Supporting Online Material for

Evolution of Mammals and Their Gut Microbes

Ruth E. Ley, Micah Hamady, Catherine Lozupone, Peter Turnbaugh, Rob Roy Ramey, J. Stephen Bircher, Michael L. Schlegel, Tammy A. Tucker, Mark D. Schrenzel, Rob Knight, Jeffrey I. Gordon*

*To whom correspondence should be addressed. E-mail: jgordon@wustl.edu

Published 22 May 2008 on *Science* Express
DOI: 10.1126/science.1155725

This PDF file includes:

Supporting Online Materials

Methods

Sample acquisition - Fecal samples were collected from captive and wild animals in the early morning and immediately frozen in N_2 or dry ice, shipped or returned to the laboratory on dry ice, and preserved at -80ºC. For conspecific animals housed in groups, 2-3 samples were obtained from separate parts of their shared nighttime living quarters, to ensure that they were produced by different individuals. All captive animals were born in captivity, with the exception of one Eastern Black and White Colobus Monkey, and one Chimpanzee who were born in the wild and housed at the St Louis Zoo.

Samples from wild animals were collected in Namibia (Hartmann's Mountain Zebra, Hamadryas Baboon, Springbok, African Elephant) by observing animals drawn to a water hole. Fresh fecal specimens were also collected from Argali Sheep in Mongolia and Bighorn Sheep in Colorado, USA, by tracking and observation. Wild animal feces were oven-dried for a minimum of 30 min at 70ºC immediately after collection, and maintained in this state until they were processed for DNA extraction.

Stable isotope analyses and diet data - Samples were homogenized and pulverized in liquid N_2 with a mortar and pestle, and freeze-dried. The percentages of C, N and their stable isotope ratios ($^{13}C/^{12}C$, $^{15}N/^{14}N$) were determined by mass spectroscopy at the Stable Isotope Ratio Facility for Environmental Research (Department of Biology, University of Utah). Detailed diet records, including food composition data, were obtained for captive animals from the Saint Louis Zoological Society and the Zoological Society of San Diego. A diet fiber index (FI) was created

from the percentages in each diet of acid-detergent fiber (ADF) and neutral-detergent fiber (NDF): FI = (%ADF +1) * (%NDF +1).

DNA preparation - An aliquot (~100mg) of each fecal sample was suspended while frozen (or added dry for oven-dried samples) in a solution containing 500 μl of DNA extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210 μl of 20% SDS, 500 μl of a mixture of of phenol:chloroform:isoamyl alcohol (25:24:1)], and 500 μl of a slurry of 0.1-mm-diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK). Microbial cells were then lysed by mechanical disruption with a bead beater (BioSpec Products) set on high for 2 min (23^{o}C), followed by extraction with phenol:chloroform:isoamyl alcohol, and precipitation with isopropanol.

Five replicate PCRs were performed for each host DNA sample. Each 25 μl reaction contained 50-100 ng of purified DNA, 10 mM Tris (pH 8.3), 50 mM KCl, 2 mM $MgSO_4$, 0.16 μM dNTPs, 0.4 μM of the bacteria-specific primer 8F (5'-AGAGTTTGATCCTGGCTCAG-3'), 0.4 μM of the universal primer 1391R (5'-GACGGGCGGTGWGTRCA-3'), 1 M betaine, and 3 units of Taq polymerase (Invitrogen). Cycling conditions were 94°C for 2 min, followed by 35 cycles of 94°C for 1 min, 55°C for 45 sec, and 72°C for 2 min, with a final extension period of 20 min at 72°C [35 cycles were used for consistency across all samples because initial trials with 20 cycles failed to yield product for a number of samples]. Replicate PCRs were pooled, concentrated with Millipore columns (Montage), gel-purified with the Qiaquick kit (Qiagen), cloned into TOPO TA pCR4.0 (Invitrogen), and transformed into *E. coli* TOP10 (Invitrogen). For each sample, at least 384 colonies containing cloned amplicons were processed for sequencing. Plasmid inserts were sequenced bi-directionally using

vector-specific primers and the internal 16S rRNA gene primer 907R (5'-CCGTCAATTCCTTTRAGTTT-3').

Analysis pipeline used for the mammalian gut 16S rRNA sequence dataset (illustrated in **Figure S5**).

Sequence assembly and chimera checking – 16S rRNA gene sequences were edited and assembled into consensus sequences using PHRED and PHRAP aided by XplorSeq (*1*), and bases with a PHRAP quality score of < 20 were trimmed. A multiple sequence alignment was generated with the NAST online tool (*2*), and chimeras identified with Bellerophon version 3 (*3*), implemented at the Greengenes website (http://greengenes.lbl.gov), with the following (default) parameters: sequences were compared to others within the same host species and to the Greengenes Core Set; similarity to the core set was set to 97%; the match length to sequence threshold was set to 1250 bp; the window size was set to 300; the count of similar sequences to search for each window was 7; hypervariable regions were masked using the LaneMaskPH (http://greengenes.lbl.gov); the parent to fragment ratio was 90%; and the divergence ratio threshold was set to 1.1. Using these stringent criteria, 17,760 sequences were retained from an input dataset of 26,072 sequences. The 8,312 putative chimeric sequences are available at our website at http://gordonlab.wustl.edu/PublicationPDFs/mammal_chimeras.zip.

UniFrac clustering – Sequences remaining after chimera-checking were added to a neighbor joining (NJ) tree available with the Greengenes core set database in Arb (http://greengenes.lbl.gov/Download/Sequence_Data/Arb_databases/greengenes.arb.gz, downloaded Dec 12 2006) using parsimony insertion. Sequences that were not part of

our study were removed, and the resulting tree used to cluster communities using the online UniFrac tool (unweighted algorithm; http://bmf.colorado.edu/unifrac; (*4, 5*): i.e., a matrix of community pair-wise distances generated by UniFrac was used to cluster samples by (i) the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method (**Figure S2**) and (ii) principal coordinates analysis (PCoA) (**Figure 2**).

OTU picking algorithm - Sequence identity was calculated using megablast (*6*), with the following parameters: E-value $1x10^{-10}$; minimum coverage, 99%; word size, 42; and minimum pairwise identity, 96%. Candidate OTUs were identified as a graph, or network, of sequences where each sequence was connected to at least one other sequence having ≥96% sequence identity. The candidate OTU was considered valid if the average density of connection was above 90%: i.e., if 90% of the possible pairwise sequences in the set had a percent identity above the threshold. If the density was lower than this threshold, we then iteratively identified subgraphs in the candidate OTU in which the density was above the threshold. A representative sequence was chosen from each valid OTU by selecting the sequence with the largest number of connections to other sequences in the OTU.

Taxonomy assignments - Taxonomy was assigned using the best megablast (*7*) hit (above threshold) against Greengenes (*2*) (E-value cutoff of $1x10^{-15}$; minimum of 90% identity over the length of the shorter sequence; word size 42) and the RDP taxonomy annotation (http://greengenes.lbl.gov/Download/Sequence_Data/Greengenes_format/greengenes16SrRNAgenes.txt.gz, downloaded Dec 03 2007 and flagged chimeras omitted).

Phylogenetic diversity (PD) measurements - To determine which mammals had the most diverse communities of gut bacteria, Phylogenetic Diversity (PD) measurements, as described by Faith (*8*), were made for each animal. PD is the total amount of branch length in the phylogenetic tree of all 16S rRNA gene sequences from all mammals studied that leads to the sequences that were found in fecal samples of one specific mammal. To account for differences in sampling effort between animals, and to estimate how far we were from sampling the diversity of each mammal completely, we plotted the accumulation of PD (branch length) with sampling effort, in a manner analogous to species rarefaction curves. We generated the PD rarefaction curve for each mammal by applying custom python code (http://bayes.colorado.edu/unifrac) to the Arb parsimony insertion tree. For each animal, we first removed all sequences that were not from that animal from the global Arb parsimony insertion tree, and calculated the total remaining branch length (defined as the total PD for that animal). We then sequentially removed five sequences, chosen at random, and recorded the branch length at each step, until there were fewer than five sequences remaining in the tree. The plotted values (**Figure S3**) are averages over 25 replicate trials.

Testing for basal carnivorous lineages - Because ancestral mammals were carnivores, we tested whether the ancestral environment of the bacterial lineages was in a carnivorous host, and whether this state subsequently switched to herbivorous or omnivorous gut environments. To do so, we determined whether predicted switches from a carnivorous host to an herbivorous or omnivorous host were deeper in the phylogenetic tree than the other four possible types of switches. In order to perform this analysis, we first used the Fitch parsimony algorithm (*9*) to infer ancestral states for the internal nodes.

In this algorithm, the tree is traversed from the tips to the root. If the intersection is not empty (if the same state or set of states is in all of the 'child' nodes), the ancestral node is assigned to the intersection. If the intersection is empty (if there are no states shared by all of the children), the ancestor is assigned to the union of all of the children's environments. For instance, if a node has three children, and one is from a carnivore, one is from an herbivore, and one is from an omnivore, the intersection will be the empty set, and the state of the node will be set to all three diets. Since this technique results in many of the internal nodes having ambiguous assignments, the switches that occurred cannot be directly inferred.

To determine the direction of the environment switches, we started at the root of the tree and moved towards the tips. If the state of the root node was ambiguous, we picked a state at random. Each ambiguous node in the tree was assigned to its parent state (since we traversed from the root to the tips, the parent never would have had an ambiguous state). In this way, for each node in the tree, we assigned an unambiguous state and thus inferred branches in a tree where a particular type of switch occurred. Because there are many equally parsimonious solutions to internal state assignments, particularly if the root node is ambiguous, we repeated the internal state assignments for 50 replicate trials. For each of the six possible types of switches between the three diets, we identified the nodes at which this switch was predicted to occur, and calculated the distance of this node from the root of the tree. We averaged these values for all of the nodes at which switches were predicted to have occurred for the 50 replicate trials of internal state assignments, and compared them to determine whether the switches from a

carnivorous state were on average closer to the root. The average distance of carnivore to omnivore or herbivore switches from the root were not lower than the other switch types.

Network-based analysis – Each host-bacterial network was constructed as a bipartite graph, in which each node represented either a host sample or a bacterial OTU. Connections were drawn between samples and OTUs, with edge weights defined as the number of sequences from each OTU that occurred in each sample. Networks were visualized using Cytoscape 2.5.2 (*10*). The dataset for this analysis consisted of sequences with a minimum length of 400 bp, 1% maximum ambiguous characters (n=21,533). To test if mammal nodes were more connected to other mammal nodes in the same diet group, or in the same taxonomic order, than expected by chance, a G test for independence was applied. Each sample pair was classified according to whether its members shared at least one OTU, and whether they shared a category. Pairs were then tested for independence in these categories (this had the effect of asking whether pairs that shared a diet category were also equally likely to share an OTU). This procedure provides a parametric estimate of the p-value for the association, although factors such as sampling can also affect this estimate.

Co-evolution between mammals and their gut bacterial communities - We performed UniFrac, recursively, on the entire bacterial tree in a procedure that had the effect of asking, at each node, whether the bacterial lineages stemming from that node mirrored the mammalian phylogeny. For the de-replicated, chimera-screened tree containing 18,237 sequences [17,760 generated in this study plus 38 from the wild African Gorilla (see **Table S1**) and 439 from human samples (lean controls from (*11*))], we began at the root, and performed UniFrac at each node in a post-order traversal

through the tree, using a file mapping each bacterial sequence back to the mammalian species from which it came. We calculated a UPGMA tree from the resulting UniFrac distance matrix at each step, and compared this tree to a reference tree of the mammals (*12*) using the method of overlapping subsets. This yielded a distance from 0 to 1 for each subtree, with 1 indicating that none of the monophyletic groups were the same in the two trees, and 0 indicating that all of them were the same. This method was more robust than simply asking whether subtrees of the bacterial tree had the same topology as the mammalian tree because it allowed us to account for losses of bacterial taxa from specific hosts, multiple speciation within a host (including ancestral hosts), and repeated patterns in multiple clades that began with different ancestors but underwent the same pattern of cospeciation. These analyses were implemented using PyCogent (*13*).

Results

We performed several controls to determine the impact of chimeras, sequence length, number of sequences per sample, number of individual samples per mammalian species, and the effects of percent identity threshold selected for defining OTUs.

Gauging the impact of chimeras

Chimeric 16S rDNA sequences are concatenations that occur when two or more gene segments, each derived from different genes, recombine during the polymerase chain reaction. There is no universally accepted method to detect chimeras in 16S rRNA gene sequence datasets, although a number of software tools exist for chimera checking. We selected Bellerophon v.3 for several reasons: (i) it is the only application currently available that allows large (>1000) sequence libraries to be screened; (ii) it allows

searching for 'parent' sequences of putative chimeras against the Greengenes 'coreset' database as well as the rest of the library so that, in principle, parent sequences that were not present in the final sequence set can be used in the analysis; and (iii) it is used to screen sequences from GenBank prior to deposition in the Greengenes database (http://greengens.lbl.gov) which is reportedly free of chimeras (*14*).

As described in Supporting Online Methods, we used the parameter settings implemented by the authors of the Greengenes database (*14*): a divergence ratio of 1.1, a fragment to parent level of similarity of >90%, and a 300 bp window. 8,400 sequences were flagged as chimeric.

