PNAS
www.pnas.org

Supplementary Information for

Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia

Ashley A. Ross[1], Kirsten M. Müller[1], J. Scott Weese[2] and Josh D. Neufeld[1]

[1]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada

[2]Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

Correspondence: Josh D. Neufeld

Email: jneufeld@uwaterloo.ca

This PDF file includes:

Supplementary text
Figs. S1 to S5
Tables S1 to S3

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.1801302115

Supplementary Information Text

Control sample analysis. The 78 no-template, DNA extraction kit, and sterile swab controls were analyzed for contaminants after sequence processing (Supplementary Dataset 3). A total of 3 of 4 kits controls, 4 out of 5 sterile swabs, and 67 out of 69 no template PCR controls contained fewer reads than all other samples. The sterile swab and kit control that contained a higher number of sequences were processed with different kits, implying that contamination from an adjacent well may have impacted this kit control (1), instead of an inherent contamination within the DNA extraction reagents (the contaminated kit control was processed in a plate with a clean sterile swab, and vice versa).

The most abundant kit control contaminant was related to the *Neisseriaceae,* at 48.7% abundance in the control sample. This OTU was present in ~27% of samples in this run, the majority of which were cats. Indeed, cat #136 had a very high number of *Neisseriaceae* sequences (~42,000), and was located adjacent to the kit control well. It is therefore hypothesized that this particular kit control's high contamination was from an adjacent well via cross-contamination instead of from a source that would impact all samples, such as kit reagents, implying that there was no significant impact on all samples. Additionally, one of the contaminated no template PCR controls was dominated by an OTU affiliated with *Rhodocytophaga* (36.2%), which had only a single read in one animal sample included in the study. Although several human-associated signatures were in the no-template kit controls, such as *S. epidermidis* and *P. acnes*, they were in negative controls that were surrounded by human samples. The animal samples in these plates did not have elevated levels of human-associated OTUs, and did not group with human samples on ordinations. Indeed, these organisms were found in low abundance on animal samples (Table 2). Removing these OTUs would therefore be inappropriate. To reduce the known impact from well to well cross-contamination (2), all samples were randomized. This ensured that samples from the same animal were distributed across multiple plates and MiSeq runs, as were samples from within a mammalian species or order. Observed influences, including host taxonomy and geography, cannot be due to these groups of samples being situated proximally within the same extraction or PCR plate.

Six "run control" samples consisting of human, zoo, pet, and wild animal samples were included in each of the three runs, confirming the absence of detectable run bias (SI Appendix, Fig. S5). The low diversity observed in human samples (Figure 2) was not due to variations in Illumina run sequencing because the 37 non-human animal samples included in the first lane possessed the same diversity levels as samples from the same species that were sequenced in other lanes.

The following 19 animal swabs were removed in the mammal dataset due to failure to amplify: eight cats, two beavers, and one each of river otter, cape eland, white rhinoceros, cheetah, horse, dog, Indian flying fox, and reindeer samples. These unamplified samples represent 3.6% of total mammalian samples. There was a disproportionate number of cat samples requiring removal, which may be due to several

factors, such as pet owners sampling more lightly on cats resulting in insufficient sample collection. If the swabs were not pressed firmly against the animal's skin, it is possible that only a small number of microorganisms were collected that were below the sequencing detection limit. Alternatively, cats may possess lower overall skin microbial abundances.

