myxococcales	o LD1-PB3
f Ruaniaceae	o YS2	f Leptotrichiaceae	o Myxococcales	o WCHB1-41
f Sanguibacteraceae	o Streptophyta	c Gemm-1	o PB19	f RFP12
f Streptomycetaceae	c ML635J-21	c Gemm-3	f Syntrophobacteraceae	o WCHB1-25
f Streptosporangiaceae	f Elusimicrobiaceae	c Gemm-5	f Campylobacteraceae	o WCHB1-41
f Thermomonosporaceae	c Endomicrobia	c 3BR-5F	f Helicobacteraceae	f Verrucomicrobiaceae
f Williamsiaceae	p FBP	c BD1-5	f Aeromonadaceae	p WPS-2
f Yaniellaceae	f Fibrobacteraceae	p LD1	f Succinivibrionaceae	Other
o Actinomycetales	f TSCOR003-O20	f Victivallaceae	o Alteromonadales	
f Bifidobacteriaceae	f [Exiguobacteraceae]	c TSBW08	f [Chromatiaceae]	
o Bifidobacteriales	f Bacillales	o BA021	f 211ds20	
f Coriobacteriaceae	f Listeriaceae	f Isosphaeraceae	f Alteromonadaceae	
f Euzebyaceae	f Paenibacillaceae	f Pirellulaceae	f Idiomarinaceae	
c OPB41	f Planococcaceae	f Planctomycetaceae	f OM60	
f Rubrobacteraceae				

Supplementary Figure 22 - Color code for all 291 microbial families.



Consortium	Bacteria inserted	Bacteria detected
2-bacteria consortium	<i>Bifidobacterium animalis</i>	<i>Bifidobacterium animalis</i>
	<i>Butyrivibrio fibrosolvens</i>	<i>Butyrivibrio fibrosolvens</i>
3-bacteria consortium	<i>Bifidobacterium animalis</i>	<i>Bifidobacterium animalis</i>
	<i>Butyrivibrio fibrosolvens</i>	<i>Butyrivibrio fibrosolvens</i>
	<i>Corynebacterium vitaeruminis</i>	<i>Corynebacterium vitaeruminis</i>
4- bacteria A consortium	<i>Bifidobacterium animalis</i>	<i>Bifidobacterium animalis</i>
	<i>Butyrivibrio fibrosolvens</i>	<i>Butyrivibrio fibrosolvens</i>
	<i>Selenomonas bovis</i>	<i>Selenomonas bovis</i>
	<i>Streptococcus bovis</i>	<i>Streptococcus bovis</i>
4-bacteria B consortium	<i>Butyrivibrio fibrosolvens</i>	<i>Butyrivibrio fibrosolvens</i>
	<i>Corynebacterium vitaeruminis</i>	<i>Corynebacterium vitaeruminis</i>
	<i>Selenomonas bovis</i>	<i>Selenomonas bovis</i>
	<i>Streptococcus bovis</i>	<i>Streptococcus bovis</i>
5-bacteria consortium	<i>Bifidobacterium animalis</i>	<i>Bifidobacterium animalis</i>
	<i>Butyrivibrio fibrosolvens</i>	<i>Butyrivibrio fibrosolvens</i>
	<i>Corynebacterium vitaeruminis</i>	<i>Corynebacterium vitaeruminis</i>
	<i>Selenomonas bovis</i>	<i>Selenomonas bovis</i>
	<i>Streptococcus bovis</i>	<i>Streptococcus bovis</i>

Supplementary Figure 23 - **Relative abundance of sequenced bacteria within the different consortia.** Five different bacterial consortia were created, each consortium was repeated 3 times (except for the 4-bacteria consortium B). The adjacent table describes the bacteria inserted and identified in each consortium. Source data is provided as a Source Data file.