Chimera Test 1: addition of artificially generated chimeras to a manually curated dataset does not affect clustering of samples based on PCoA of UniFrac distances and by network analysis. This test employed the manually curated dataset of Relman and co-workers (*15*), consisting of 11,627 bacterial 16S rRNA sequences derived from six colonic mucosal sites, and one fecal sample from each of three unrelated, healthy adults (total of seven sequence libraries/individual; individuals labeled 70, 71 and 72 in **Figure S6**). This dataset has been reported to have ~2% chimeras by other researchers (*16*). We increased the size of the dataset by 100% by adding chimeras generated *in silico* from the dataset. The process of generating these artificial chimeras involved taking two different sequences, selected at random from the same library and searching them for a shared 20 nucleotide word that would be used as an *in silico* recombination breakpoint, provided that the recombined segments were at least 100 bp long. In each instance, the first half of the first sequence was combined with the second half of the second sequence. All simulated chimeric sequences were then combined with

the original dataset. The original (n=11,627) and chimera-augmented datasets (n=23,254) were then analyzed in two ways: (1) alignment with NAST followed by UniFrac analysis and PCoA of the UniFrac distance matrix, and (2) directly using OTU-based network analysis (**Figure S5**).

Samples in this dataset cluster by individual when the original published set of sequences (n=11,627) was used (panel A in **Figure S6A**). The clustering is almost identical when the dataset is increased by addition of 11,627 chimeras (panel B in **Figure S6B**). Similarly, in the network analysis, clustering of the samples is by individual for the published and chimera-augmented datasets (panels C and D in **Figure S6**). Thus, even a very high proportion of chimeras in the dataset does not affect the conclusions: samples cluster by individual.

Chimera test 2: addition of Bellerophon-flagged or artificially generated chimeras to the mammalian dataset does not affect clustering of samples in the UniFrac PCoA analysis or bacterial-mammalian network. In our study of mammalian microbiotas, we had excluded all 8,400 sequences flagged as chimeric by Bellerophon v.3. In this test, we added back these putative chimeras and similar to the process employed for Chimera Test 1, ran the UniFrac analysis and PCoA (chimeras aligned with NAST), as well as network-based analysis. The results revealed that the clustering was unchanged by addition of Bellerophon-flagged chimeras or by introducing the artificially generated chimeras (**Figure S7**).

Chimera test 3: removing 8,400 randomly selected sequences from the full dataset, containing Bellerophon v3 flagged chimeras, does not affect clustering in the network analysis. See **Figure S8**.

Chimera test 4: **artificially generated chimeras added unevenly to each sample does not affect clustering.** We tested the effect of an uneven distribution of chimeras across samples on the clustering in the network. The number of artificially generated chimeras added to a given sample (either a +0%, +25%, +50%, +75% or +100% increase above the number of sequences in the Bellerophon v3-processed dataset) was randomly determined. **Figure S9** shows that clustering in the network analysis was not affected by this treatment.

Other comments - In the network diagrams, unique OTUs are collapsed into diamond-shaped nodes that are scaled accordingly: i.e., the larger the diamonds, the more unique OTUs were estimated for the associated host sample (**Figure 1A**). A comparison of diamond-shaped nodes in the network diagrams for the chimera-augmented versus the chimera-screened mammalian datasets generated for tests 2-4 revealed that, as expected, the chimera-augmented datasets contain a greater number of unique OTUs. Nonetheless, the clustering pattern remained unchanged (data not shown) (This was also true for the dataset Eckburg et al (15) (see panels C and D in **Figure S6)**.

Figure S10 shows five identical runs of the Bellerophon screened mammalian gut dataset in which animal nodes are colored by diet. The use of networks to display microbial ecology datasets was developed for this study. The spring-embedded algorithm used for network figures is stochastic; multiple runs in Cytoscape of the same dataset using the same parameters will yield different layouts each time, but the network statistics and clustering trends are unaffected.

Evaluating the impact of 16S rRNA sequence length

To test the effect of sequence length on network clustering, we performed the analysis using three different length cut-offs: only including sequences ≥400 bp, or ≥800 bp or ≥1000 bp (see **Figures 1B, S10**, **Figure S11 (**panels A and B**)**). The results revealed no effect on clustering (the sequence length cut-off used in the analysis described in the main text was ≥400 bp).

Gauging the effect of the number of sequences used per fecal sample

We jackknifed the UniFrac tree analysis shown in **Figure S2** to assess how often cluster nodes were recovered when smaller, even sets of sequences were sampled from the host samples. We performed 10 jackknife analyses, one for each of 10 specified numbers of sequences (n=22, 50, 75, 100, 125, 150, 175, 200, 225, and 250). For each jackknife analysis, we did 100 permutations in which we randomly sampled the specified number of sequences from each host sample and re-clustered the data. Starting with the smallest sequence set (n=22), we recorded, with symbols on the tree, which nodes were recovered in >50% of permutations. The results show support for many important nodes in the tree with a sequence number below 200. For instance, most of Herbivore Group 2 is supported with a minimum of 50 sequences, as are nodes in Herbivore Group 1 and the human cluster; Carnivore Group 1 is supported with a minimum of 75 sequences; many other nodes are supported with 125 sequences per sample or fewer.

Furthermore, we jackknifed the PCoA analysis for the chimera-screened mammalian dataset by randomly selecting 21 and 100 sequences from each mammal sample and in each case, performing UniFrac analysis followed by PCoA 100 times. Samples with fewer sequences than the jackknife value were removed prior to the analysis. The results show that repeated random sampling of 21 or 100 sequences from

each sample does not change the overall pattern of clustering by diet (panels A and B in **Figure S12**)

We also performed the network analysis four times with a different specified number of randomly selected sequences per sample each time: 20 (all 106 samples have >20 sequences); 50 (96 have >50 sequences); or 100 (83 have >100 sequences. The major clusters are recovered in each instance (panels C-E in **Figure S12):** e.g., even with 20 sequences per sample, the Carnivora cluster together, as do the Artiodactyla, the Primates, and the Perissodactyla.

Assessing the effect of the number of individual samples available per host species

We addressed this issue by only selecting one sample per species (the sample with the largest sequence count). There was no effect on clustering by taxonomic order in the network analysis (**Figure S13**).

Effect of percent identity threshold used for OTUs

In the analysis described in the main text, a threshold of 96% identity was used to delimit genus-level OTUs. To gauge the effect of percent identity on the analysis, we re-ran the network analysis on the mammalian dataset (without chimeras) with OTUs defined at 95%, 97% and 98% identity. Clustering by taxonomy was unaffected (**Figure S14**).

Figures

Figure S1 – The percentage of sequences from each fecal sample assigned to different phyla. (A) 100% of sequences. (B) An enlargement of the upper portion of panel A highlighting rarer phyla. Hosts are clustered by taxomomic order. Animal names are colored according to diet: green, herbivore; blue, omnivore; red, carnivore. See **Table S1** for additional details.

Figure S2 - Unweighted pair group method with arithmetic mean (UPGMA) clustering of bacterial communities for each host based on pair-wise differences determined using the UniFrac metric. UniFrac is based on the premise that related communities share an evolutionary history that can be estimated as the fraction of shared branch length in a common phylogenetic tree (the 21,619 16S rRNA gene sequence neighbor-joining tree). Labels are colored according to diet (carnivores, red; herbivores, green; omnivores, blue). Vertical bars located to the left of animal names indicate co-clustering of conspecific hosts. Non-clustering conspecifics are indicated with same-color stars. Details concerning the human samples are provided in parentheses and include sample ID from Table S1, descriptors used in the original studies plus PubMed ID for each study where available (*e.g.*, T0 and T4 refer to the initial and one-year time point samples for lean control subjects 13 and 14 in PubMed ID 17183309). Additional information about the samples can be found in **Table S1**. The circles and squares at internal nodes in the tree indicate jackknife support of ≥50% for 100 iterations; the key at

the upper right corner of the figure shows the minimum number of sequences retained per sample for each jackknife analysis.

Figure S3 - Phylogenetic diversity (PD) and OTU rarefaction curves. (**A**) PD rarefaction. The total amount of branch length added to a phylogenetic tree with sequences from each sample is plotted. (**B**) OTU rarefaction. For each sample, the number of OTUs (≥96% identity) per sequence is tallied.

Figure S4 - Testing for co-diversification between mammals and their fecal bacterial communities. The y-axis shows the distance from 0 (most similar) to 1 (most dissimilar) between the UniFrac tree (performed recursively at each node in the bacterial tree) and the mammalian phylogeny, compared using the method of overlapping subsets. The x-axis shows the number of mammalian samples involved in the comparison. The blue points are the real data, red points are for a randomized mammalian phylogeny, and purple refers to overlapping data points. The two distributions are significantly different (paired t-test $P=1.8 \times 10^{-11}$, t=-6.73, n=12,787; note that some bacterial clades were excluded because they were found in ≤2 mammalian samples making the clustering technique inapplicable). Although there is substantial overlap between the red and the blue distributions, the blue distribution contains most of the smallest distances, indicating that co-diversification has led to more concordance between the trees than would be expected by chance. The two distributions converge at larger numbers of taxa, suggesting that co-diversification is seen primarily in specific clades of mammals rather than in the tree overall.

Figure S5 - Analysis pipeline. Boxes refer to software tools, red ellipses to steps from which results were outputted, and arrows and red letters to analysis pathways. Assembled sequences ("original dataset" or "full dataset") were aligned using NAST, then Bellerophon v.3 was applied to screen for chimeras (Greengenes on-line tools, Box1). The chimera-screened dataset was processed two ways (A or B, red arrows). (A): The aligned sequences were added to an existing NJ tree in Arb with hypervariable regions masked (Box 2); the NJ tree is exported for the UniFrac analysis and PCoA (A_1, Box 3: online UniFrac), and to the co-evolution test (A_2, Box 4: custom python software) and PD rarefaction (A_3, Box 5: custom python software). (B): The alignment was removed and OTUs were chosen from the unaligned sequences (Box 6: see **Methods** in Supporting Online Materials). After assigning sequences to OTUs, several paths were taken. These were (B_1), Network analysis, and (B_2), Rarefaction analysis. Analysis pathways labeled with an "A" are not affected by OTU estimates, but are based on the NAST alignment and are potentially affected by any misalignments. Conversely, analysis pathways labeled with a "B" are not affected by misalignment.

Figure S6 - Chimera test 1: addition of artificially generated chimeras to a manually curated dataset does not affect clustering of samples based on PCoA of UniFrac distances. PCoA plots of UniFrac distance matrix for the dataset published in ref. 15 without added (**A**), and with added artificial chimeras (**B**). PC1 and PC2 are plotted and the percent variation explained is indicated on the axes. Network diagrams of datasets

with (**C**) and without (**D**) artificial chimeras. Each colored circle in the PCoA plots or circular node in the network diagrams represents a community associated with a colonic mucosal biopsy or a fecal sample. Sample nodes are colored according to the individual from which they were obtained (70, red; 72 green; 71, blue); small grey rounded-square nodes in the network diagrams are OTUs (see legend for **Figure 1A** and the main text for a full explanation of network components).

Figure S7 - Chimera test 2: addition of Bellerophon-flagged or artificially generated chimeras to the mammalian dataset does not affect network clustering. UniFrac PCoA plots of the mammalian bacterial dataset with and without 8,400 Bellerophon v3 flagged chimeric sequences (**A** and **B**, respectively). Network diagrams with and without Bellerophon-flagged chimeras (**C** and **D**, respectively) and with artificially produced chimeras (**E**; each sample was doubled in size with the addition of chimeras from that sample). In all cases, clustering is qualitatively similar. Animal nodes are colored by phylogenetic order. For an explanation of network symbols and taxon labels, see the legend for **Figure 1A** in the main text.

Figure S8 - Chimera test 3: removal of 8,400 randomly selected sequences from the full dataset (not screened for chimeric sequences) does not affect network clustering. The starting mammalian dataset contained the 8,400 chimeras identified by Bellerophon v3. 8,400 sequences were then randomly removed from the "full dataset" and a network analysis performed.

Figure S9 - Chimera test 4: an uneven proportion of chimeras artificially added to each sample does not affect network clustering. (A, B) Network analysis for the mammalian dataset augmented with artificially generated chimeras: the number of artificial chimeras added to each sample was uneven. Host nodes are colored by taxonomic order in panel **A** and by the proportion of added artificial chimeras in panel **B** (light blue: +0%, dark blue: +25%, magenta: +50%, red: +75% or yellow: +100%). Samples cluster by order and not based on their proportion of artificial chimeras.

Figure S10 - Network diagrams generated from repeated identical runs. (**A**-**E**) Replicate runs of the same mammalian chimera-screened dataset. Stochastic differences are evident.

Figure S11 – Assessing the effects of 16S rRNA sequence length on the network analysis. The minimum length of 16S rRNA sequences required for inclusion in the network analysis was (**A**) 800 bp, and (**B**) 1000 bp. Clustering by taxonomic order is not altered: compare to **Figure S10** where the minimum sequence length was 400 bp for all runs.

Figure S12 - Effect of sequence count in the UniFrac PCoA and network analysis. (**A**,**B**) Jackknifed UniFrac PCoA plots for the chimera-screened mammalian dataset. Panel A, 21 sequences randomly chosen for each mammal (100 jackknife repetitions performed). Panel B, 100 sequences chosen at random for each mammal (100 jackknife repetitions). PC1 and PC2 are plotted on the x- and y-axes, respectively, together with the

percentage of variation explained by the plotted principal coordinate. The points plotted are the average position for each mammal sample in the 100 replicate runs; ellipses around the points are interquartile ranges. (**C-E**) Network diagrams for the chimera-screened dataset with a randomly selected even number of sequences per sample. Panel C, 20 sequences/sample (all 106 fecal samples had >20 sequences). Panel D, 50 sequences/sample (96 samples had >50 sequences). Panel E, 100 sequences/sample (83 samples had >100 sequences). Clustering by taxonomic order is evident even when the number of sequences/sample is only 20.