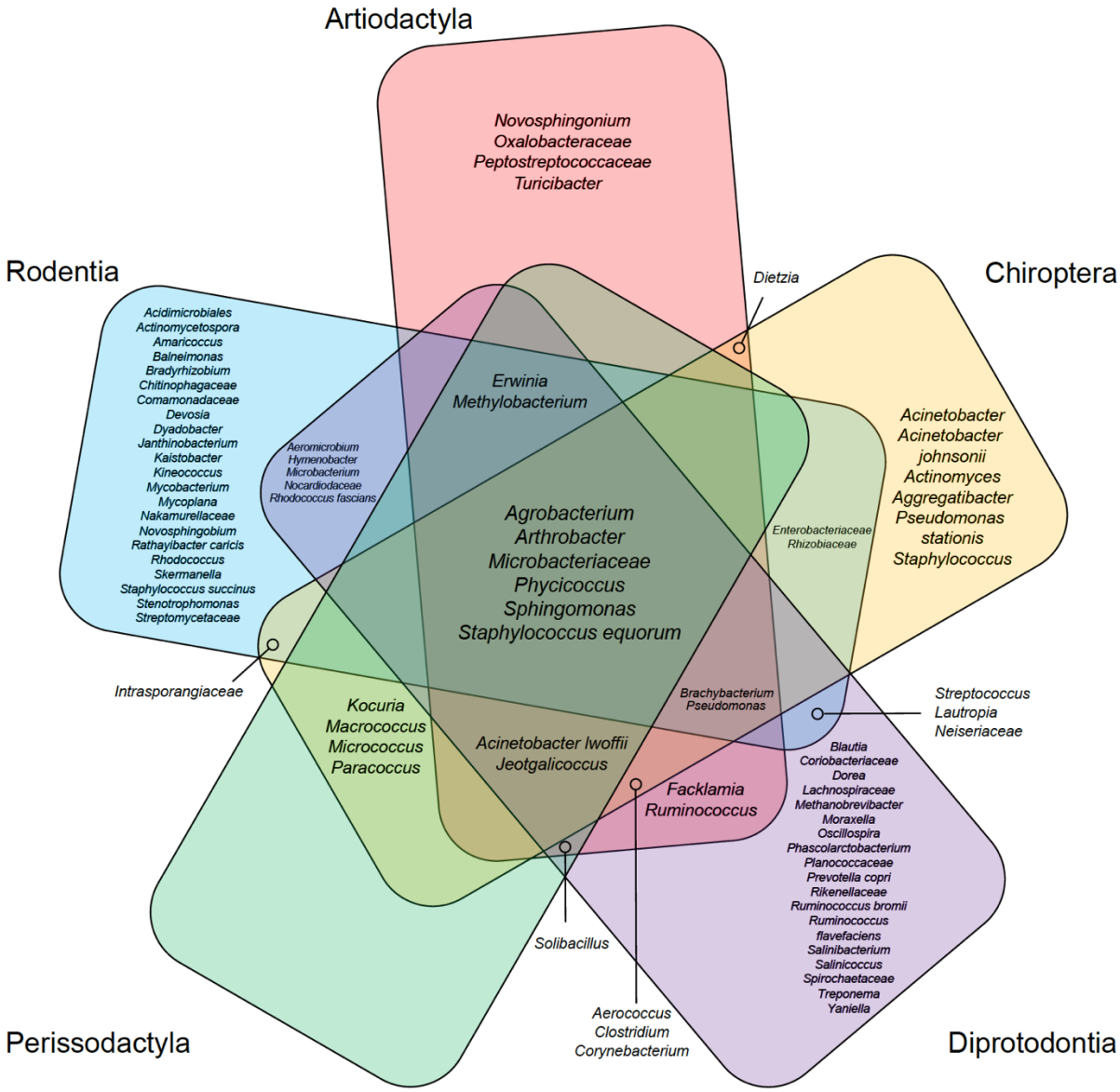

Fig. S1. Venn diagram representing the results of a core operational taxonomic unit (OTU) analysis. Core OTUs were defined as being present in >90% of samples in a designated category. Five mammalian orders were included because they were each associated with samples obtained from multiple species, and these orders did not include "indoor" animals, such as humans, cats, and dogs. The most resolved taxonomic ranking for each OTU was included in the diagram.

