PNAS

Supporting Information for

Eco-evolutionary processes structure milk microbiomes across the mammalian tree of life

Mia M. Keady, Randall R. Jimenez, Morgan Bragg, Jenna Wagner, Sally L. Bornbusch, Michael L. Power, Carly R. Muletz-Wolz

Corresponding author: Carly R. Muletz-Wolz
Email: muletzc@si.edu

This PDF file includes:

Supporting text
Figures S1 to S11
Tables S1 to S7
Legends for Datasets S1
SI References

Other supporting materials for this manuscript include the following:

Datasets S1 Structural equation modeling output.

Supporting Information Text

Milk nutrient content analysis

Briefly, dried samples were combusted in an elemental gas analyzer (Model 2400, Perkin Elmer, Norwalk, CT) to determine total nitrogen (TN) content. TN was used to estimate protein in each milk sample using a conversion factor of 6.38 (3). Fat was assayed using a micro-modification of the Rose-Gottlieb procedure that involves sequential lipid extraction with ethanol, diethyl ether, and petroleum ether (4). Sugar was measured using the phenol-sulphuric acid colorimetric procedure with lactose monohydrate standards and was read at 490 nm on a microplate reader (Model ELX808, Biotek, Winooski, VT) (5, 6).

Milk gross energy (GE) was calculated for each milk sample as the sum of the energy from protein, fat, and sugar using: 5.86 kcal/g for protein, 9.11 kcal/g for fat, and 3.95 kcal/g for sugar (7). This method of GE calculation has been shown to closely correlate with experimentally measured gross energy using adiabatic bomb calorimetry for milk from species as diverse as aardvarks (Orycteropus afer) (8), bongos (Tragelaphus eurycerus) (9), and rhesus macaques (Macaca mulatta) (10). We calculated the mg/kcal GE for each nutrient (protein, sugar, and fat) by dividing the percent nutrient by total GE and multiplying by 1,000. We hereon refer to the nutrient mg/kcal GE as protein GE, sugar GE and fat GE. In analyses, we were unable to use fat GE as a factor as it was confounded with protein GE and sugar GE (linear correlation: fat-protein = -0.61, far-sugar = -0.66).

For milk nutrient content, we did not have data for 10 individuals because we did not have enough milk to run assays. For individuals for which we had milk nutrient content data (n = 73), we conducted k-means cluster analysis on sugar GE and protein GE to assign individuals to three clusters (based on scree plot analysis); for individuals without data (n = 10), we assigned them their cluster based on individuals within their species (n = 7) (11) or based on expert opinion (n = 3; M. Power). The three clusters corresponded to high protein, high sugar, and high fat (Figure S9) and were the three categories used in categorical analyses of milk nutrient content. For quantitative analyses of milk nutrient content, sugar GE and protein GE were used.

Null model analysis

Briefly, we calculated phylogenetic beta diversity for each pairwise sample type bacterial community using β-mean-nearest taxon distance (βNTID) (12, 13). Then, we calculated βNTI from 1000 random phylogenetic trees. We applied a cut-off $|\beta NTI| > 2$ to identify pairs of communities that were more similar than expected by chance in terms of phylogeny, meaning that observed difference between communities could be determined by environmental selection (13). Then, we calculated the RCBray between pairs of bacterial communities, and combined with βNTI to generate βRCBray. If $\beta RCBray > 0.95$ it indicates community variation is influenced by dispersal limitation, while if $\beta RCBray < 0.95$ it indicates mass effects. When βRCBray is between -0.95 and +0.95 it indicates the variation of the community is determined by ecological drift (13, 14).

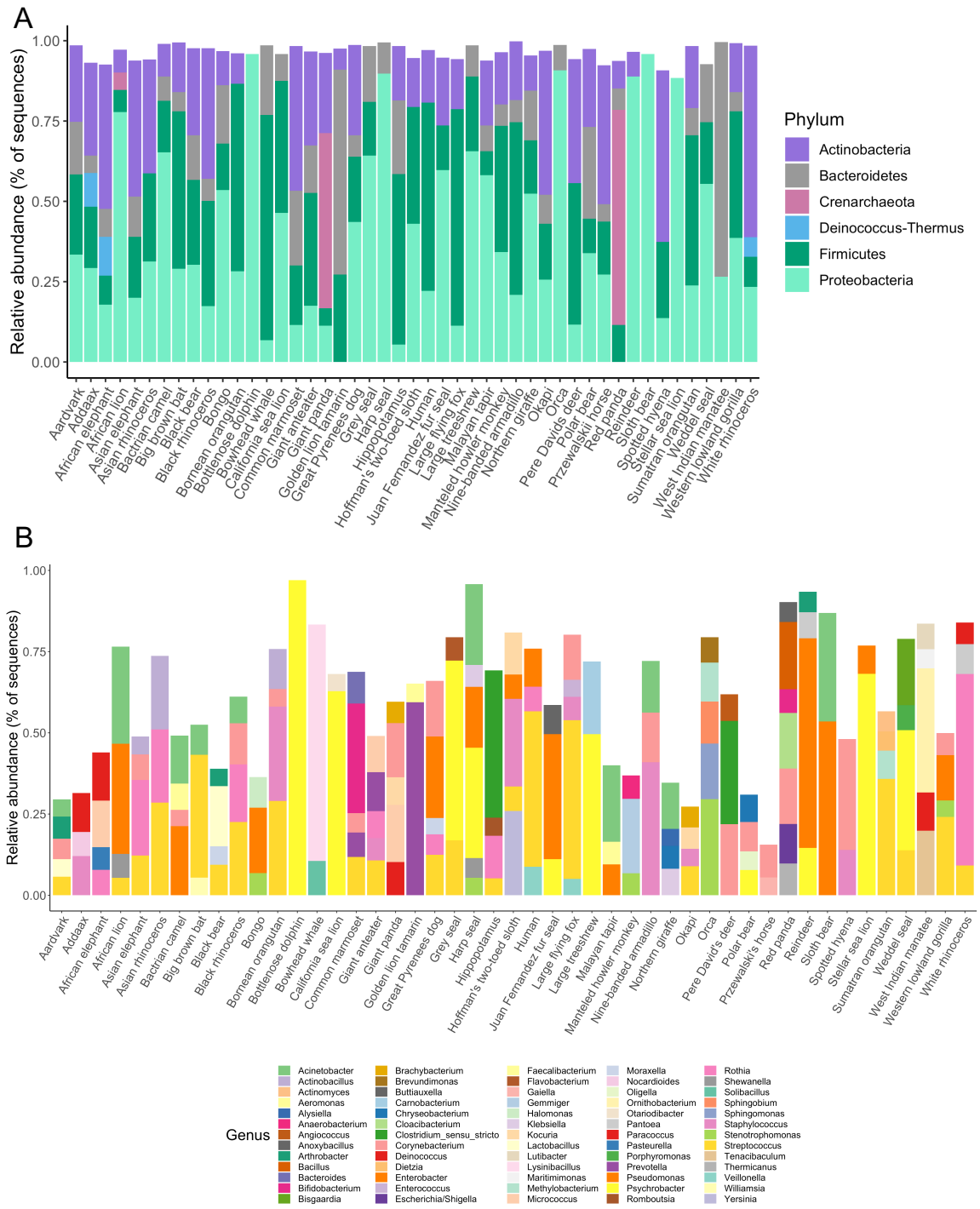


Figure S1. Stacked bar plot of relative abundance of dominant microbial (A) phyla and (B) genera among mammalian species. Full dataset, merged by common name. At least 5% relative abundance of phyla and genera.