Figure S13 - Effect of reducing the dataset to one sample per mammal species. Network diagram for the chimera-screened mammalian dataset with one sample per species. Clustering by taxonomic order is still evident: e.g., the single human specimen (HumEckA) clusters with the other primates.

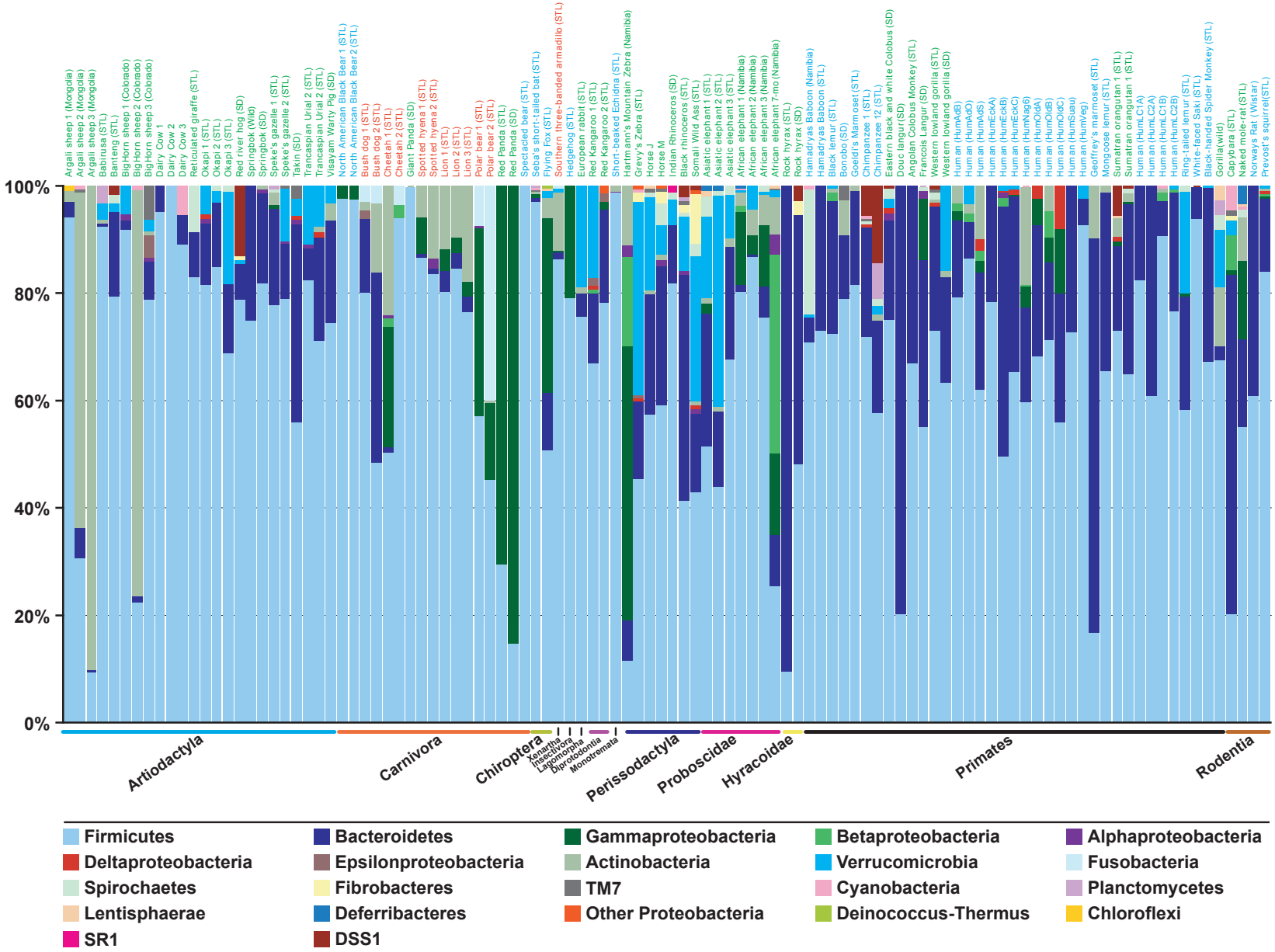
Figure S14 - Effect of varying threshold cutoff for OTU percent identity. Network diagrams for the Bellerophon-screened dataset. Furthest-neighbor-like OTUs include sequences with pairwise percent identities of (**A**) ≥ 95%; (**B**) ≥97%; (**C**): ≥98%.

Table S1 – Mammals used in the study: samples and metadata. Sample IDs, which are also the prefix used for each sequence obtained for any given sample and deposited in GenBank, are listed, as well as the labels used for each sample in the network diagrams. Multiple sample IDs refer to multiple individuals per species. Column designations are largely self-explanatory: 'Total OTUs 96% ID', the number of OTUs at 96% 16S rRNA gene sequence identity obtained per sample; 'Number of OTUs Unique to Sample', number of OTUs only found in that sample; '%Unique OTUs', the percentage of unique OTUs for each sample; Provenance, where the animals were living at time of sampling [SD= San Diego Zoo and San Diego Zoo's Wild Animal Park, ST= St Louis Zoo, W=wild or domesticated (2 horses and humans)]; Diet, H=herbivore, C=carnivore, O=omnivore; Gut physiology, FG=foregut fermenter, HG=hindgut fermenter, S=simple gut; Stable isotope measurements of feces, $d^{13}C$, $d^{15}N$; carbon and nitrogen content of feces (%C, %N); ADF, percentage of acid-detergent fiber in the diet (a measure of hemicellulose); NDF, percentage of neutral detergent fiber (a measure of cell wall content); Fiber Index, an index of total fiber calculated from ADF and NDF; Fiber Index Category, categories based on the fiber index ranges obtained. Dietary ADF and NDF were provided by the St Louis and San Diego zoos for captive animals. Animals for which sequence information was generated in this study are listed in the upper table. Information for previously published sequence data obtained from GenBank is listed in the lower table, with PubMed IDs or Author/Year of publication. N/A, data not available.

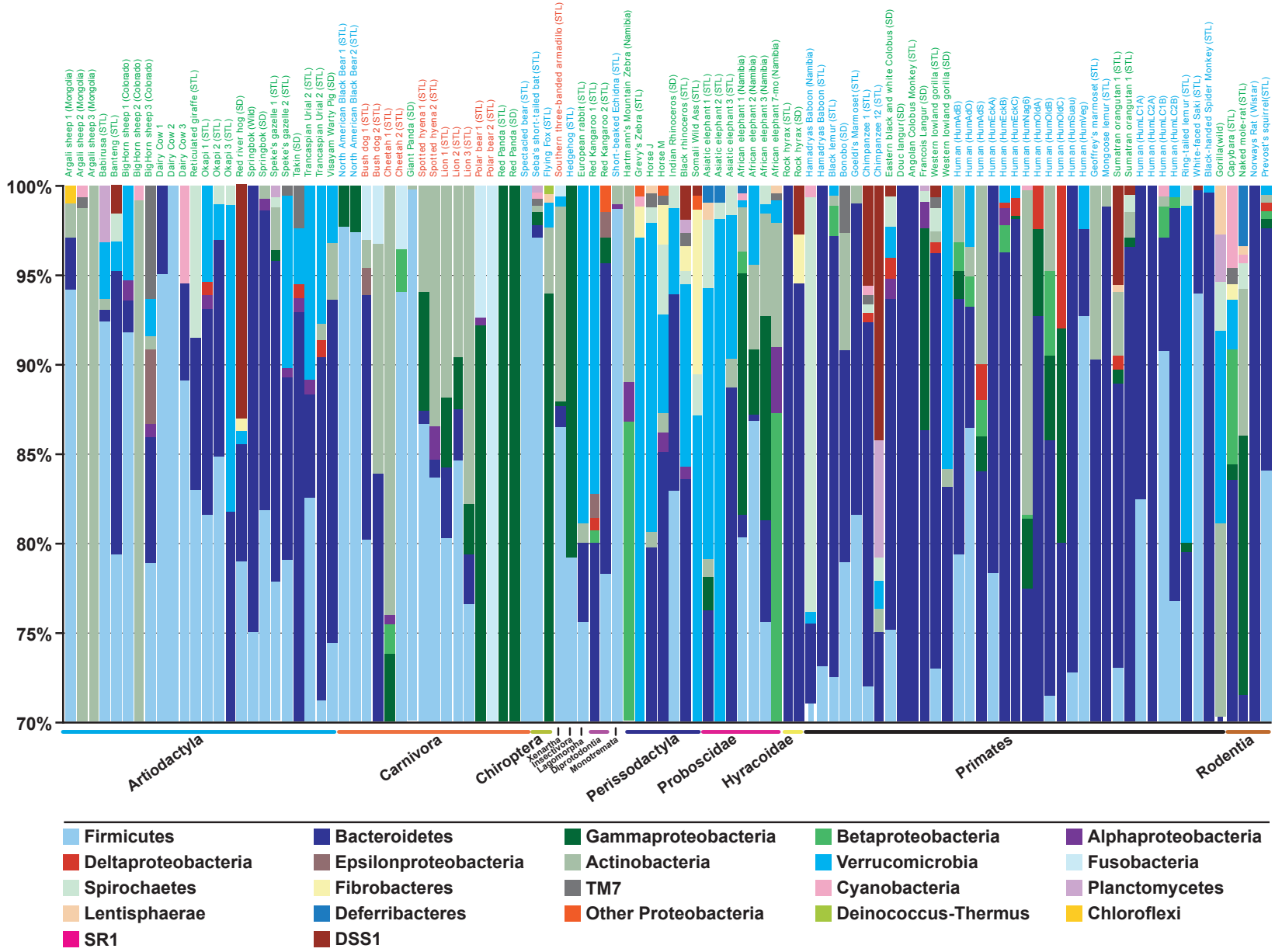
References

1. D. N. Frank *et al.*, *Proc Natl Acad Sci U S A* **104**, 13780 (2007).
2. T. Z. DeSantis, Jr. *et al.*, *Nucleic Acids Res* **34**, W394 (2006).
3. T. Huber, G. Faulkner, P. Hugenholtz, *Bioinformatics* **20**, 2317 (2004).
4. C. Lozupone, M. Hamady, R. Knight, *BMC Bioinformatics* **7**, 371 (2006).
5. C. Lozupone, R. Knight, *Appl Environ Microbiol* **71**, 8228 (2005).
6. S. McGinnis, T. L. Madden, *Nucleic Acids Res* **32**, W20 (2004).
7. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
8. D. P. Faith, *Biological Conservation* **61**, 1 (1992).
9. W. M. Fitch, *Systematic Zoology* **20**, 406 (1971).
10. P. Shannon *et al.*, *Genome Res* **13**, 2498 (2003).
11. R. E. Ley, P. J. Turnbaugh, S. Klein, J. I. Gordon, *Nature* **444**, 1022 (2006).
12. U. Arnason *et al.*, *Proc Natl Acad Sci U S A* **99**, 8151 (2002).
13. R. Knight *et al.*, *Genome Biol* **8**, R171 (2007).
14. T. Z. DeSantis *et al.*, *Appl Environ Microbiol* **72**, 5069 (2006).
15. P. B. Eckburg *et al.*, *Science* **308**, 1635 (2005).
16. K. E. Ashelford, N. A. Chuzhanova, J. C. Fry, A. J. Jones, A. J. Weightman, *Appl Environ Microb* **72**, 5734 (2006).