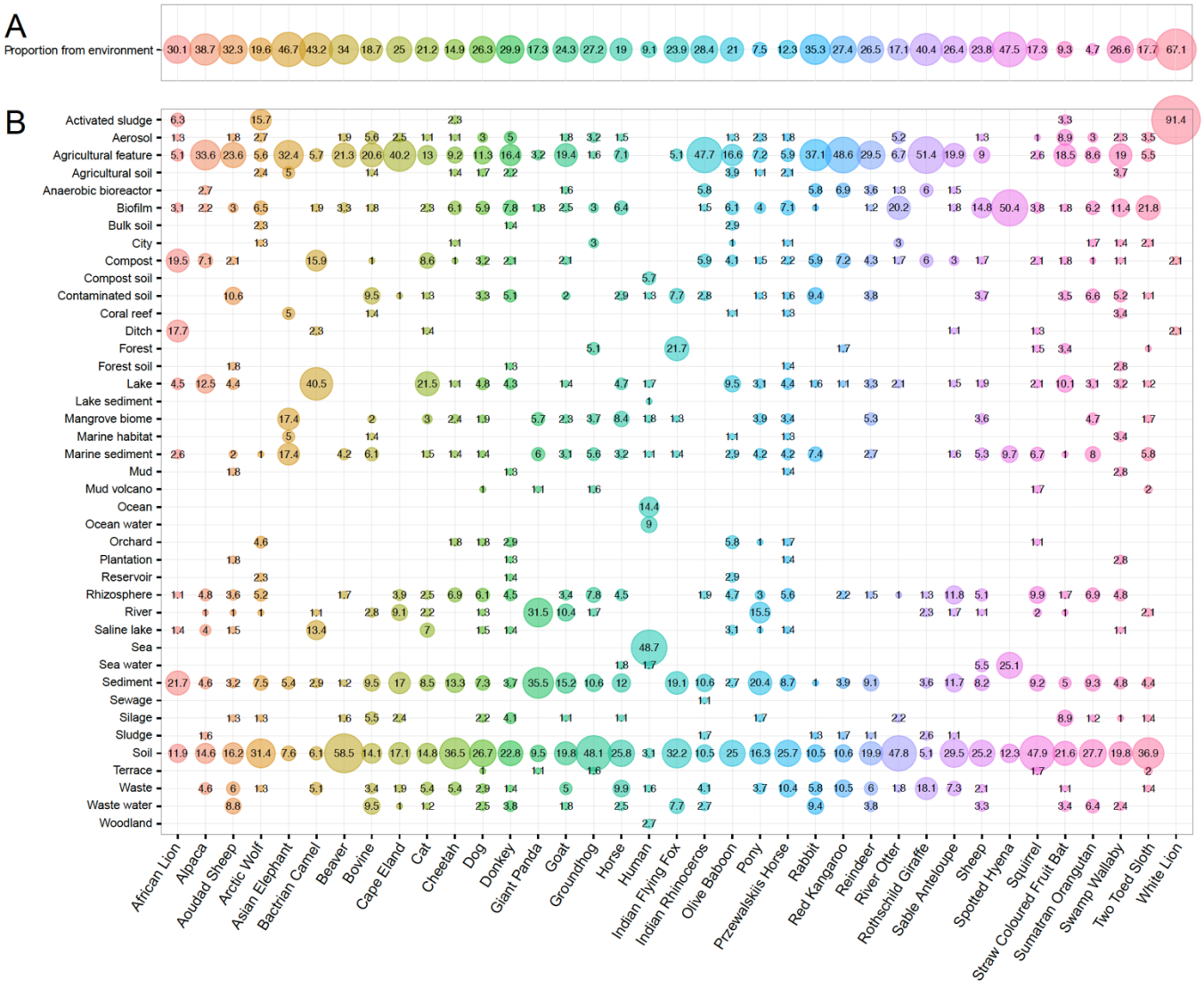

Fig. S2. Bubbleplot representing the proportion of sequences associated with a non-skin habitat for each mammalian species, according to a SeqEnv analysis. A. Proportion of all sequences for each sampled species that were not associated with skin as a habitat ("environment"). B. Detailed distribution sequences for each sampled species across non-skin environmental habitats. Only environments represented by >1% overall relative abundance are shown.