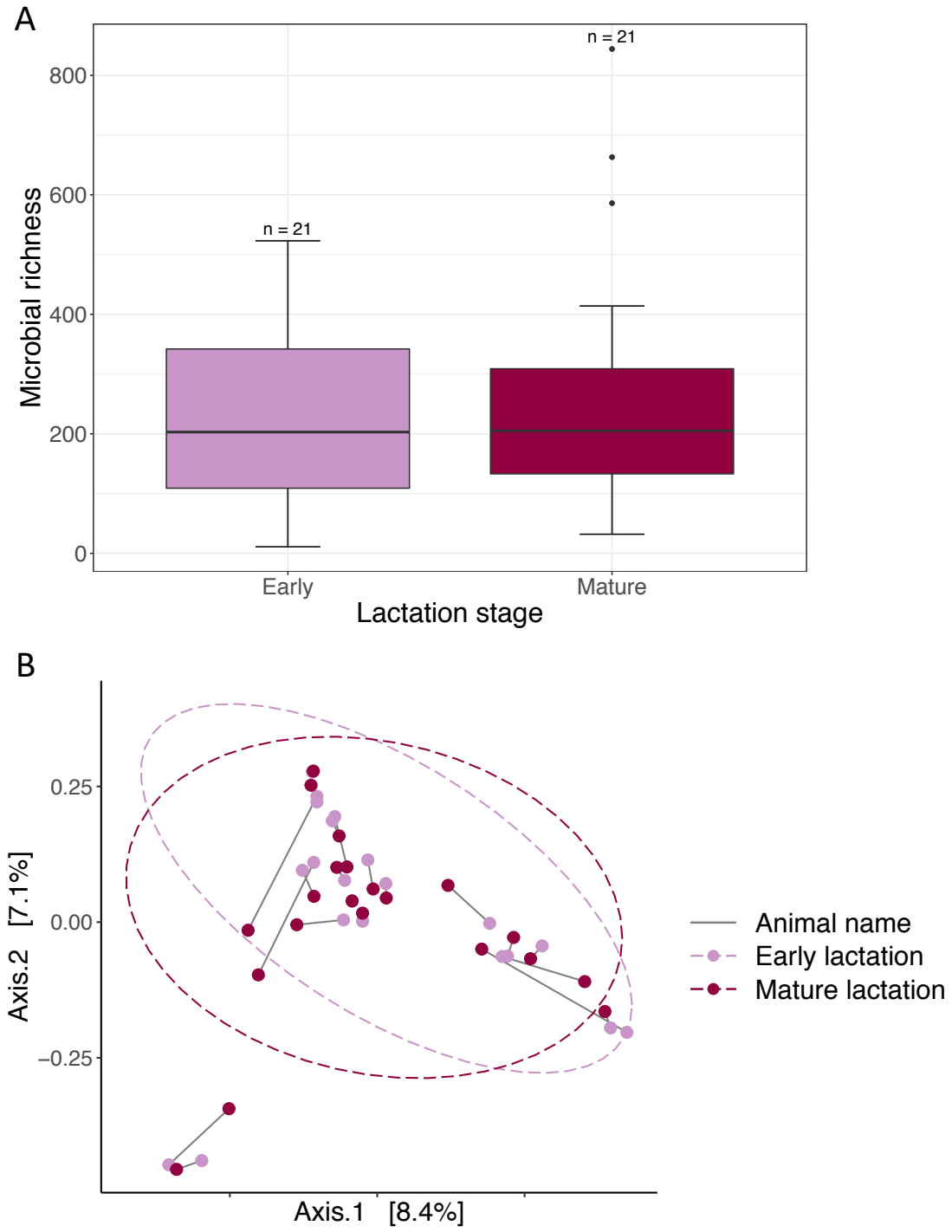


Figure S2. Similar milk microbial (A) richness and (B) composition among early and mature lactation stages (LMM $p > 0.05$; PERMANOVA $p > 0.05$).