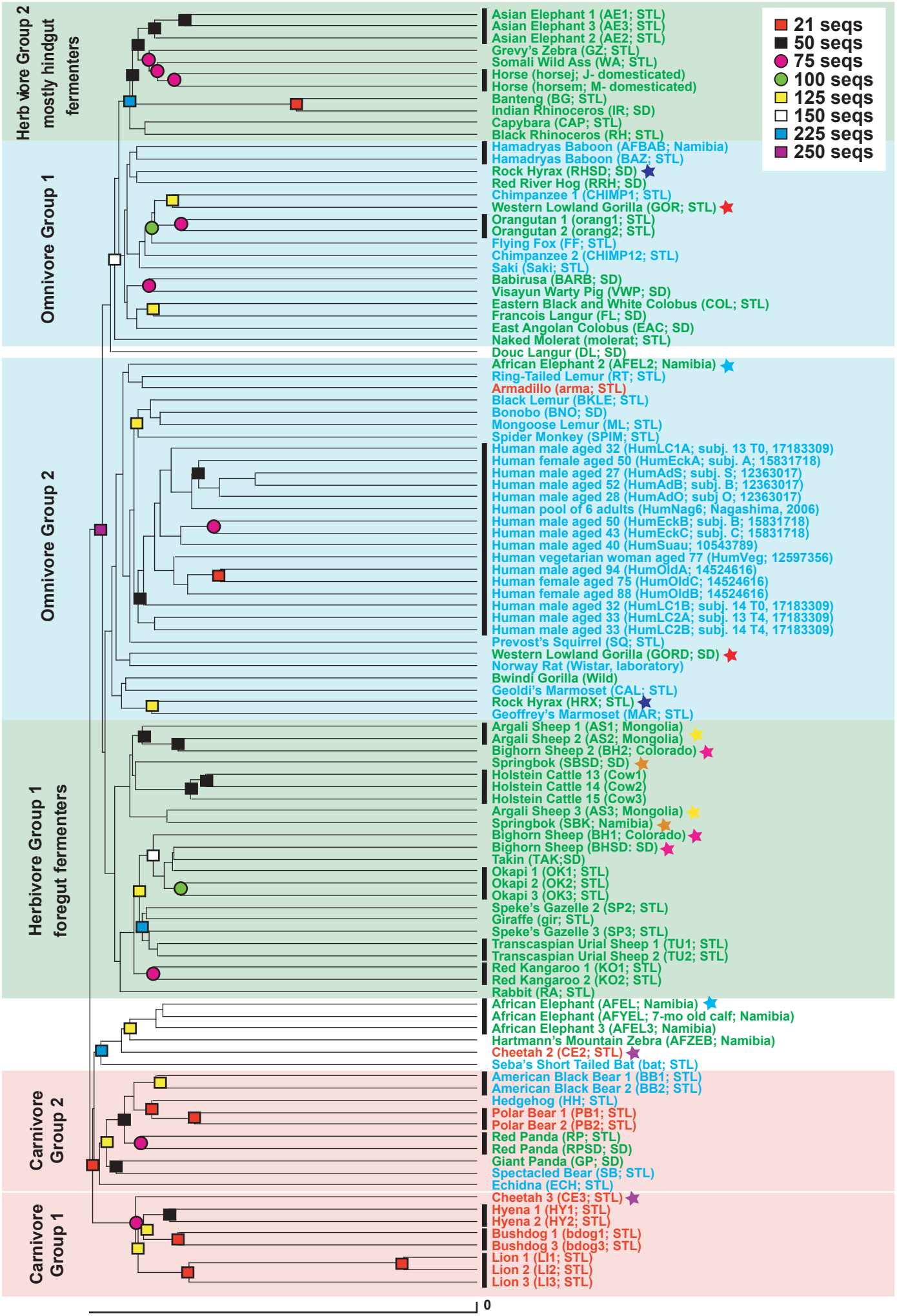
Ley *et al*, Figure S1A



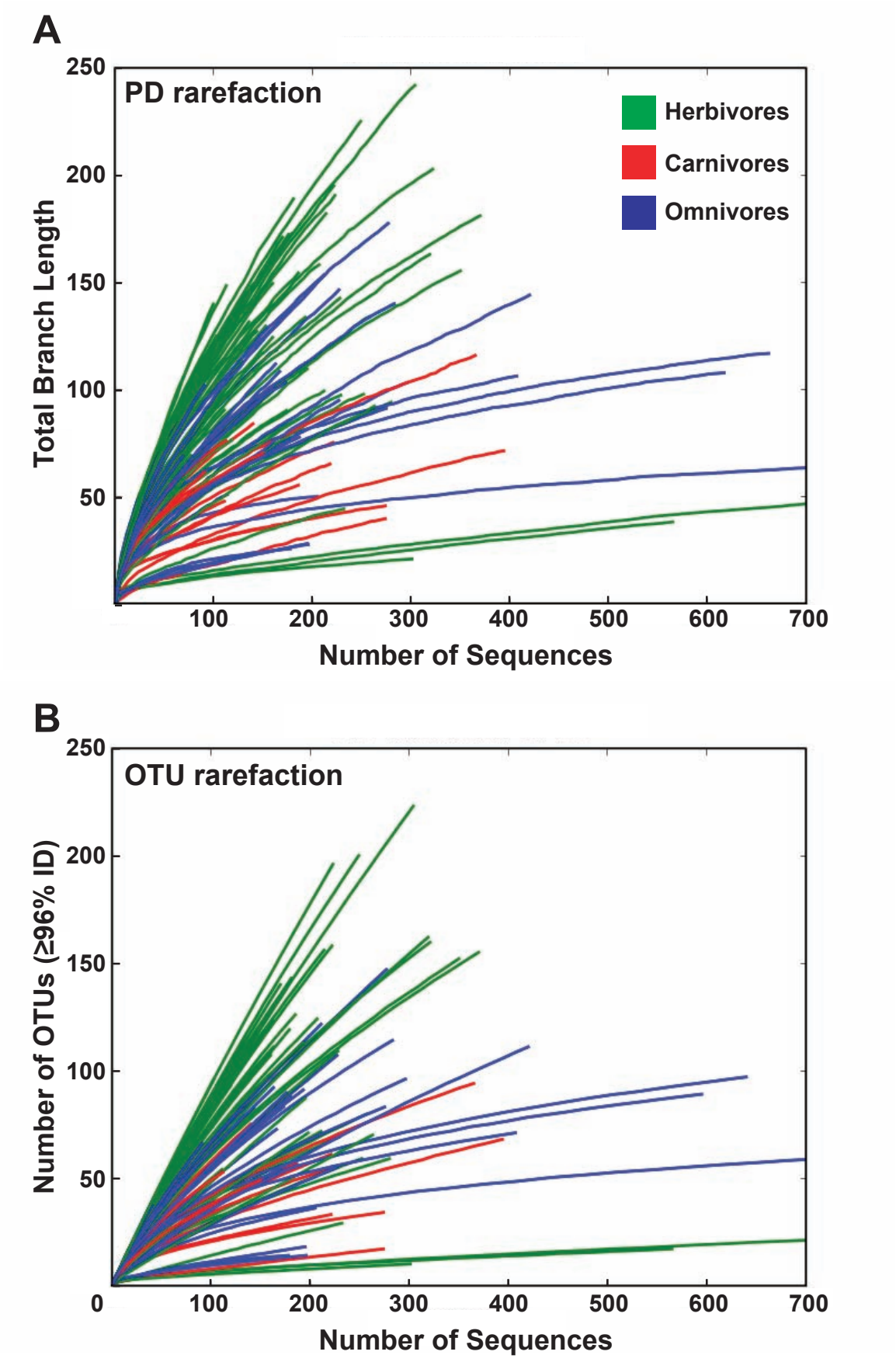
Ley *et al*, Figure S1B



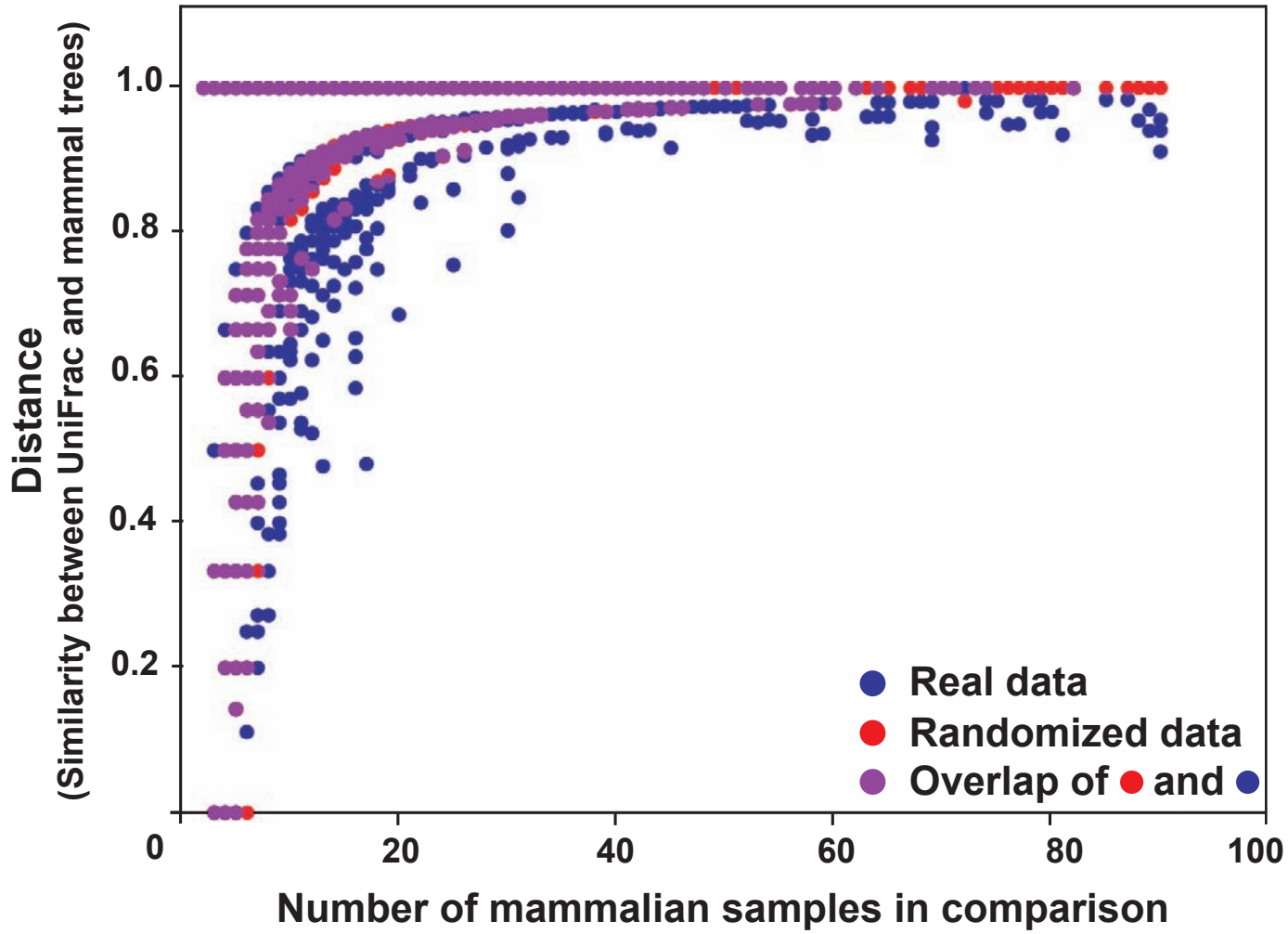
Ley *et al*, Figure S2



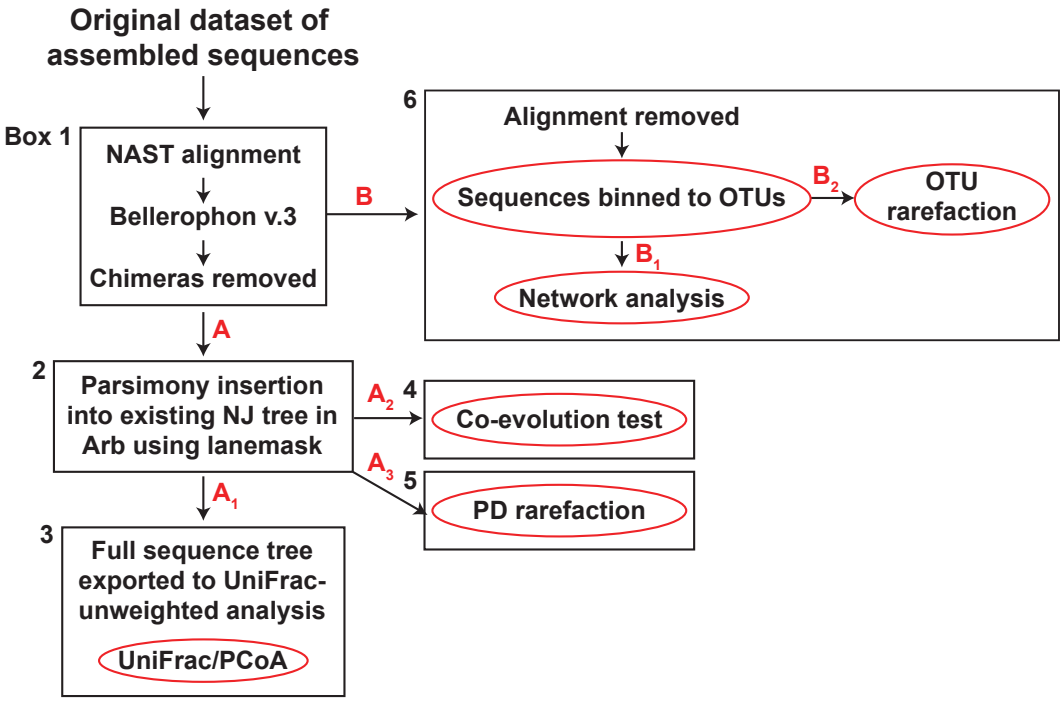
Ley *et al*, Figure S3