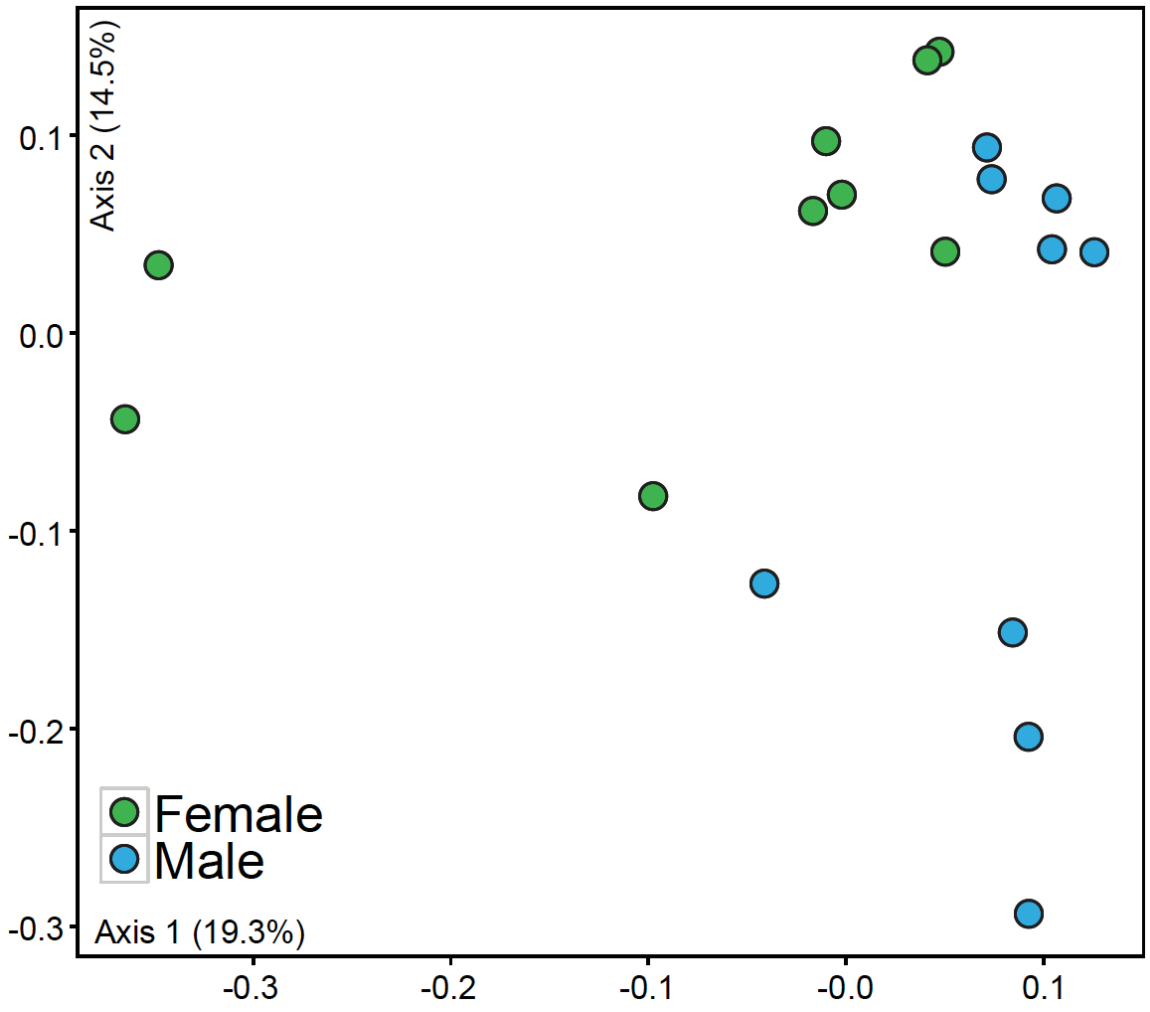

Fig. S3. Principle coordinate analysis (PCoA) ordination generated by using the Bray-Curtis dissimilarity metric for all sampled body locations of red kangaroos. Samples are colored according to biological sex.