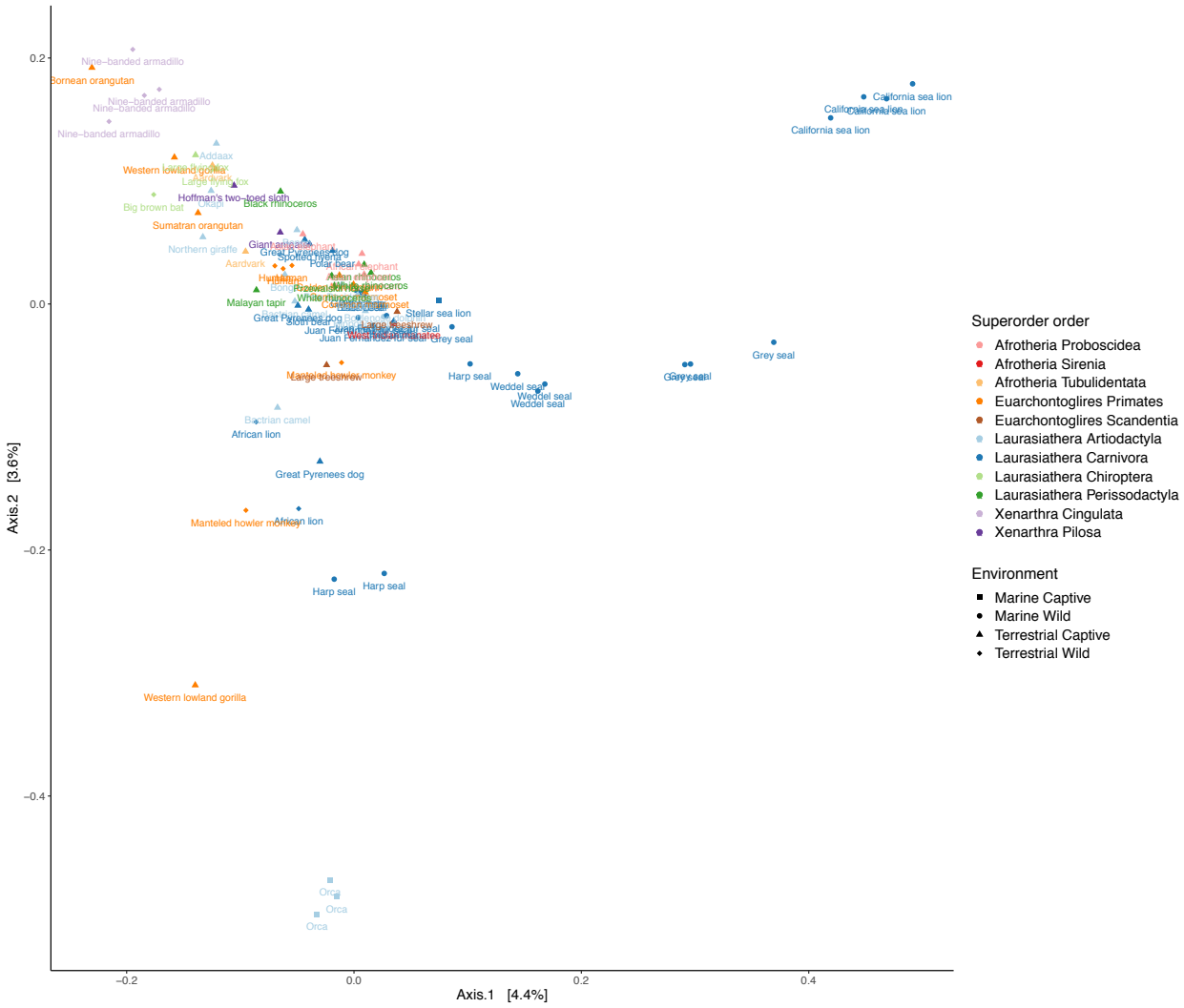


Figure S3. Principal coordinate analysis of Bray-Curtis distances for independent measures. Dataset identical to Figure 2, but labelled by common name of mammal.

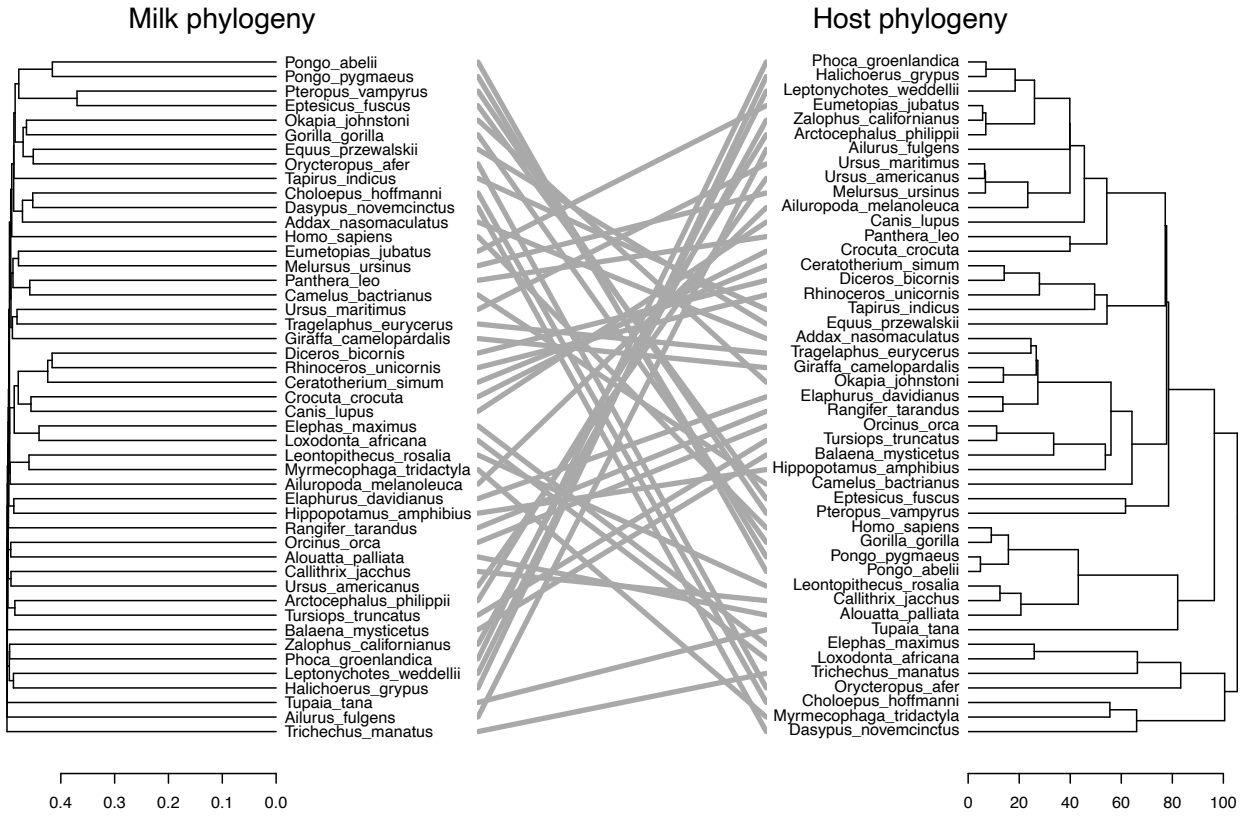
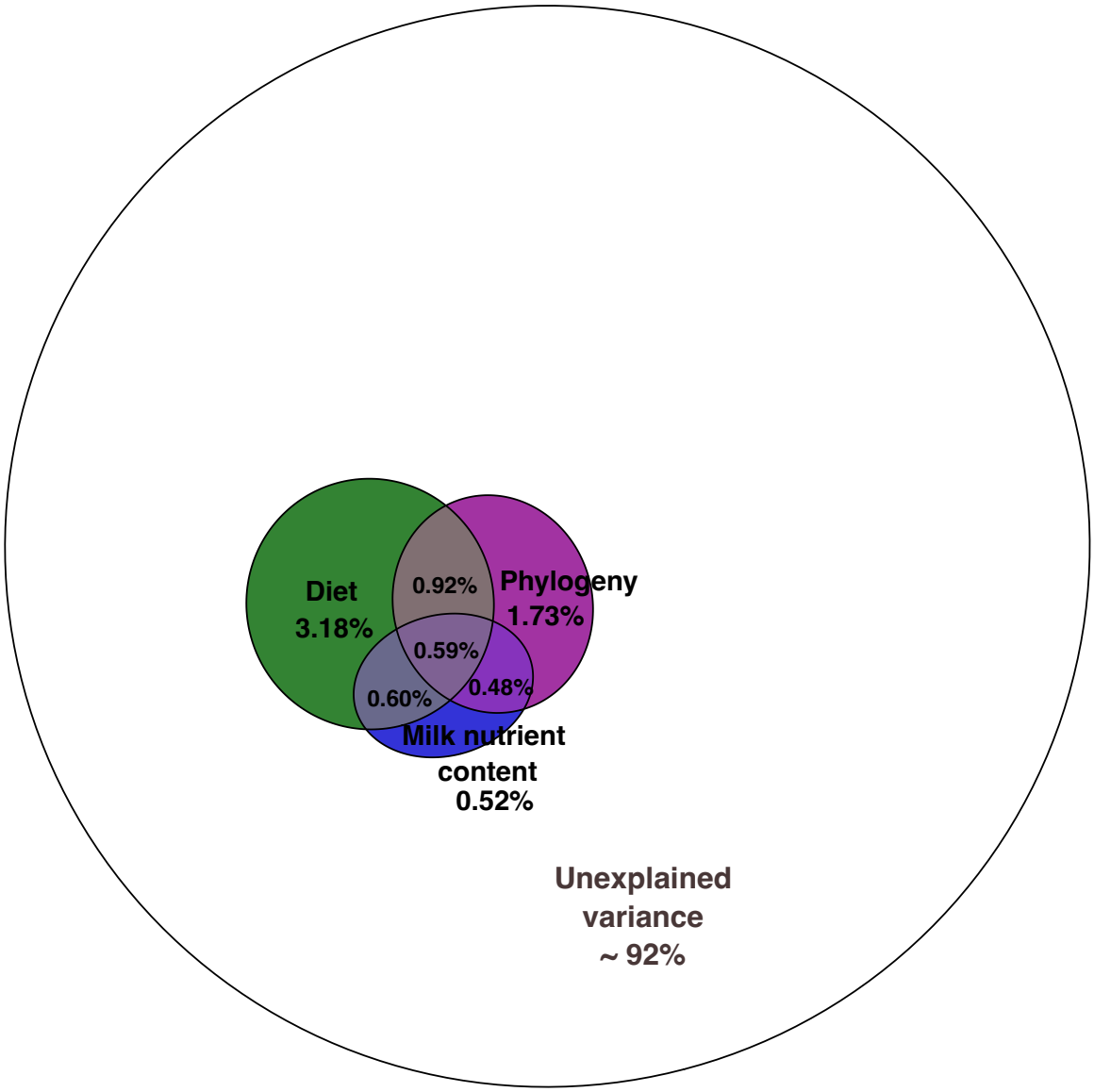