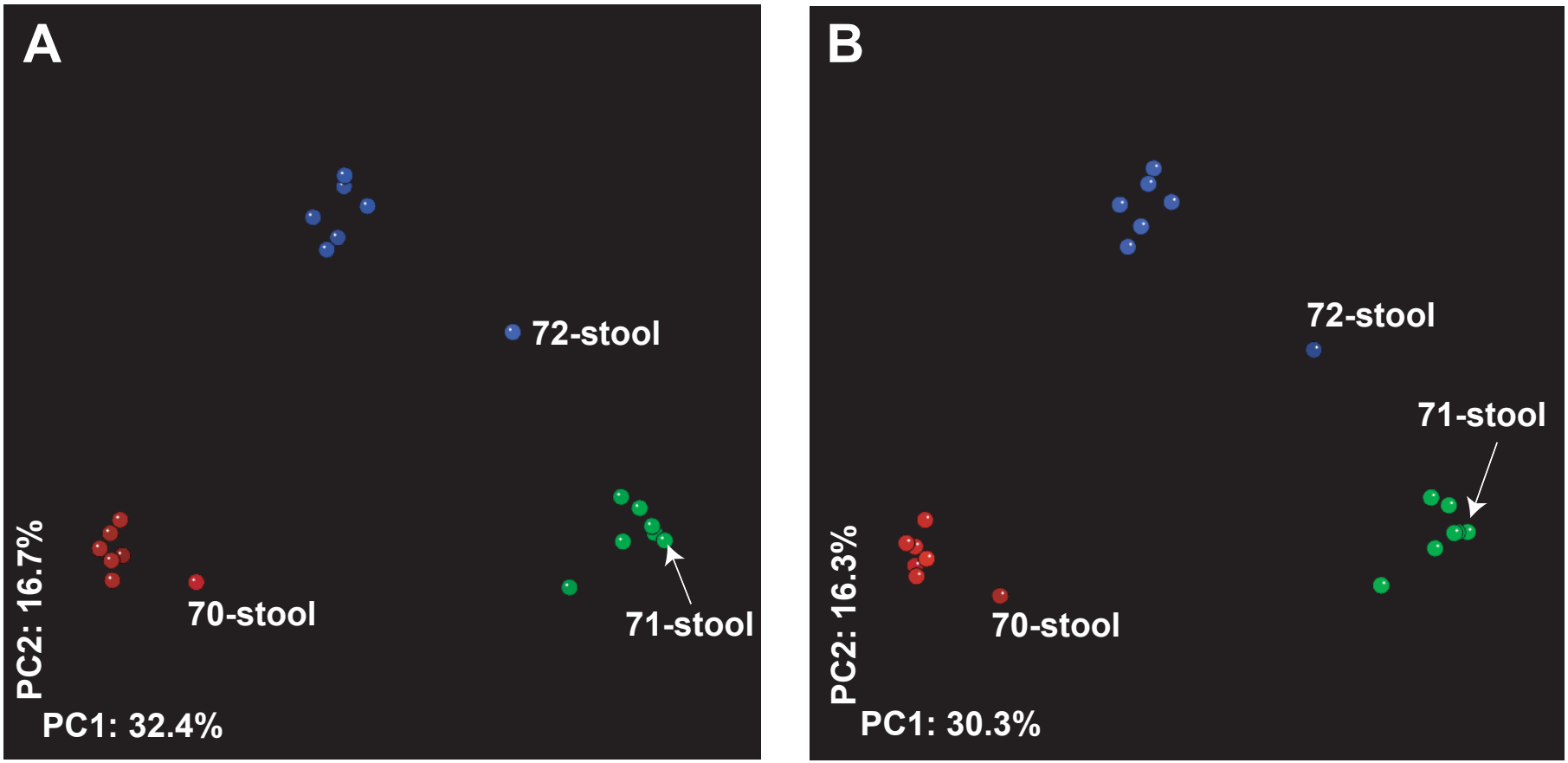
Ley *et al*, Figure S4



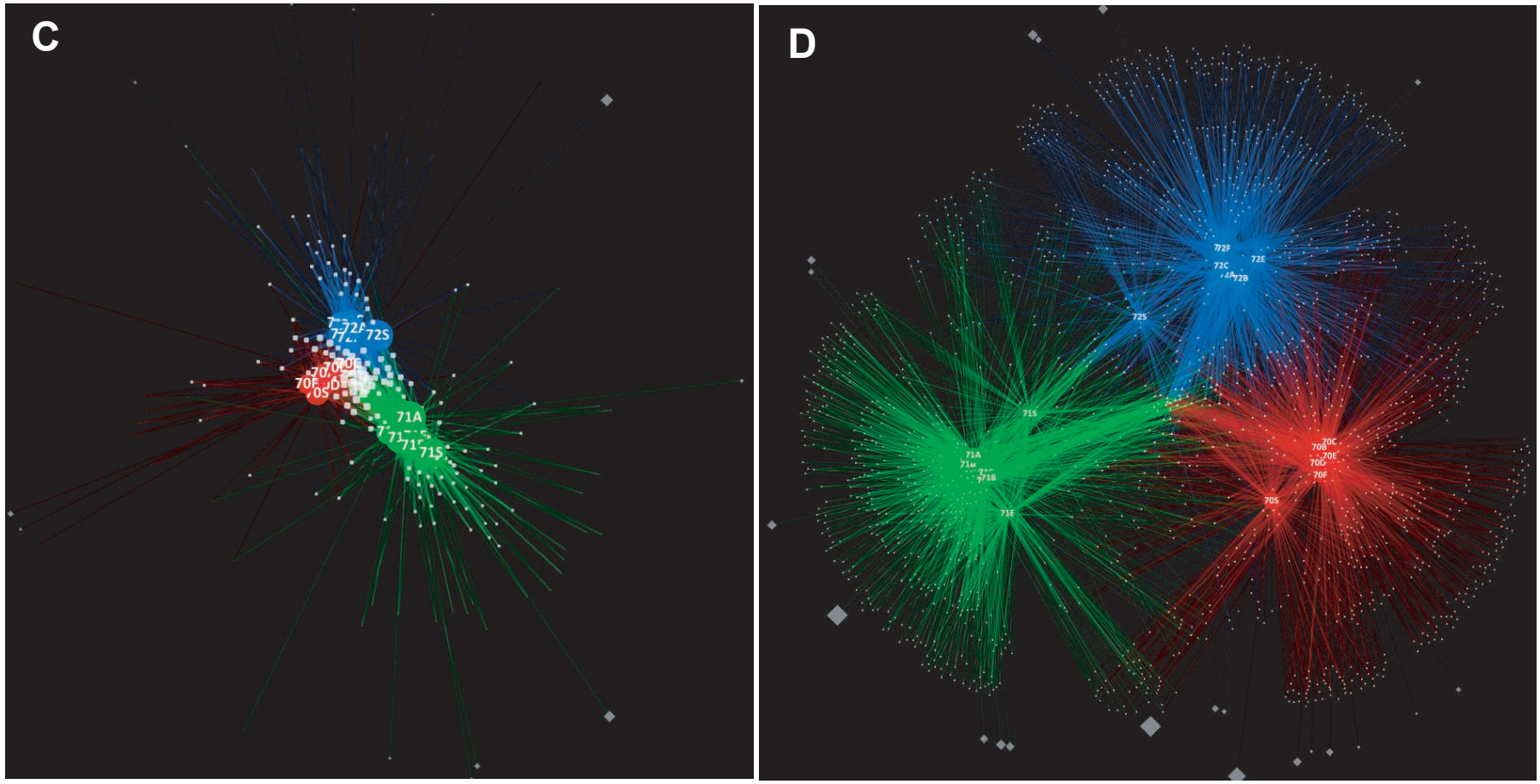
Ley *et al.*, Figure S5



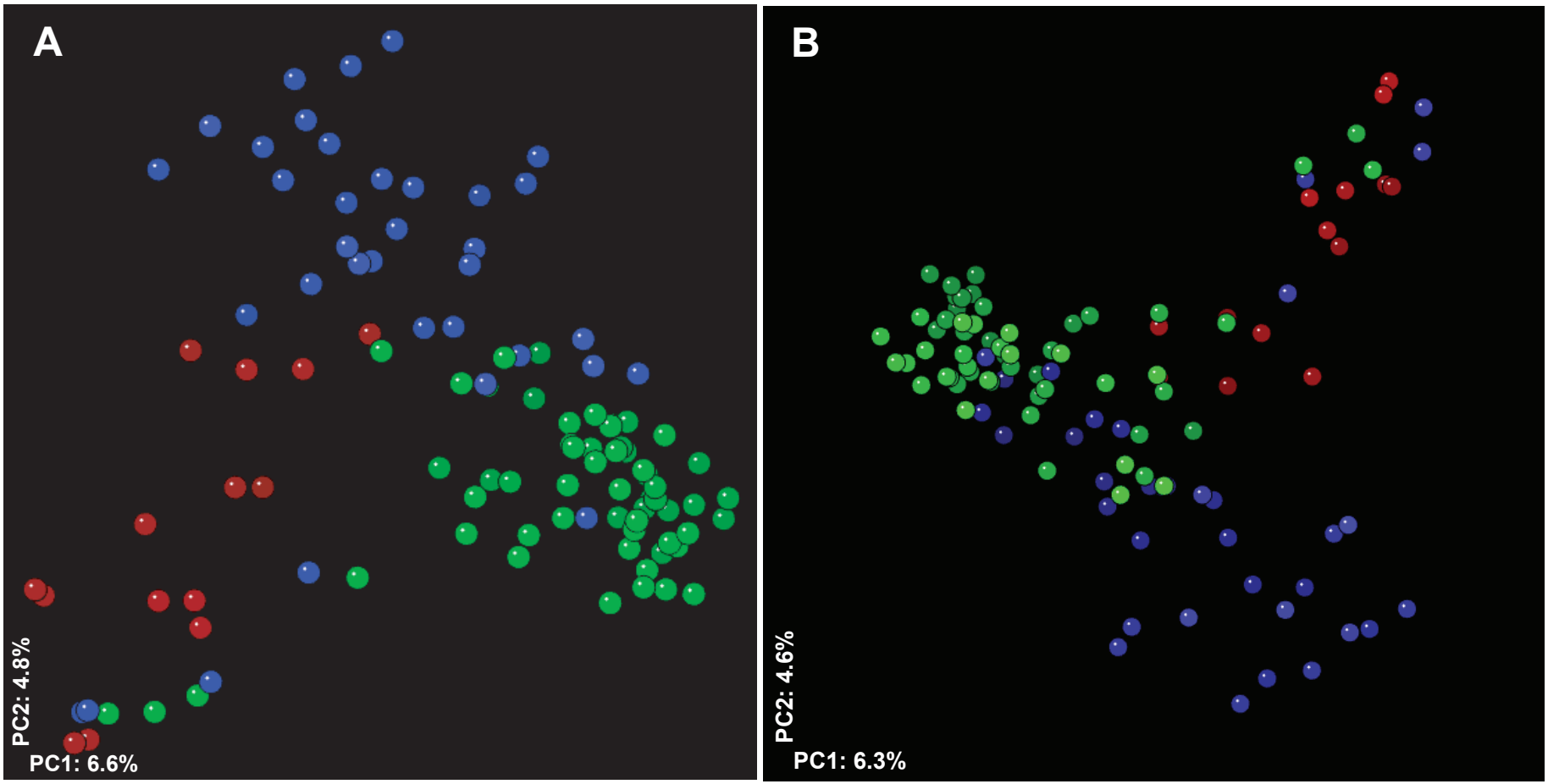
Ley *et al.*, Figure S6



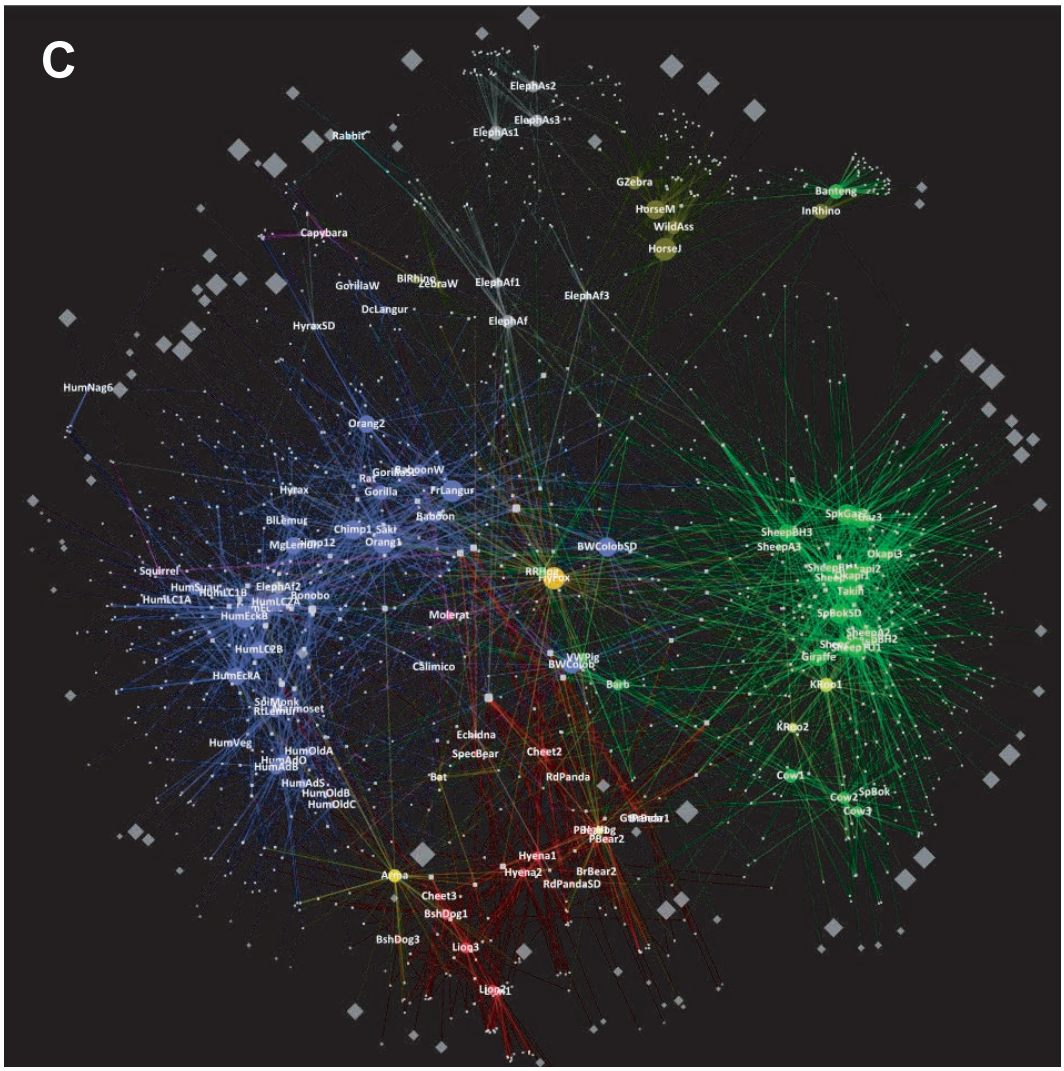