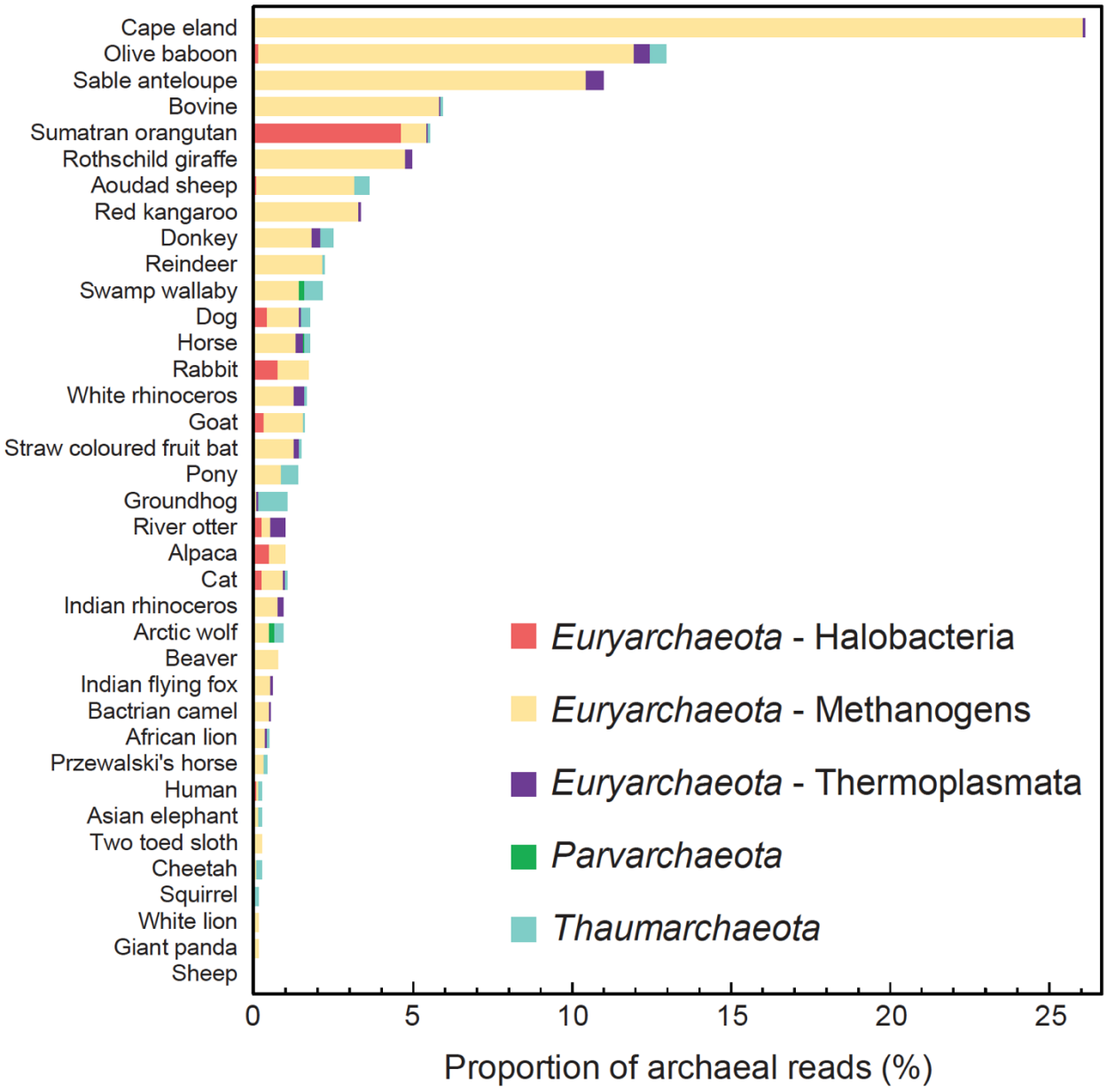

Fig. S4. Summary of archaeal sequence taxonomy. The normalized proportions represent the total proportion of all 6,509 archaeal sequences (together representing ~0.1% of all sequences analyzed for this study). Each mammalian species was normalized to the number of individuals sampled to account for unequal species-specific sampling depths. To calculate normalized proportions, we took the number of archaeal sequences for each species and divided this by the number of animals within the species, providing an average number of sequences per species. The average for each species was divided by the total summed average of all species, then multiplied by 100 to generate the displayed normalized percentages.