Figure S4. Topological comparisons between milk microbiome dendrogram (Jaccard distances) and host phylogeny (MYA) showing no evidence for phylosymbiosis.



Bray-Curtis

Figure S5. Individual and collective contributions of host phylogeny, diet, and milk nutrient content on milk microbiome structure (Bray-Curtis). Individual and shared variance among variables collectively explained 8% of variation (shared 2.6%, individual 5.4%).

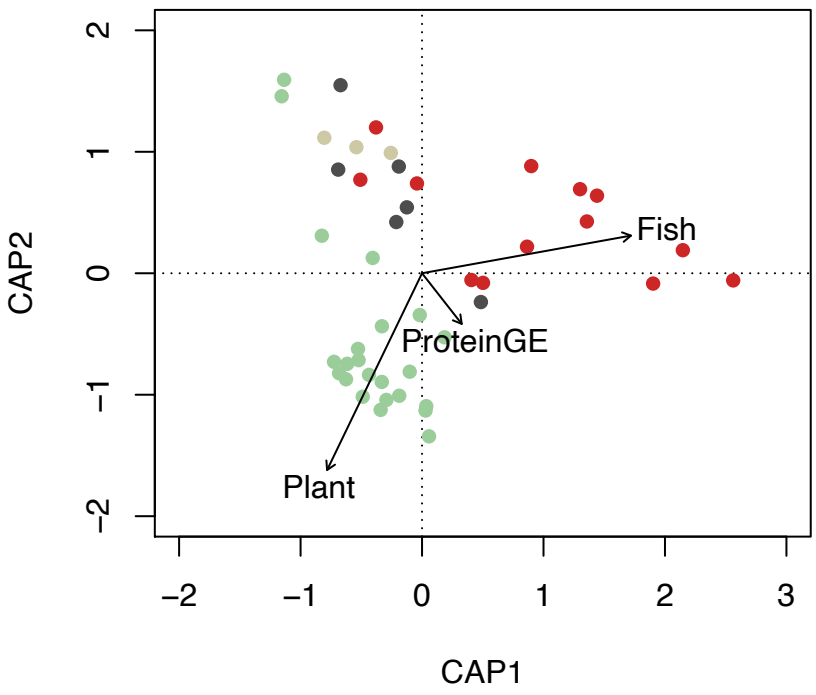


Figure S6. Constrained analysis of principal coordinates showing the relationship of milk microbial composition with host diet and milk nutrient content for Jaccard distances (similar results for UniFrac distances not shown). Individuals are shown colored based on their diet category as in Figure 2.

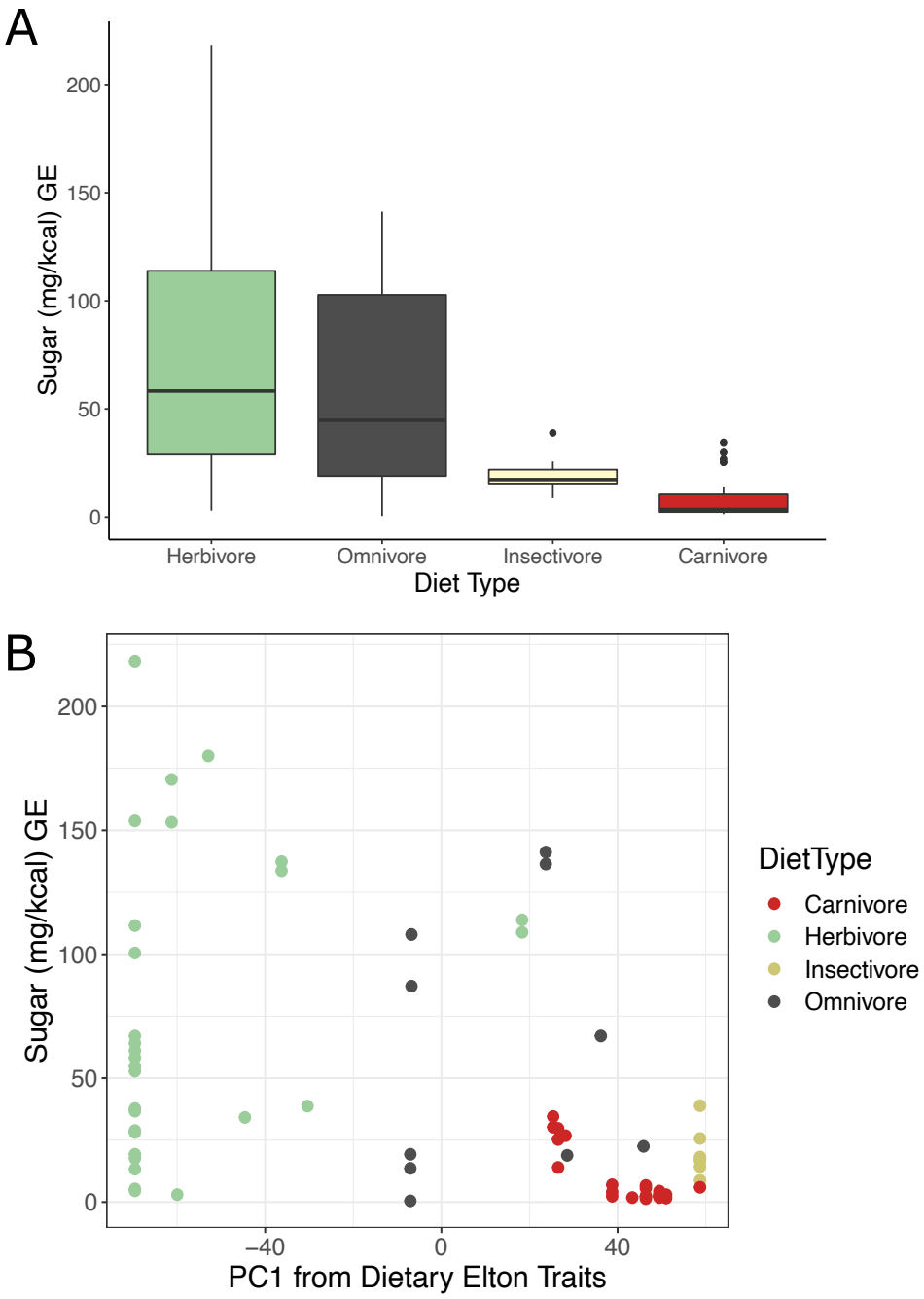


Figure S7. Relationship between diet and milk sugar GE. (A) Boxplot depicting milk sugar GE by dietary categories (herbivores, omnivores, insectivores, carnivores). (B) Principal coordinate scores (PC1) from quantitative dietary traits (Elton traits) plotted against milk sugar GE.