Ley *et al.*, Figure S6



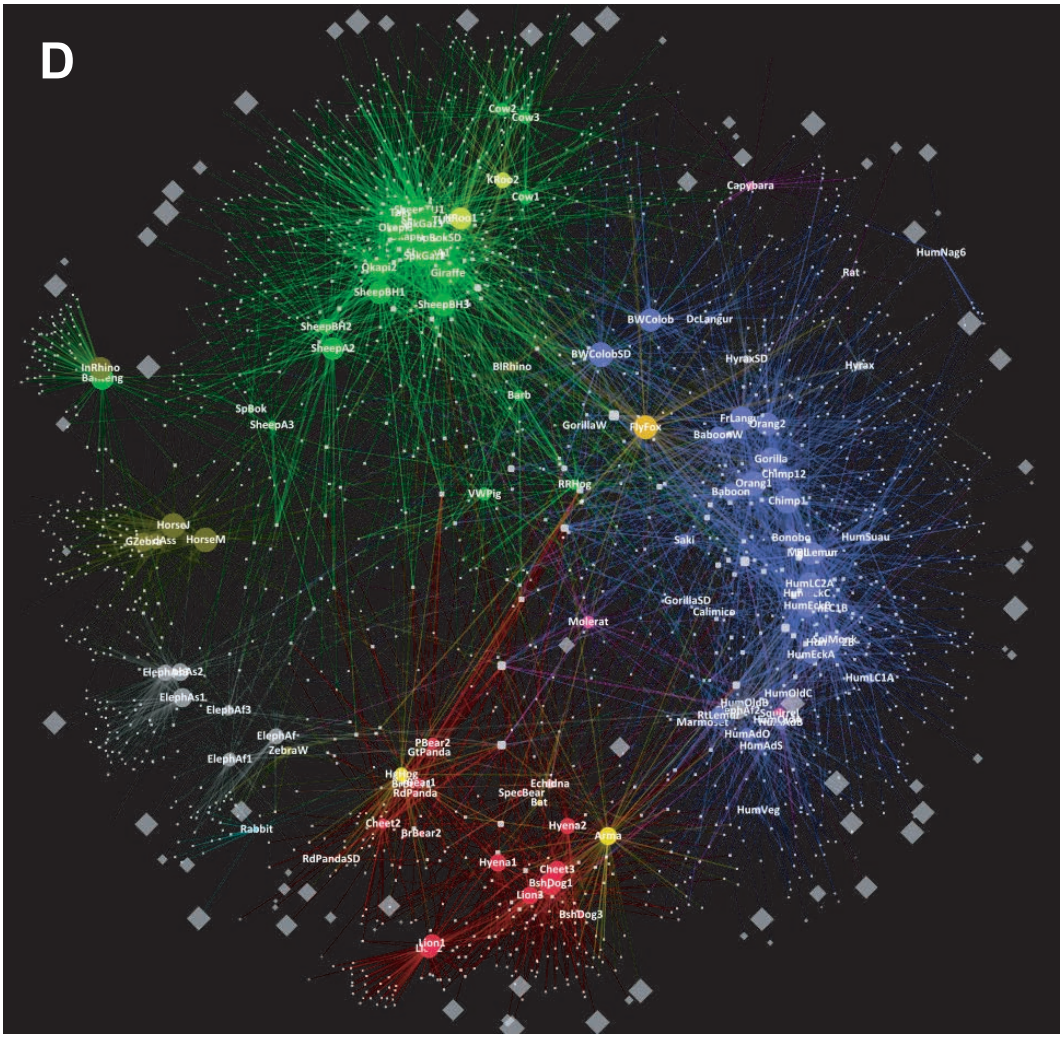
Ley *et al.*, Figure S7



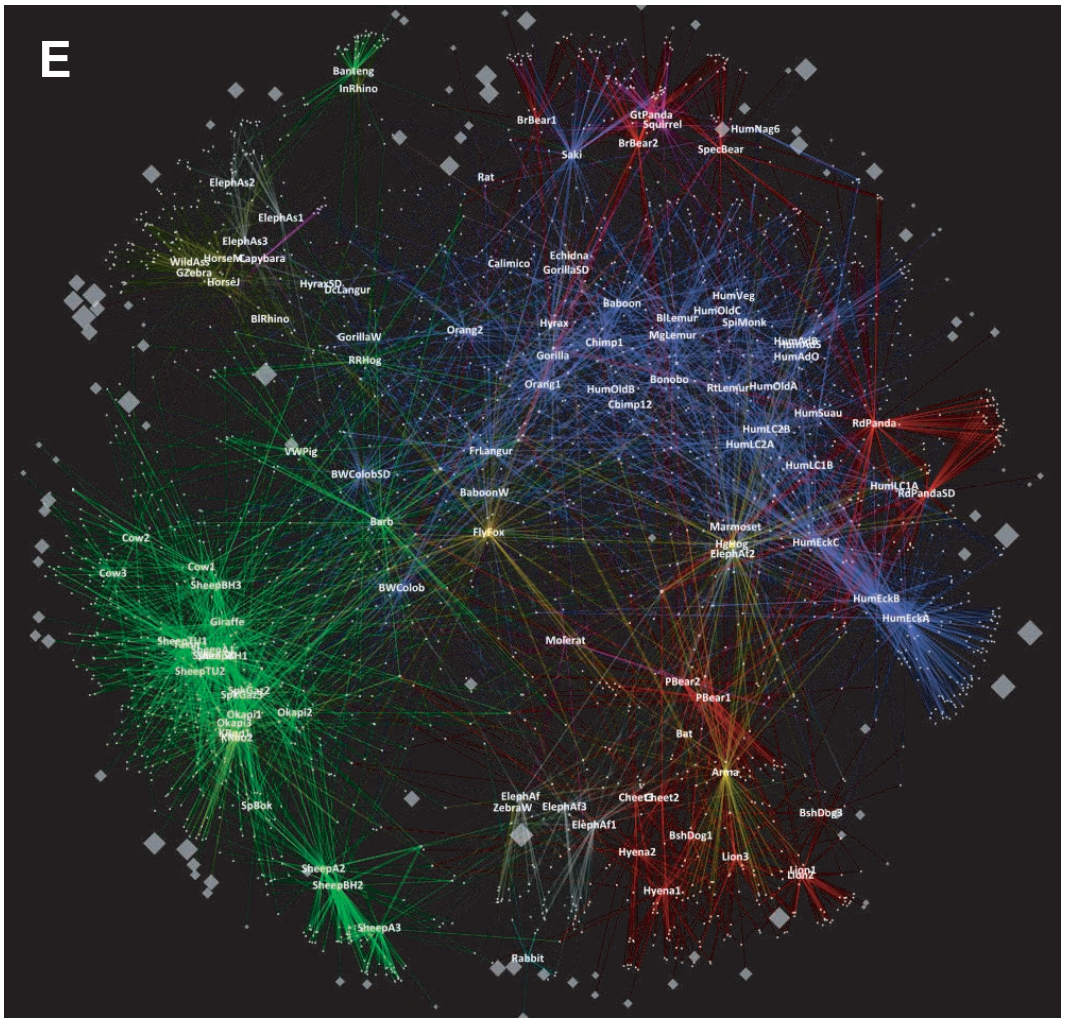
Ley *et al.*, Figure S7



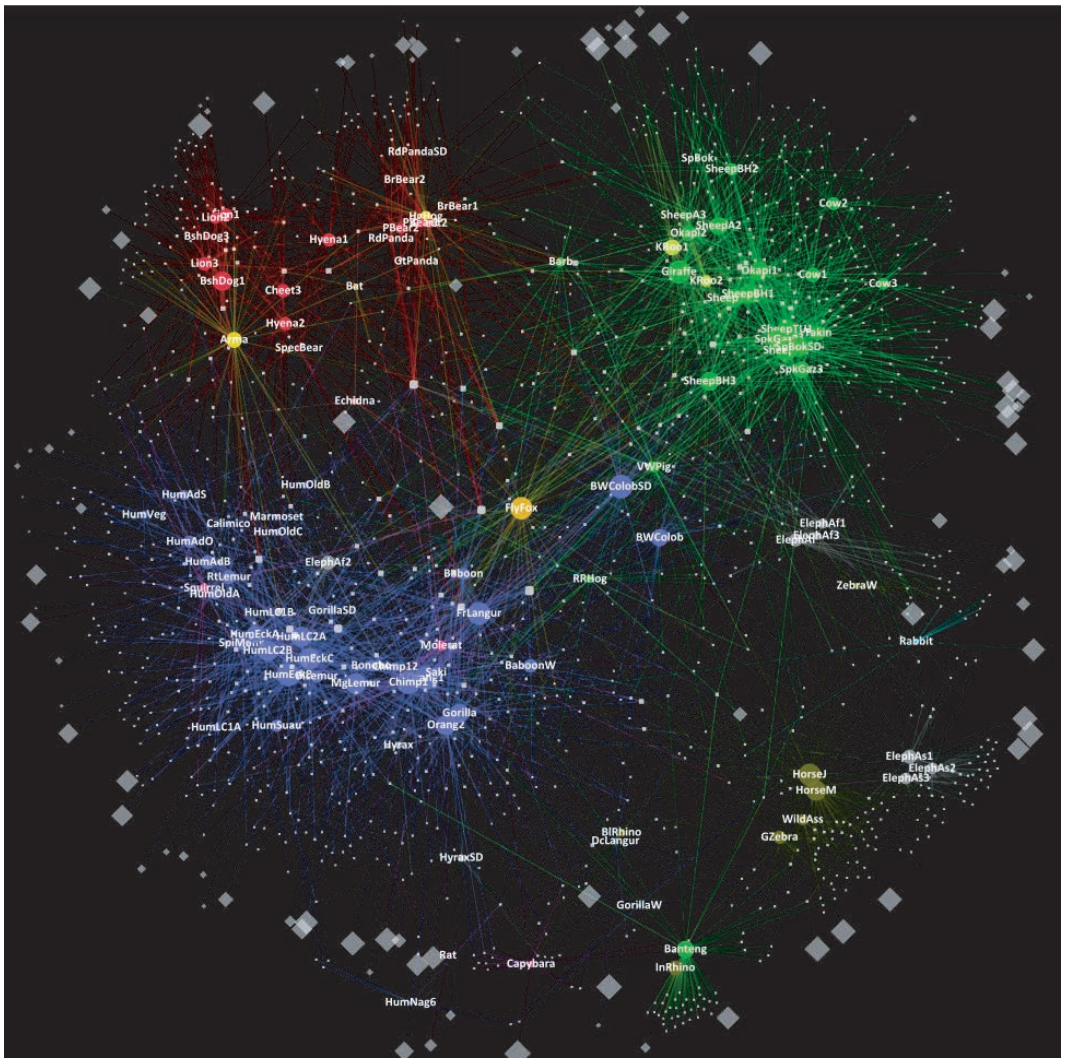
Ley *et al.*, Figure S7



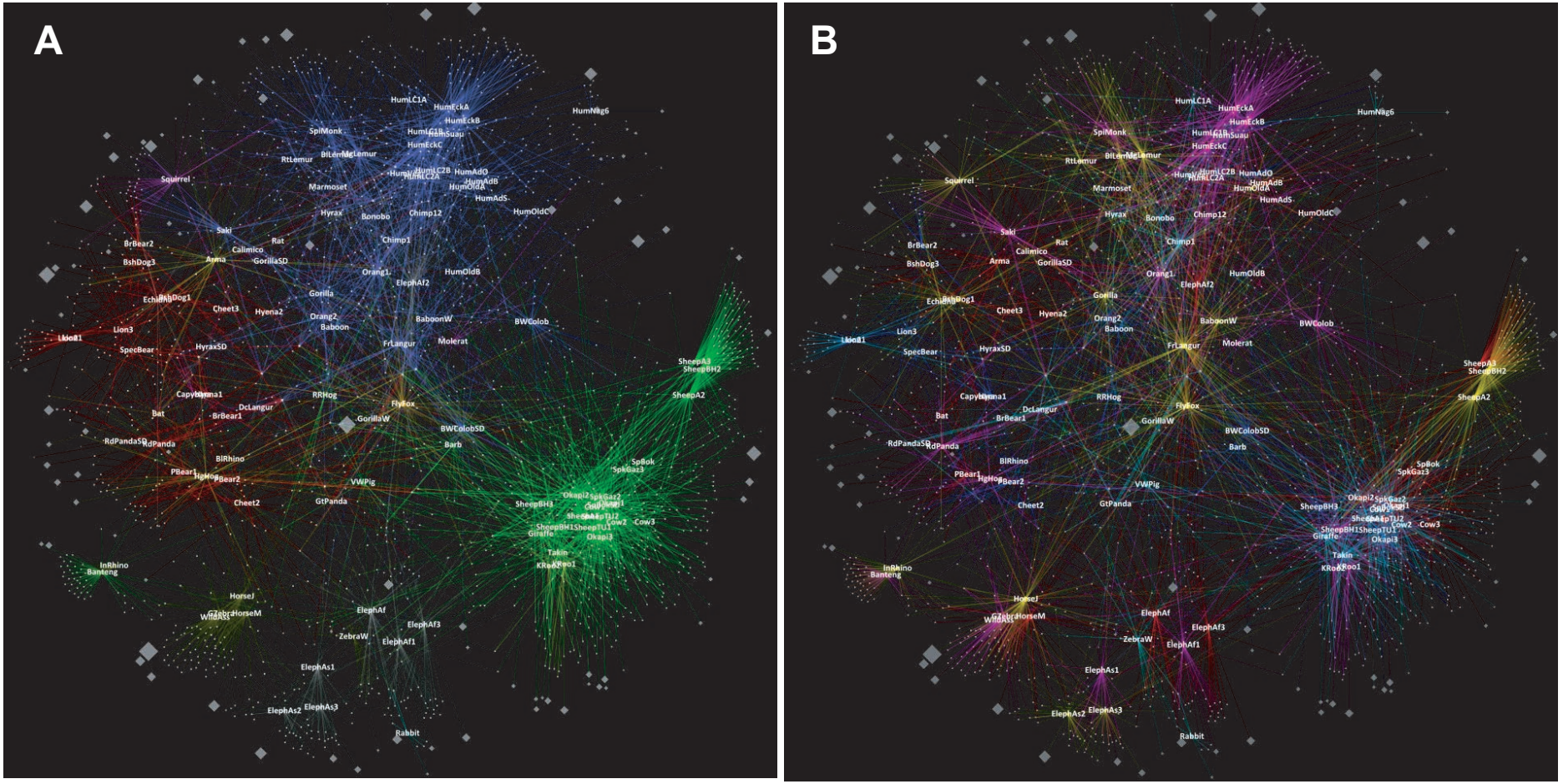