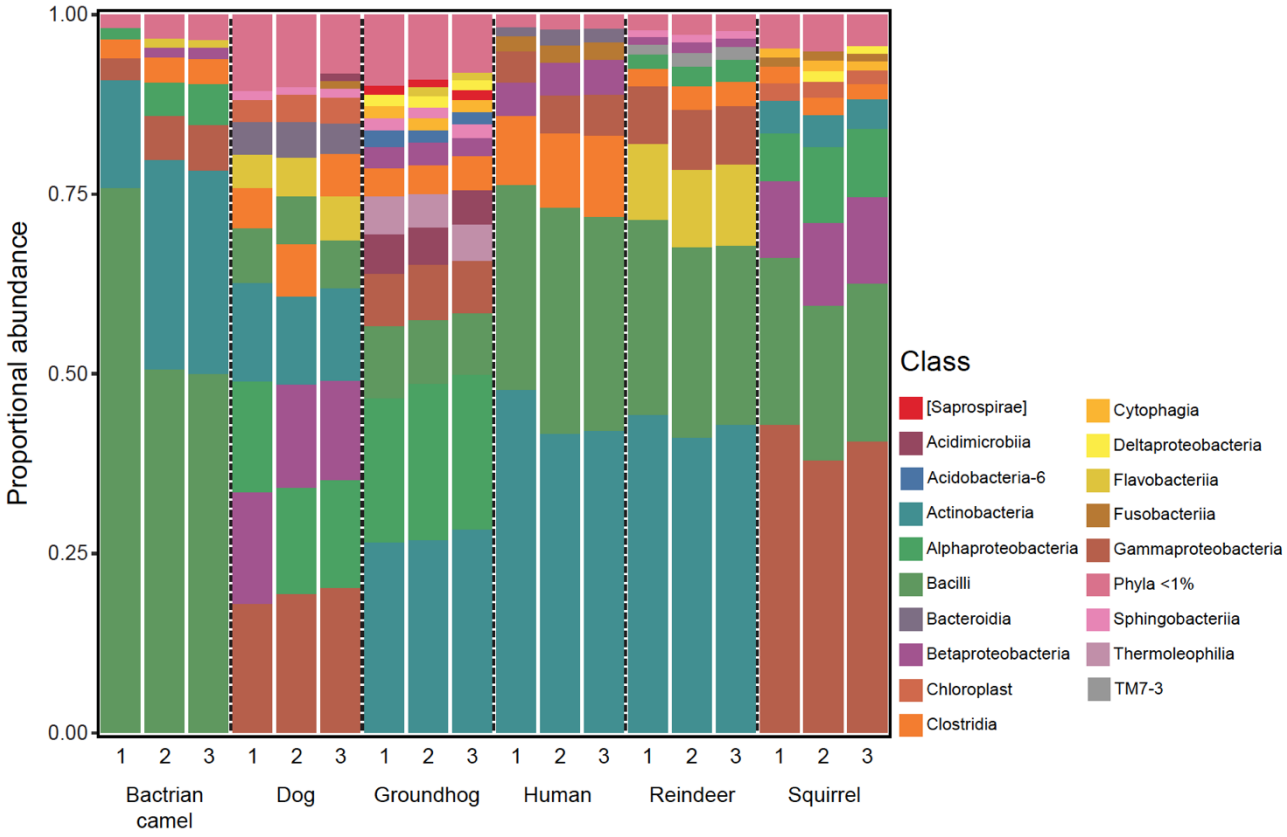

Fig. S5. Taxanomy bar graph of six "run control" samples that were included in each of the three MiSeq runs, ordered from left to right for each species. Taxonomy shown only for operational taxanomic units (OTUs) at >1% relative abundance.

Table S1. Summary of most abundant operational taxonomic units (OTUs) associated with each mammalian species sampled for this study.