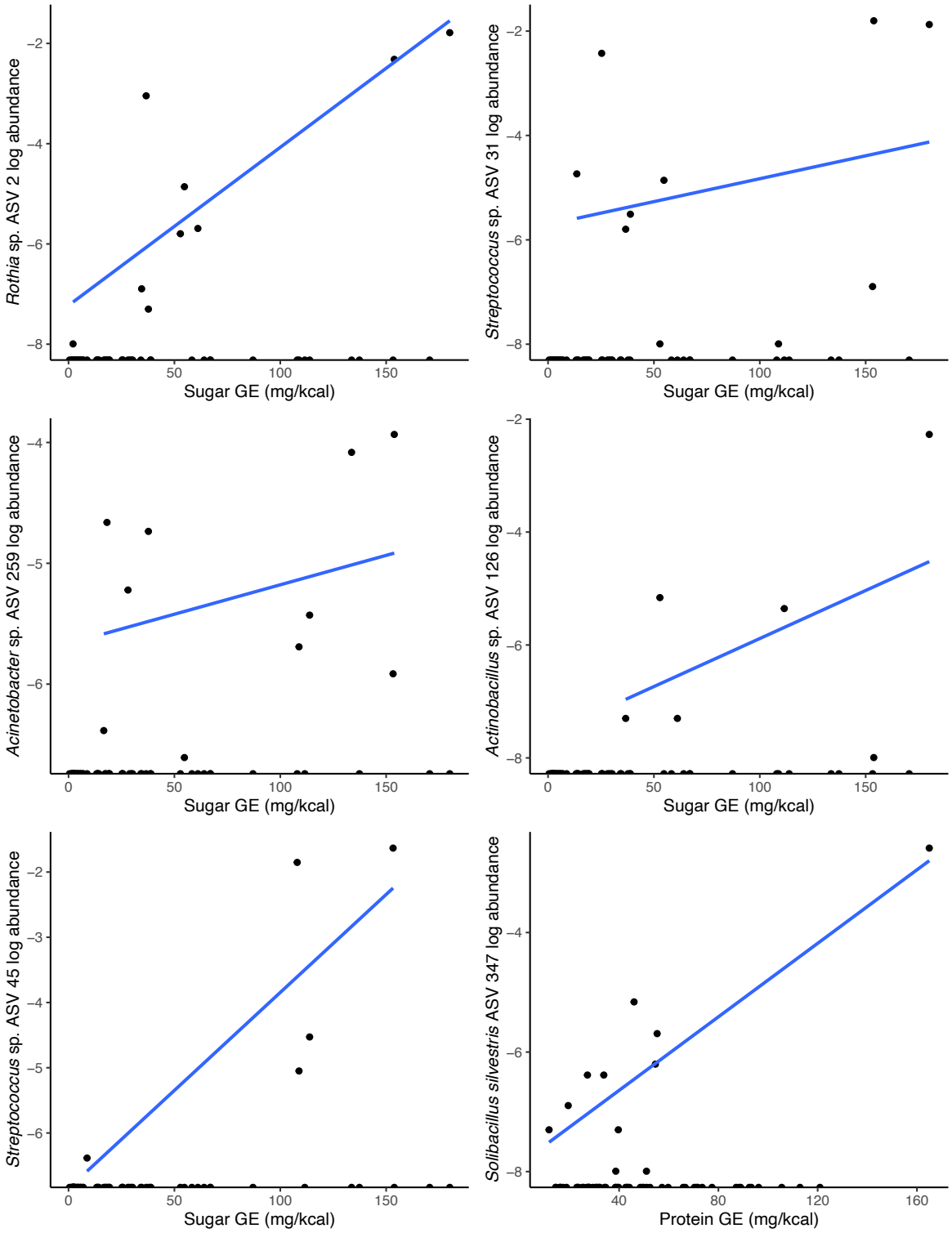


Figure S8. Correlation between microbial ASV abundance and milk sugar GE and milk protein GE. Six bacterial ASVs were correlated with milk sugar GE, and one with milk protein GE. See Table S5 for full taxonomic assignment. Microbial abundances were log transformed and had at least 5% relative abundance and occurred at least five times (n = 135 ASVs).