Ley *et al.*, Figure S7



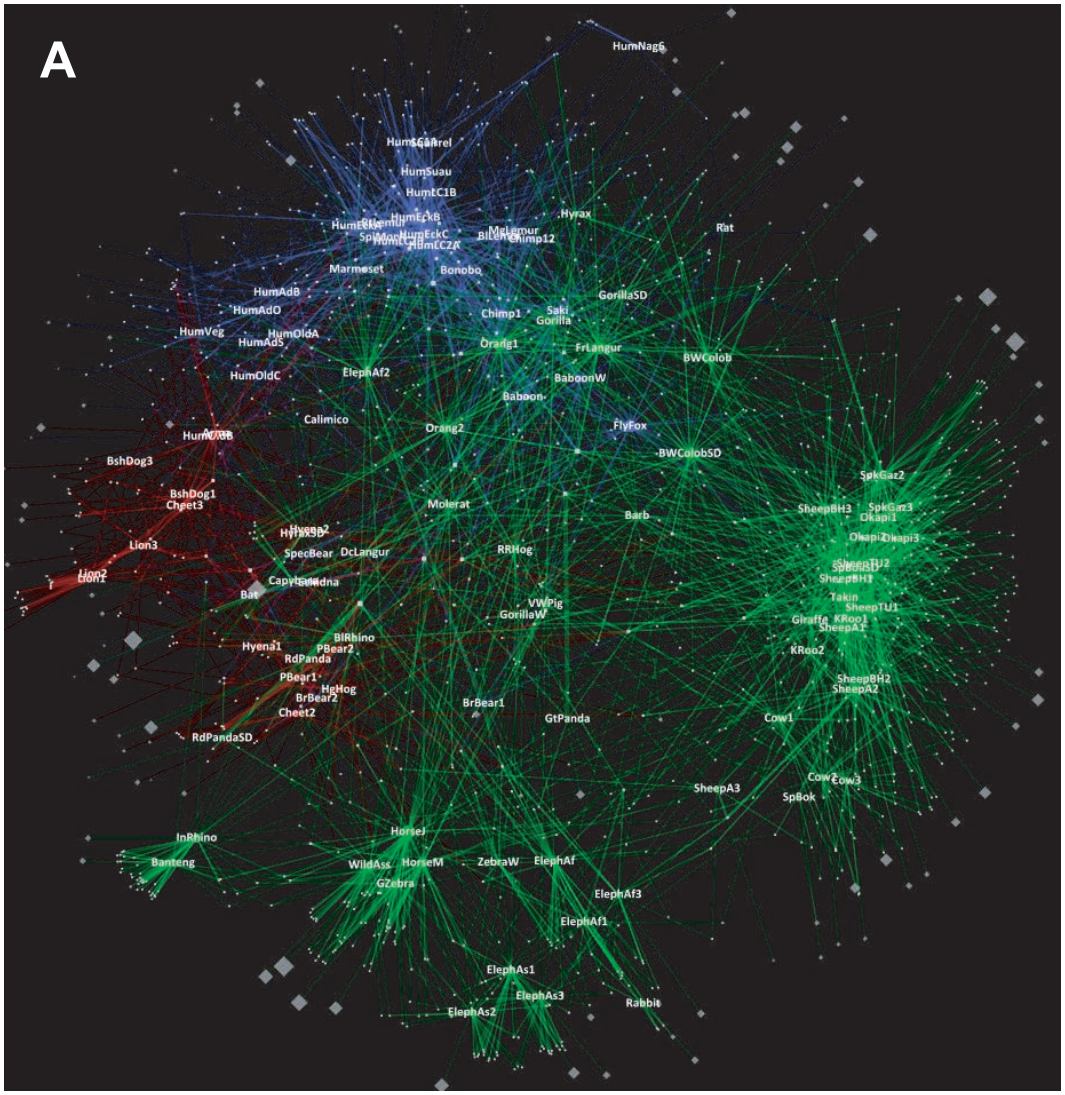
Ley *et al*, Figure S8



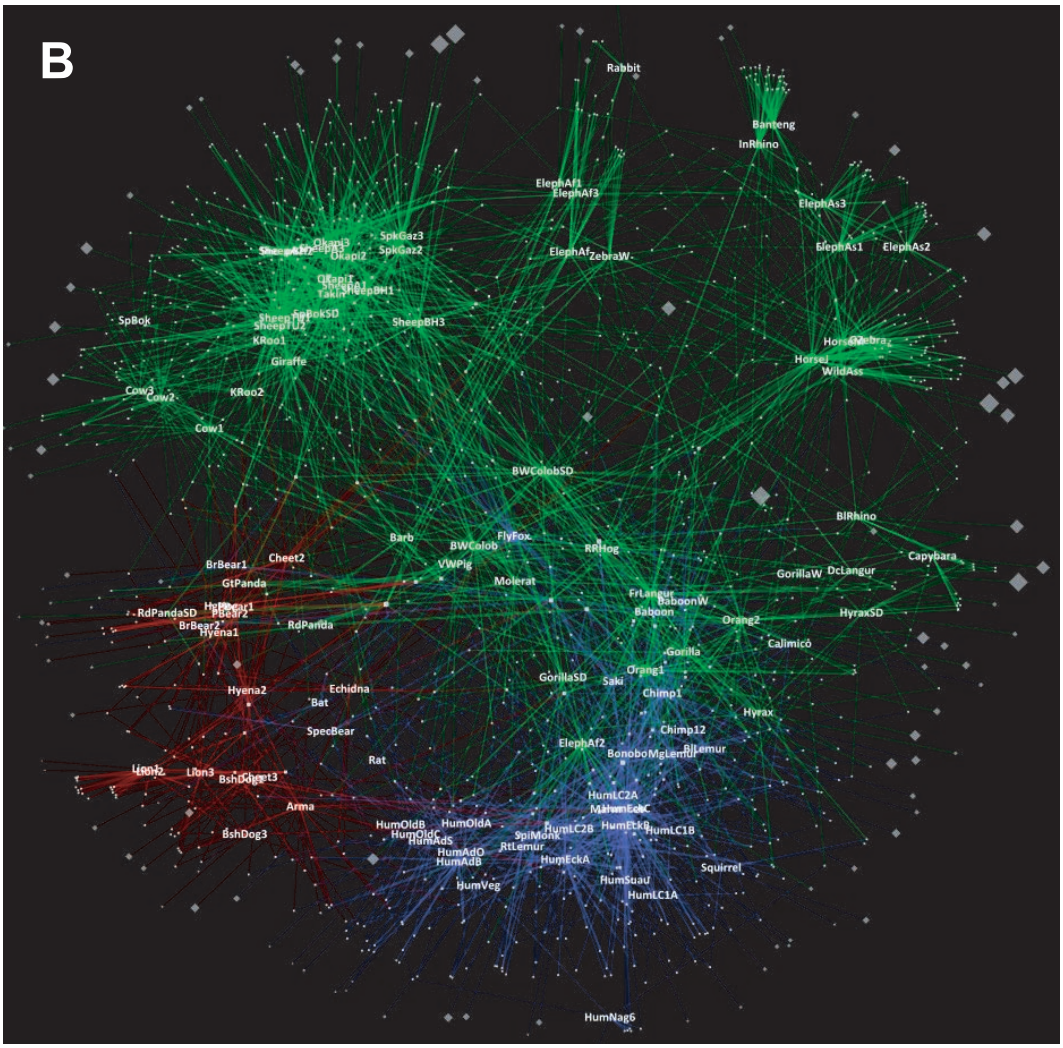
Ley *et al.*, Figure S9



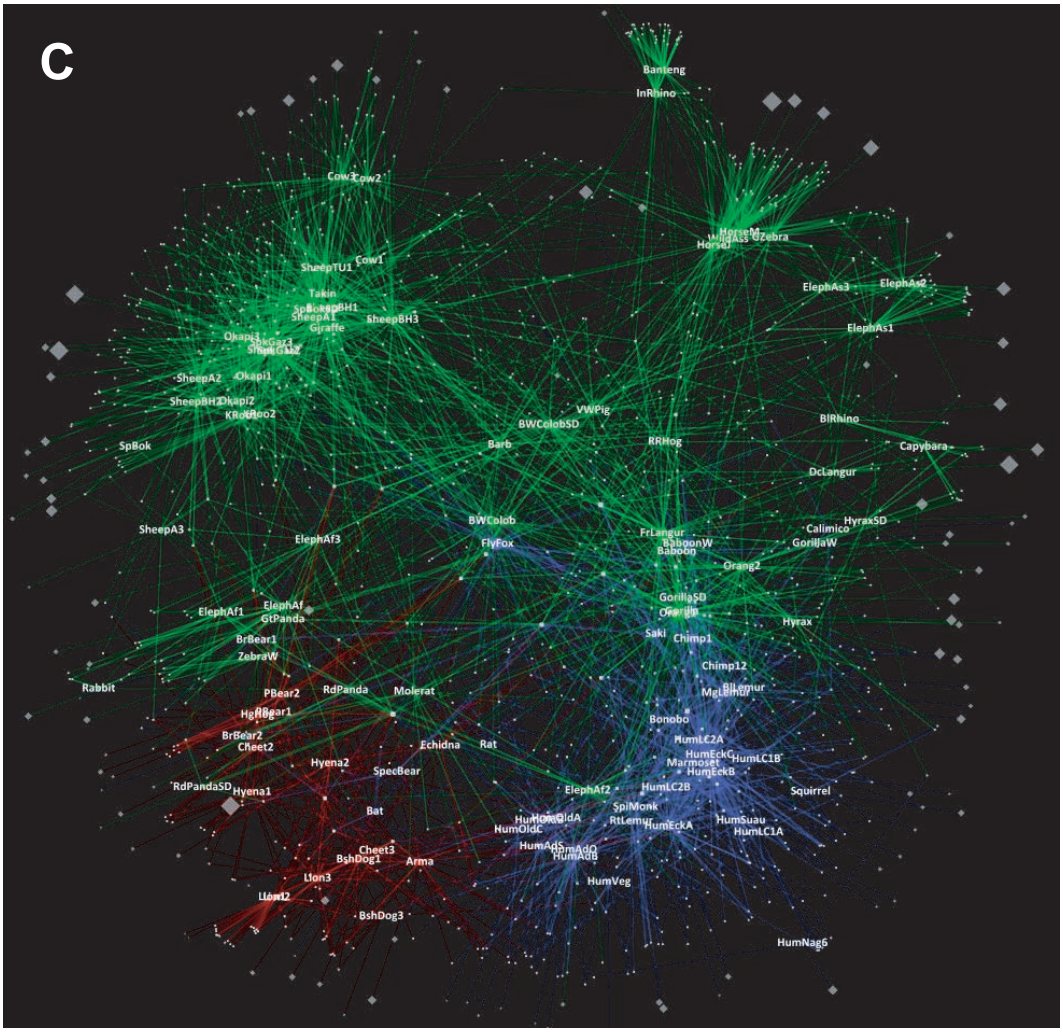
Ley *et al*, Figure S10