Order	Common name	Sampled individuals	Taxonomy of most abundant OTU	Relative abundance of most abundant OTU (%)	Number of OTUs with >1% relative abundance	Number of unique OTUs
Artiodactyla	Alpaca	3	*Macrococcus*	11.4	18	1136
	Aoudad Sheep	9	*Staphylococcus*	5.0	13	2088
	Bactrian Camel	15	*Planomicrobium*	16.5	19	1372
	Bovine	45	*Staphylococcus*	7.4	15	4182
	Cape Eland	11	*Ruminococcaceae*	4.2	12	1840
	Goat	6	*Staphylococcus*	14.2	10	1539
	Reindeer	18	*Alkanindiges*	12.9	18	1295
	Rothschild Giraffe	9	*Corynebacterium*	5.5	20	1395
	Sable Antelope	3	*Oligella*	2.7	17	1086
	Sheep	3	*Corynebacterium*	8.1	14	988
Carnivora	African Lion	9	*Psychrobacter sanguinis*	6.9	17	1481
	Arctic Wolf	9	*Weeksellaceae*	5.9	11	2021
	Cat	48	*Neisseriaceae*	6.7	12	3399
	Cheetah	20	*Enhydrobacter*	11.9	12	2277
	Dog	35	*Macrococcus*	2.4	7	4356
	Giant Panda	6	*Clostridium*	27.6	15	946
	River Otter	2	*Rhodococcus*	9.2	19	417
	Spotted Hyena	3	*Actinobacillus*	7.3	23	436
	White Lion	6	*Psychrobacter*	22.1	14	823
Chiroptera	Indian Flying Fox	18	*Streptococcus*	16.1	21	927
	Straw Coloured-Fruit Bat	9	*Clostridium butyricum*	21.2	14	871
Diprotodontia	Red Kangaroo	18	*Sharpea*	6.0	10	2115
	Swamp Wallaby	3	*Flavobacteriaceae*	7.5	15	958
Lagomorpha	Rabbit	7	*Staphylococcus succinus*	22.1	15	997
Perissodactyla	Donkey	21	*Macrococcus*	6.1	11	5036
	Horse	68	*Corynebacterium*	9.1	8	5645
	Indian Rhinoceros	6	*Actinomycetales*	5.4	25	893
	Pony	3	*Gemellaceae*	15.0	13	916
	Przewalski's Horse	15	*Macrococcus*	34.5	6	2153
	White Rhinoceros	14	*Corynebacterium*	18.0	19	1314
Primates	Human	77	*Propionibacterium acnes*	16.5	15	1628
	Olive Baboon	15	*Lactobacillus*	4.9	14	1890
	Sumatran-Orangutan	9	*Neisseriaceae*	15.4	14	1219
Proboscidea	Asian Elephant	15	*Micrococcus*	8.3	17	1224
Rodentia	Beaver	1	*Moraxellaceae*	7.7	15	319
	Groundhog	6	*Macrococcus*	3.3	12	1955
	Squirrel	21	*Escherichia coli*	5.5	11	2906
Xenarthra	Two Toed Sloth	3	*Kocuria*	7.6	13	922

Table S2. Summary of animal samples with similar microbial communities to humans as determined by ordination analysis.

Animal ID #	Species	Owner also sampled?	Number of samples grouped with humans	Body regions that grouped with human
52	Cat	Yes	3	Back, inner thigh, torso
54	Cat	Yes	2	Back, torso
55	Cat	Yes	2	Back, inner thigh
56	Cat	Yes	2	Inner thigh, torso
57	Dog	Yes	1	Back
58	Cat	Yes	3	Back, inner thigh, torso
69	Dog	No	1	Back
70	Cat	No	3	Back, inner thigh, torso

Table S3. TestPrime 1.0 comparison of Pro341F/Pro805R primer set to the SILVA database, identifying overall primer set mismatches to representative sequences of all archaea, thaumarchaeota specifically, and bacteria.

Mismatches to either primer	Cumulative proportion of archaea (%)	Cumulative proportion of thaumarchaeota (%)	Cumulative proportion of bacteria (%)
0	64.8	11.9	85.7
1	89.0	93.2	94.6
2	94.9	95.5	96.1

Additional data table S1 (separate file)

Table of operational taxonomic units (OTUs) for all samples rarefied (tab 1), all samples with humans removed (tab 2), and non-rarefied OTU counts for archaea-affiliated sequences (tab 3). Sample codes can be linked to metadata by consulting with the metadata file (Additional data table S2). Summary count data, consensus lineage information, and representative sequences are all included in the rightmost columns.

Additional data table S2 (separate file)

Metadata table linking sample identifiers (#SampleID) with information related to sequencing and sampling.

Additional data table S3 (separate file)

Non-rarefied table of operational taxonomic units (OTUs) for DNA extraction kit controls ("Kit"), no-template controls ("NTC"), and sterile swab controls ("SS"). Summary count data, consensus lineage information, and representative sequences are all included in the rightmost columns.

References

1. Novakova E, et al. (2017) Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile virus. *Front Microbiol* 8(April):1–17.
2. Galan M, et al. (2016) 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife: the importance of cleaning post-sequencing data before estimating positivity, prevalence and co-infection. *mSystems* 1(4):e00032-16.