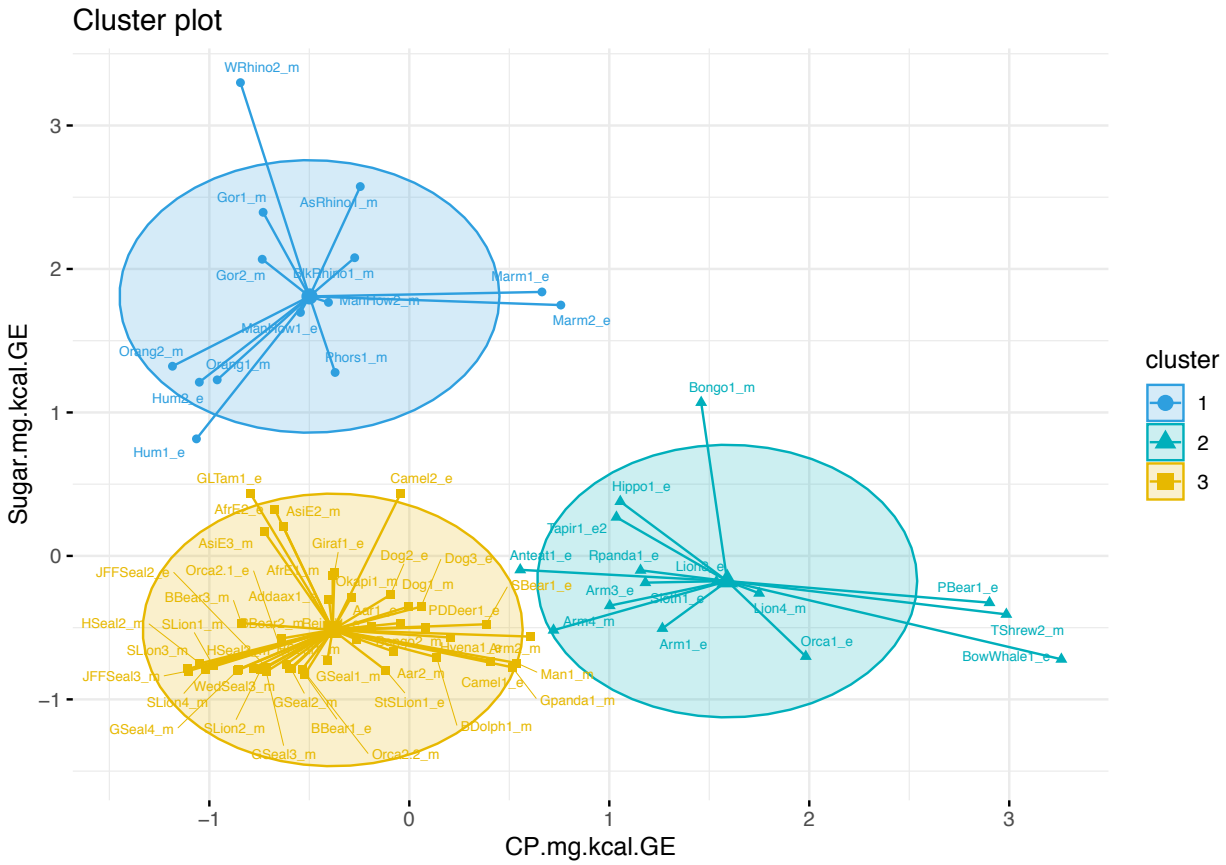


Figure S9. **Visualization of k-means clusters for milk nutrient content (milk sugar GE, milk protein GE, milk fat GE).** Plot created using function fviz_cluster (factoextra package). Confidence ellipses on euclidean distances from the center are shown. ID represents species name, unique ID and early or mature milk sample. The three clusters corresponded to high sugar (cluster 1), high protein (cluster 2), and high fat (cluster 3). Fat was not included in cluster analysis because it was highly negatively correlated with Sugar (Pearson correlation coefficient = -0.67) and Protein (Pearson correlation coefficient = -0.62).

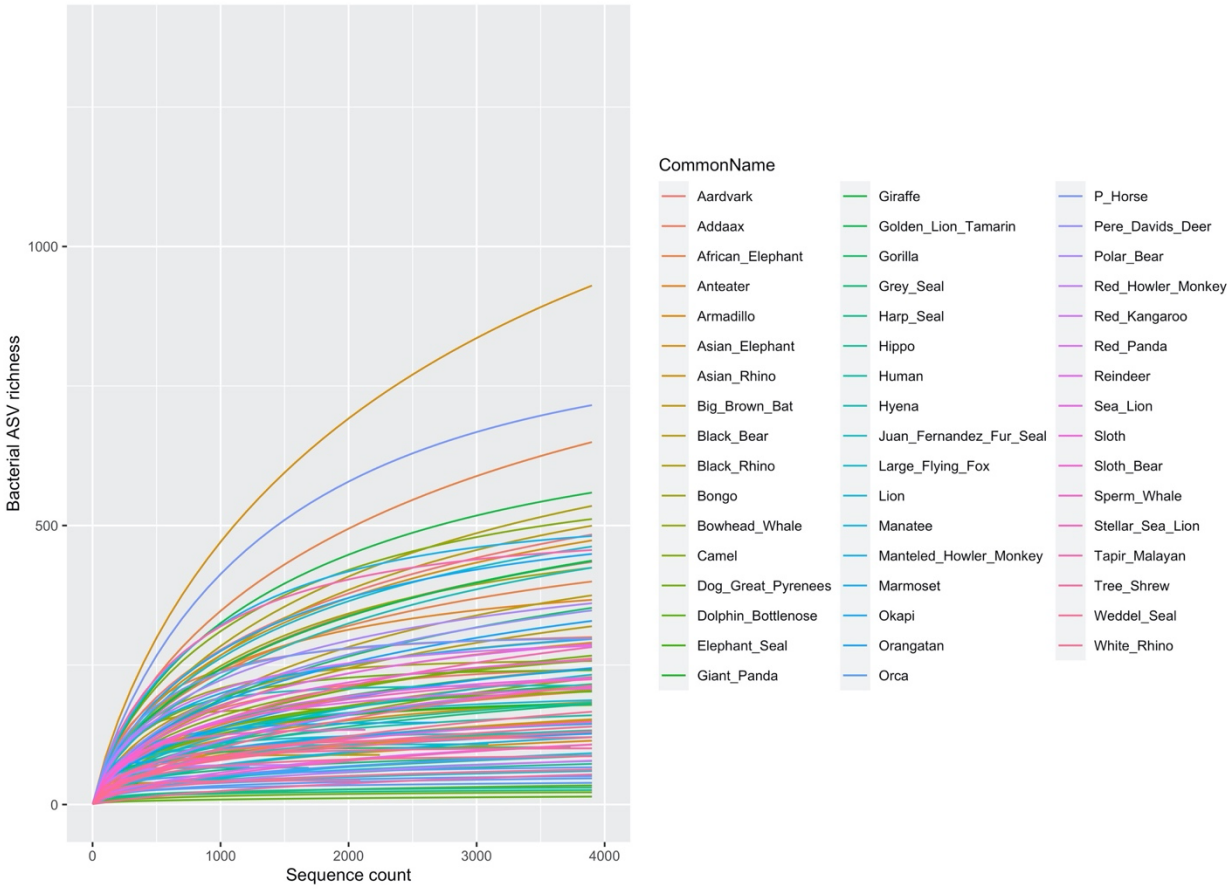


Figure S10. **Rarefaction curve depicting microbial richness by sequence count.** Samples were rarified to 3,000 total reads per sample given high variation in sequencing depth. Most host species have reached maximum microbial diversity by 3,000 reads.

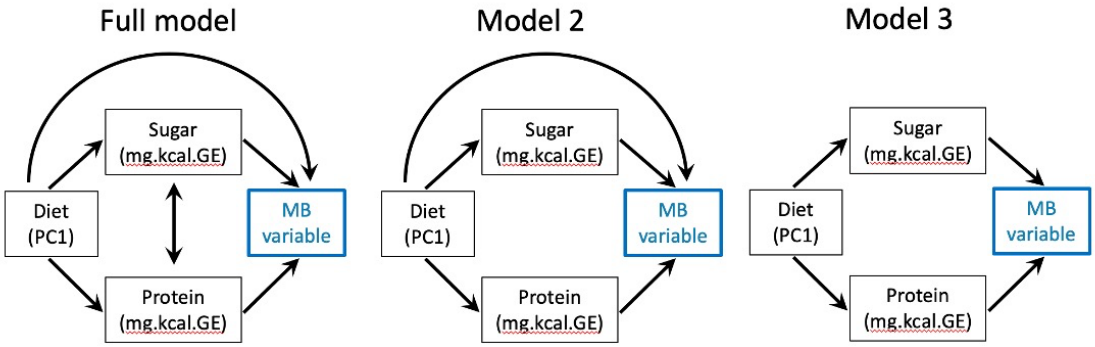


Figure S11. Three a *priori* models used in structural equation modeling (SEM). We assessed each model for four measures of microbiome structure (MB variable): species richness, Faith's phylogenetic diversity, Bray-Curtis, and unweighted UniFrac. The full model represents all possible relationships between variables, model 2 does not include a bidirectional relationship between milk sugar GE and milk protein GE, and model 3 assumes there is no direct relationship between diet and milk microbiome structure.

Table S1. **Sample sizes per species per dataset.**

Species name	Common name	Eco-evolutionary analysis	Lactation stage analysis	n
Addax nasomaculatus	Addaax	Y	N	1
Ailuropoda melanoleuca	Giant panda	Y	N	1
Ailurus fulgens	Red panda	Y	N	1
Alouatta palliata	Manteled howler monkey	Y	N	2
Arctocephalus philippii	Juan Fernandez fur seal	Y	N	3
Balaena mysticetus	Bowhead whale	Y	N	1
Callithrix jacchus	Common marmoset	N	Y	1
Callithrix jacchus	Common marmoset	Y	N	1
Callithrix jacchus	Common marmoset	Y	Y	1
Camelus bactrianus	Bactrian camel	N	Y	2
Camelus bactrianus	Bactrian camel	Y	Y	2
Canis lupus	Great Pyrenees dog	N	Y	1
Canis lupus	Great Pyrenees dog	Y	N	2
Canis lupus	Great Pyrenees dog	Y	Y	1
Ceratotherium simum	White rhinoceros	N	Y	2
Ceratotherium simum	White rhinoceros	Y	Y	2
Choloepus hoffmanni	Hoffman's two-toed sloth	Y	N	1
Crocuta crocuta	Spotted hyena	Y	N	1
Dasypus novemcinctus	Nine-banded armadillo	Y	N	4
Diceros bicornis	Black rhinoceros	Y	N	1
Elaphurus davidianus	Pere Davids deer	Y	N	1
Elephas maximus	Asian elephant	N	Y	2
Elephas maximus	Asian elephant	Y	Y	2
Eptesicus fuscus	Big brown bat	Y	N	1
Equus przewalskii	Przewalskii horse	N	Y	1
Equus przewalskii	Przewalskii horse	Y	Y	1
Eumetopias jubatus	Stellar sea lion	N	Y	1
Eumetopias jubatus	Stellar sea lion	Y	Y	1
Giraffa camelopardalis	Northern giraffe	N	Y	1
Giraffa camelopardalis	Northern giraffe	Y	Y	1
Gorilla gorilla	Western lowland gorilla	N	Y	2
Gorilla gorilla	Western lowland gorilla	Y	Y	2
Halichoerus grypus	Grey seal	Y	N	4
Hippopotamus amphibius	Hippopotamus	Y	N	1
Homo sapiens	Human	Y	N	3
Leontopithecus rosalia	Golden lion tamarin	Y	N	1
Leptonychotes weddellii	Weddel seal	Y	N	3
Loxodonta africana	African elephant	N	Y	2
Loxodonta africana	African elephant	Y	Y	2
Melursus ursinus	Sloth bear	Y	N	1

Myrmecophaga tridactyla	Giant anteater	Y	N	1
Okapia johnstoni	Okapi	Y	N	1
Orcinus orca	Orca	N	Y	2
Orcinus orca	Orca	Y	N	1
Orcinus orca	Orca	Y	Y	2
Orycteropus afer	Aardvark	N	Y	1
Orycteropus afer	Aardvark	Y	N	1
Orycteropus afer	Aardvark	Y	Y	1
Panthera leo	African lion	Y	N	2
Phoca groenlandica	Harp seal	Y	N	3
Pongo abelii	Sumatran orangutan	Y	N	1
Pongo pygmaeus	Bornean orangutan	N	Y	1
Pongo pygmaeus	Bornean orangutan	Y	Y	1
Pteropus vampyrus	Large flying fox	N	N	2
Pteropus vampyrus	Large flying fox	Y	N	2
Rangifer tarandus	Reindeer	Y	N	1
Rhinoceros unicornis	Asian rhinoceros	Y	N	1
Tapirus indicus	Malayan tapir	N	N	1
Tapirus indicus	Malayan tapir	Y	N	1
Tragelaphus eurycerus	Bongo	N	Y	1
Tragelaphus eurycerus	Bongo	Y	N	1
Tragelaphus eurycerus	Bongo	Y	Y	1
Trichechus manatus	West Indian manatee	Y	N	1
Tupaia tana	Large treeshrew	Y	N	2
Tursiops truncatus	Bottlenose dolphin	N	Y	1
Tursiops truncatus	Bottlenose dolphin	Y	N	1
Tursiops truncatus	Bottlenose dolphin	Y	Y	1
Ursus americanus	Black bear	Y	N	3
Ursus maritimus	Polar bear	Y	N	1
Zalophus californianus	California sea lion	Y	N	4
			TOTAL =	**107**

Table S2. Sample sizes per factor in repeated measures dataset. 21 independent females sampled at early and mature lactation stages.

Superorder	Diet type	Environment	n
Afrotheria	Herbivore	Terrestrial Captive	8
Afrotheria	Insectivore	Terrestrial Captive	2
Euarchontoglires	Herbivore	Terrestrial Captive	6
Euarchontoglires	Omnivore	Terrestrial Captive	2
Laurasiathera	Carnivore	Marine Captive	8
Laurasiathera	Carnivore	Terrestrial Captive	2
Laurasiathera	Herbivore	Terrestrial Captive	14
		TOTAL =	**42**

Table S3. Sample sizes per factor in independent measures dataset.

Superorder	Diet type	Environment	n
Afrotheria	Herbivore	Marine Captive	1
Afrotheria	Herbivore	Terrestrial Captive	4
Afrotheria	Insectivore	Terrestrial Captive	2
Euarchontoglires	Herbivore	Terrestrial Captive	4
Euarchontoglires	Herbivore	Terrestrial Wild	2
Euarchontoglires	Omnivore	Terrestrial Captive	5
Euarchontoglires	Omnivore	Terrestrial Wild	3
Laurasiathera	Carnivore	Marine Captive	5
Laurasiathera	Carnivore	Marine Wild	19
Laurasiathera	Carnivore	Terrestrial Captive	5
Laurasiathera	Carnivore	Terrestrial Wild	2
Laurasiathera	Herbivore	Terrestrial Captive	20
Laurasiathera	Herbivore	Terrestrial Wild	1
Laurasiathera	Omnivore	Terrestrial Captive	1
Laurasiathera	Omnivore	Terrestrial Wild	3
Xenarthra	Herbivore	Terrestrial Captive	1
Xenarthra	Insectivore	Terrestrial Captive	1
Xenarthra	Insectivore	Terrestrial Wild	4
		TOTAL =	**83**

Table S4. Summary statistics for variance partitioning performed with multiple regression on dissimilarity matrices (MRM). We assessed individual and shared variance between host phylogeny, diet, and milk nutrition on milk microbiome structure. We included samples with complete nutrient data for milk sugar GE and milk protein GE (n = 73).

Variable	Bray-Curtis		Jaccard		Unifrac	
	R2	p-value	R2	p-value	R2	p-value
Phylogeny	1.73%	0.001	1.88%	0.001	0.29%	0.213
Diet	3.18%	0.001	2.86%	0.001	1.31%	0.001
Nutrition	0.52%	0.001	0.47%	0.001	0.08%	0.339
Phylogeny, Diet, & Nutrition	0.59%	p = 0.001	0.57%	p = 0.001	0.15%	p = 0.157
Phylogeny & Diet	0.92%		0.91%		0.25%	
Phylogeny & Nutrition	0.48%		0.48%		0.08%	
Diet & Nutrition	0.60%		0.54%		0.16%	
SUM:	**8.01%**		**7.70%**		**2.32%**	

Table S5. Microbial ASV abundance (log transformed) correlated with milk sugar GE and milk protein GE.
Microbial abundances were log transformed and had at least 5% relative abundance and occurred at least five times (n = 135 ASVs).

Bacterial ASV Correlated with Milk Sugar GE

ASV	Phylum	Class	Order	Family	Genus	Species	Adjusted p-value	Coefficient
ASV154	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	*Rothia*	sp.	0.004	4.99E-05
ASV2	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	*Rothia*	sp.	0.004	2.25E-04
ASV31	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	sp.	0.01	2.51E-04
ASV259	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*	sp.	0.021	3.04E-05
ASV126	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pasteurellaceae	*Actinobacillus*	sp.	0.034	9.84E-05
ASV45	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	sp.	0.051	2.17E-04

Bacterial ASV Correlated with Milk Protein GE

ASV	Phylum	Class	Order	Family	Genus	Species	Adjusted p-value	Coefficient
ASV347	Firmicutes	Bacilli	Bacillales	Planococcaceae	*Solibacillus*	*silvestris*	0.006	1.41E-04

Table S6. **Dispersion summary statistics**. We tested distances between group centroids for each explanatory variable and microbial measure. The least dispersion between groups was observed using UniFrac in the balanced dataset.

Independent dataset (n = 83)

Variable	Distance metric	DF	F value	p value
Super Order	Bray-Curtis	79	59.522	< 0.001
	Jaccard	79	48.687	< 0.001
	UniFrac	79	6.0689	0.0009016
Environment	Bray-Curtis	79	12.184	< 0.001
	Jaccard	79	15.113	< 0.001
	UniFrac	79	4.9714	0.003271
Diet Type	Bray-Curtis	79	26.916	< 0.001
	Jaccard	79	30.463	< 0.001
	UniFrac	79	1.1123	0.3492
Milk Nutrient Content	Bray-Curtis	80	22.047	< 0.001
	Jaccard	80	19.213	< 0.001
	UniFrac	80	6.1934	0.003156
Lactation Stage	Bray-Curtis	81	< 0.001	0.791
	Jaccard	81	0.2406	0.6251
	UniFrac	81	2.3464	0.1295

Balanced dataset (n = 51)

Variable	Distance metric	DF	F value	p value
Super Order	Bray-Curtis	47	25.896	< 0.001
	Jaccard	47	17.995	< 0.001
	UniFrac	47	1.7145	0.1768
Environment	Bray-Curtis	47	6.4379	0.0009633
	Jaccard	47	6.8027	0.000667
	UniFrac	47	2.8914	0.04513
Diet Type	Bray-Curtis	47	14.095	< 0.001
	Jaccard	47	11.945	< 0.001
	UniFrac	47	0.5888	0.6254
Milk Nutrient Content	Bray-Curtis	48	27.55	< 0.001
	Jaccard	48	26.637	< 0.001
	UniFrac	48	2.7015	0.0773
Lactation Stage	Bray-Curtis	49	0.2531	0.6171
	Jaccard	49	0.4823	0.4907
	UniFrac	49	1.29	0.2604

Table S7. Comparing mammalian skin, gut, and milk microbiome datasets used in null model analysis. We subset Song et al. (1) dataset to the same species used in the milk microbiome dataset, while Ross et al. (2) was subset to match host order of the milk microbiome dataset. Within the skin dataset (2), we chose samples from the inner thigh with no duplicate samples per individual.

	Skin Microbiome (Ross et al. 2018)	Gut microbiome (Song et al. 2020)	Milk Microbiome (Keady et al. 2023)
Bacterial taxa	8,993	20,696	13,413
No. Samples	106	134	107
No. Host Species	31	32	47
No. Super Orders	4	4	4

Data set S1 (separate file). Structural equation modeling output. Excel file contains model outputs for four measures of microbial structure (microbial richness, Faith's phylogenetic diversity, Bray-Curtis, and unweighted UniFrac). We tested three a *priori* models: a full model with all possible relationships between variables, model 2 lacks a bidirectional relationship between milk sugar GE and milk protein GE, and model 3 assumes no direct relationship between diet and milk microbiome structure.

SI References

1. S. J. Song, *et al.*, Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *mBio* **11** (2020).
2. A. A. Ross, K. M. Müller, J. S. Weese, J. D. Neufeld, Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. *PNAS* **115**, E5786–E5795 (2018).
3. D. B. Jones, Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins. *United States Department of Agriculture* (1931).
4. W. R. Hood, O. T. Oftedal, "Methods of measuring milk composition and yield in small mammals" in *Ecological and Behavioral Methods for the Study of Bats*, (Johns Hopkins University Press, 2009), pp. 539–553.
5. Michel. DuBois, K. A. Gilles, J. K. Hamilton, P. A. Rebers, Fred. Smith, Colorimetric Method for Determination of Sugars and Related Substances. *Anal. Chem.* **28**, 350–356 (1956).
6. J. R. Marier, M. Boulet, Direct Analysis of Lactose in Milk and Serum. *Journal of Dairy Science* **42**, 1390–1391 (1959).
7. D. R. Perrin, The calorific value of milk of different species. *Journal of Dairy Research* **25**, 215–220 (1958).
8. E. S. Wenker, E. A. Himschoot, B. Henry, B. Toddes, M. L. Power, Macronutrient composition of longitudinal milk samples from captive aardvarks (Orycteropus afer). *Zoo Biology* **38**, 405–413 (2019).
9. C. Petzinger, *et al.*, Proximate composition of milk of the bongo (Tragelaphus eurycerus) in comparison to other African bovids and to hand-rearing formulas. *Zoo Biology* **33**, 305–313 (2014).
10. K. Hinde, M. L. Power, O. T. Oftedal, Rhesus Macaque Milk: Magnitude, Sources, and Consequences of Individual Variation Over Lactation. *Am J Phys Anthropol* **138**, 148–157 (2009).
11. J. MacQueen, Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* **5.1**, 281–298 (1967).
12. P. V. A. Fine, S. W. Kembel, Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography* **34**, 552–565 (2011).
13. J. C. Stegen, *et al.*, Quantifying community assembly processes and identifying features that impose them. *The ISME Journal* **7**, 2069–2079 (2013).
14. S. Langenheder, *et al.*, Bacterial metacommunity organization in a highly connected aquatic system. *FEMS Microbiol Ecol* **93** (2017).