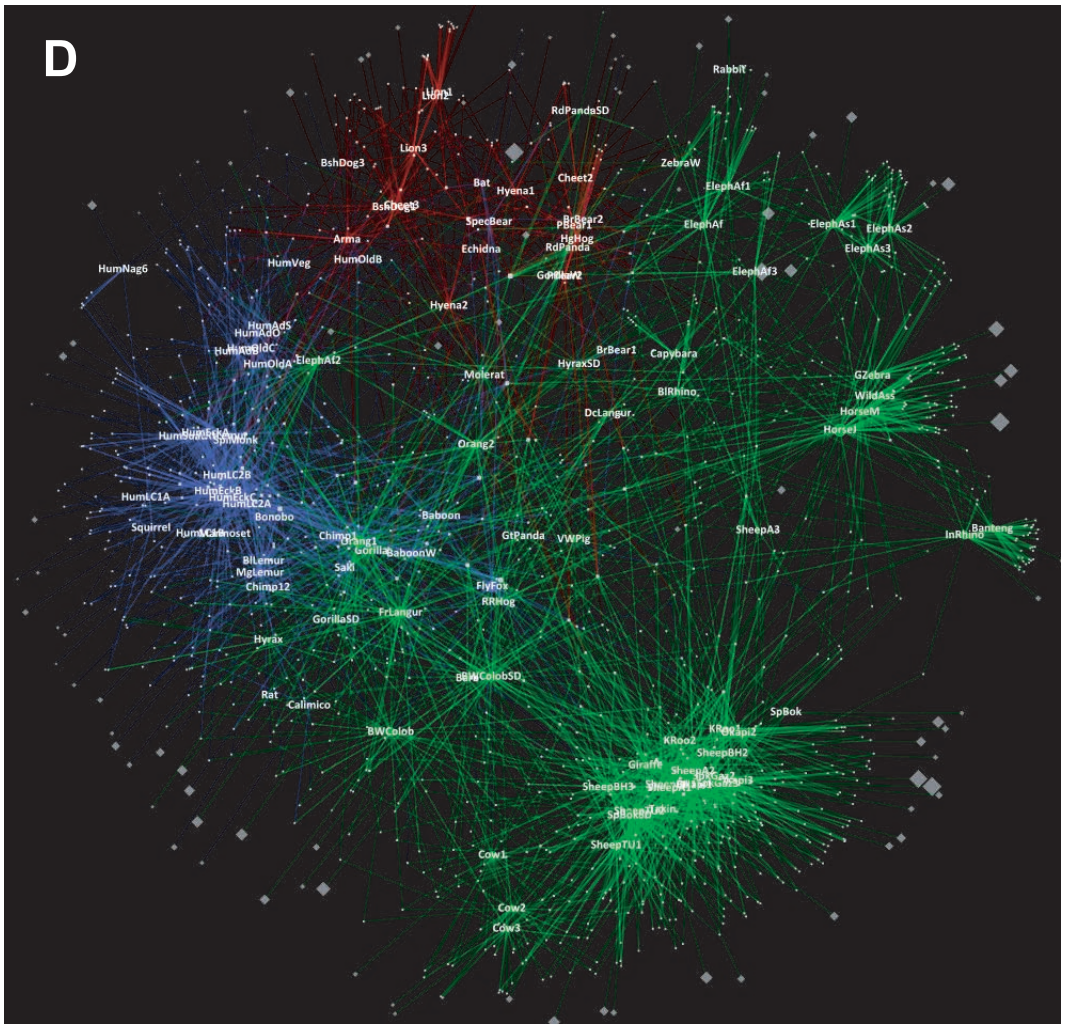
Ley *et al.*, Figure S10



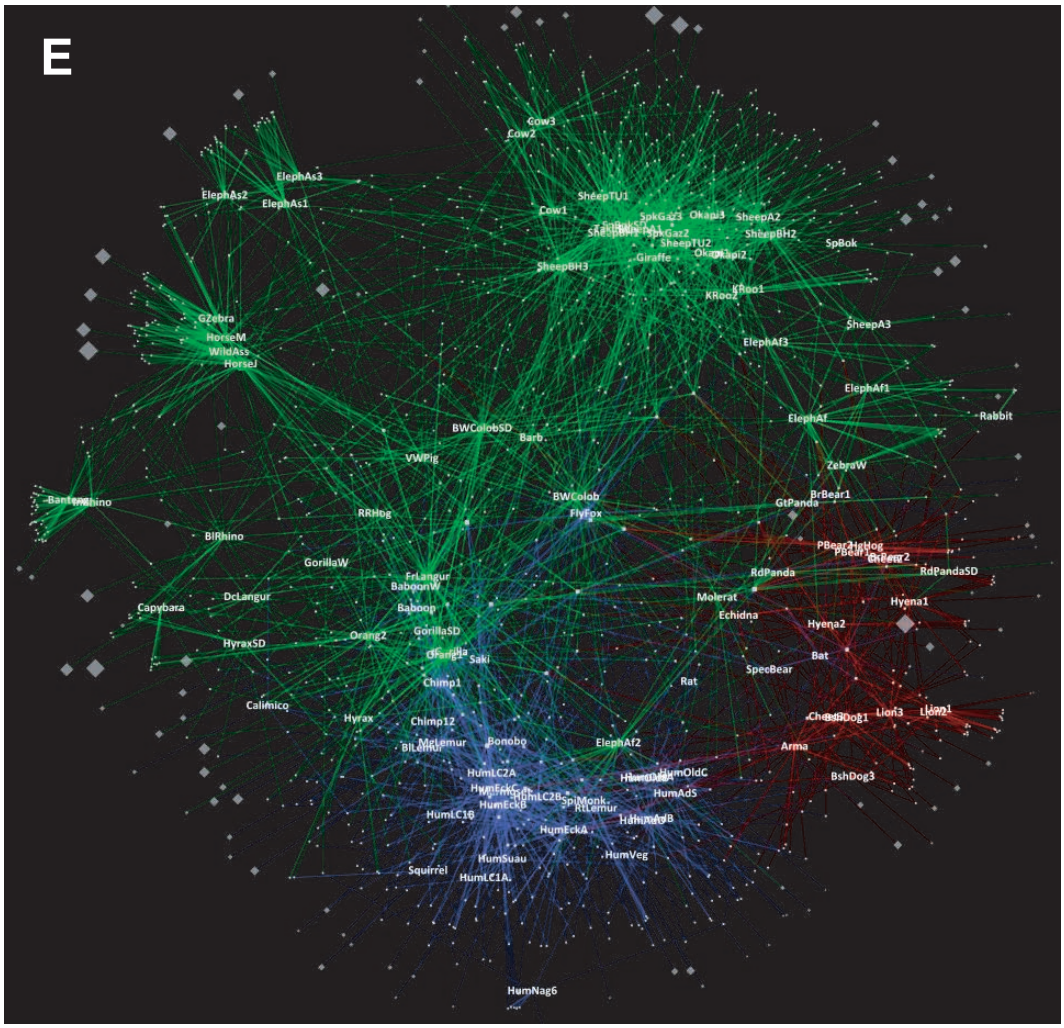
Ley *et al*, Figure S10



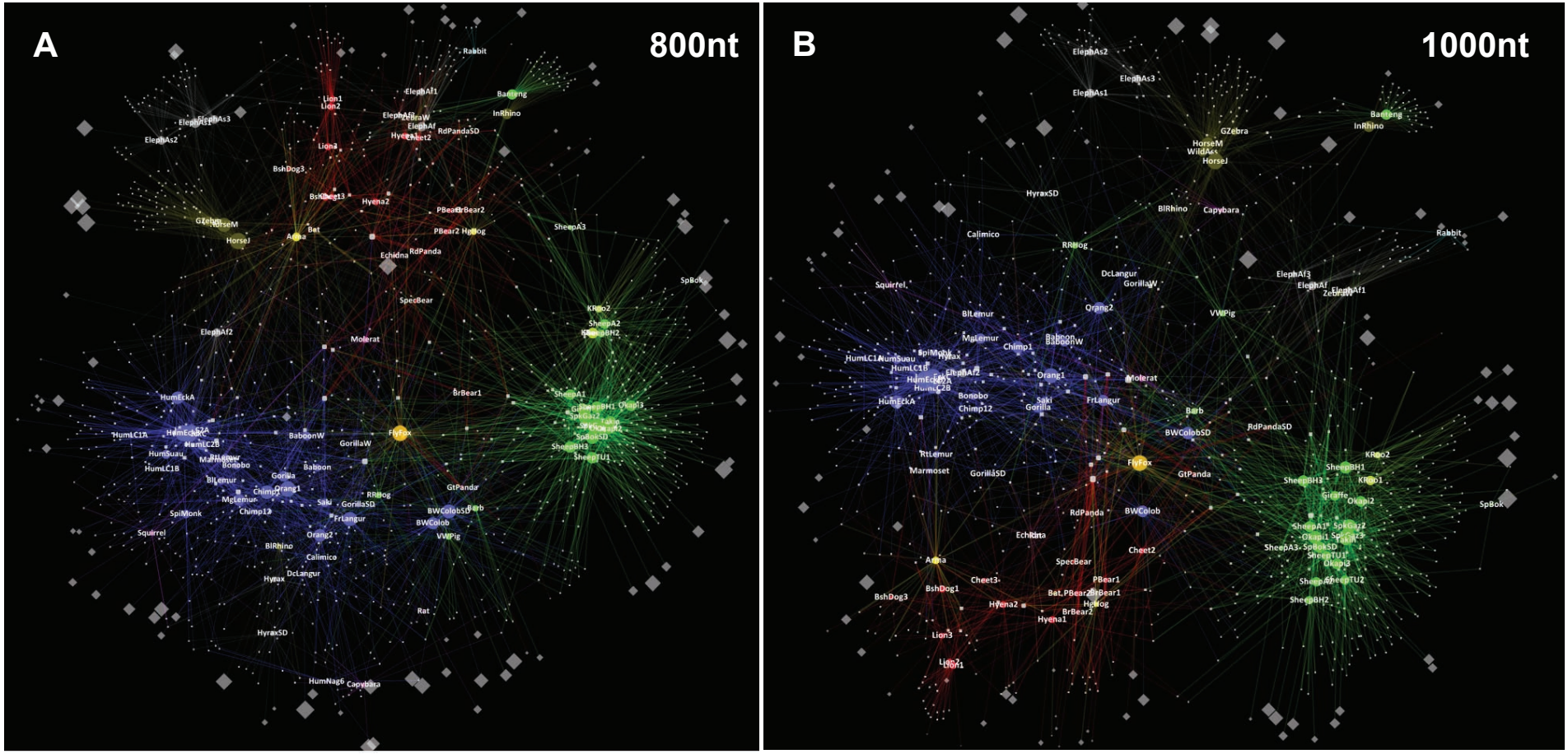
Ley *et al.*, Figure S10



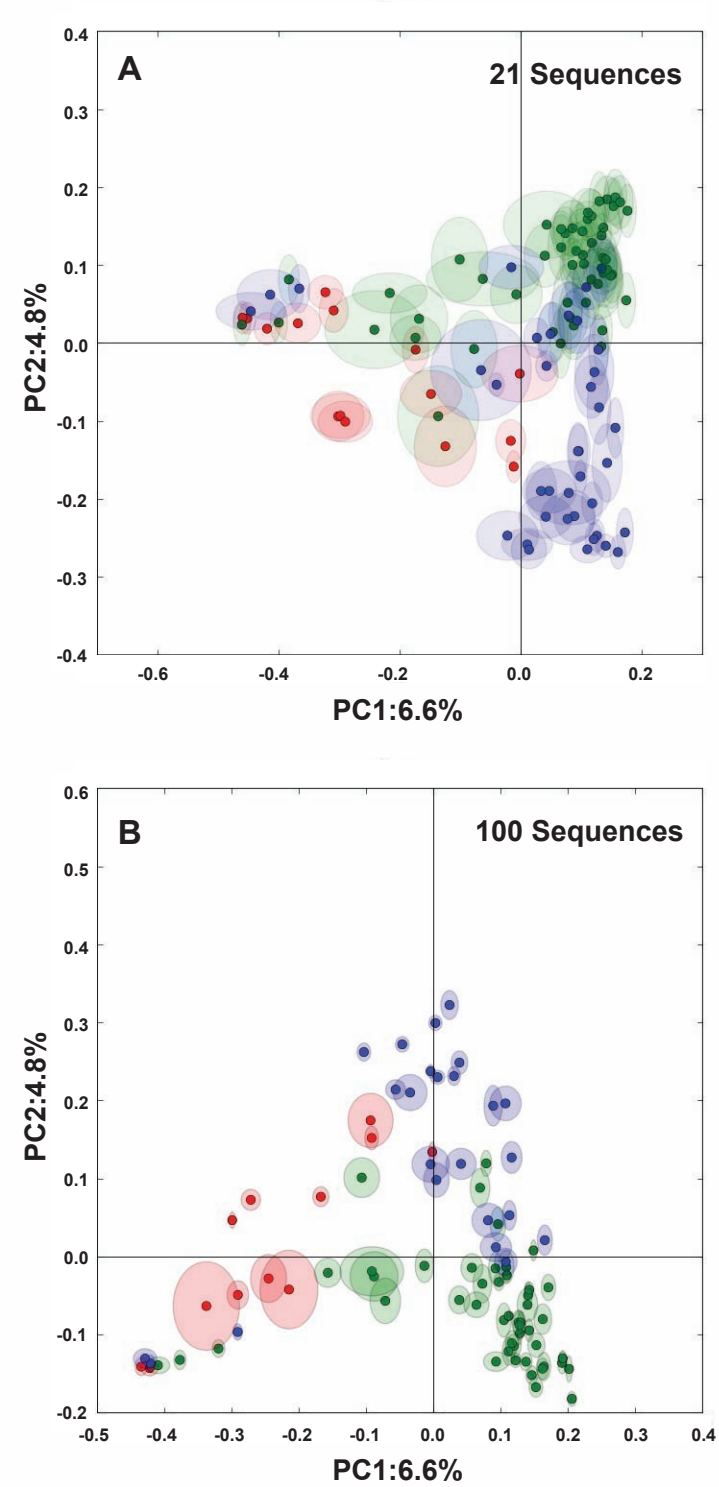
Ley *et al.*, Figure S10



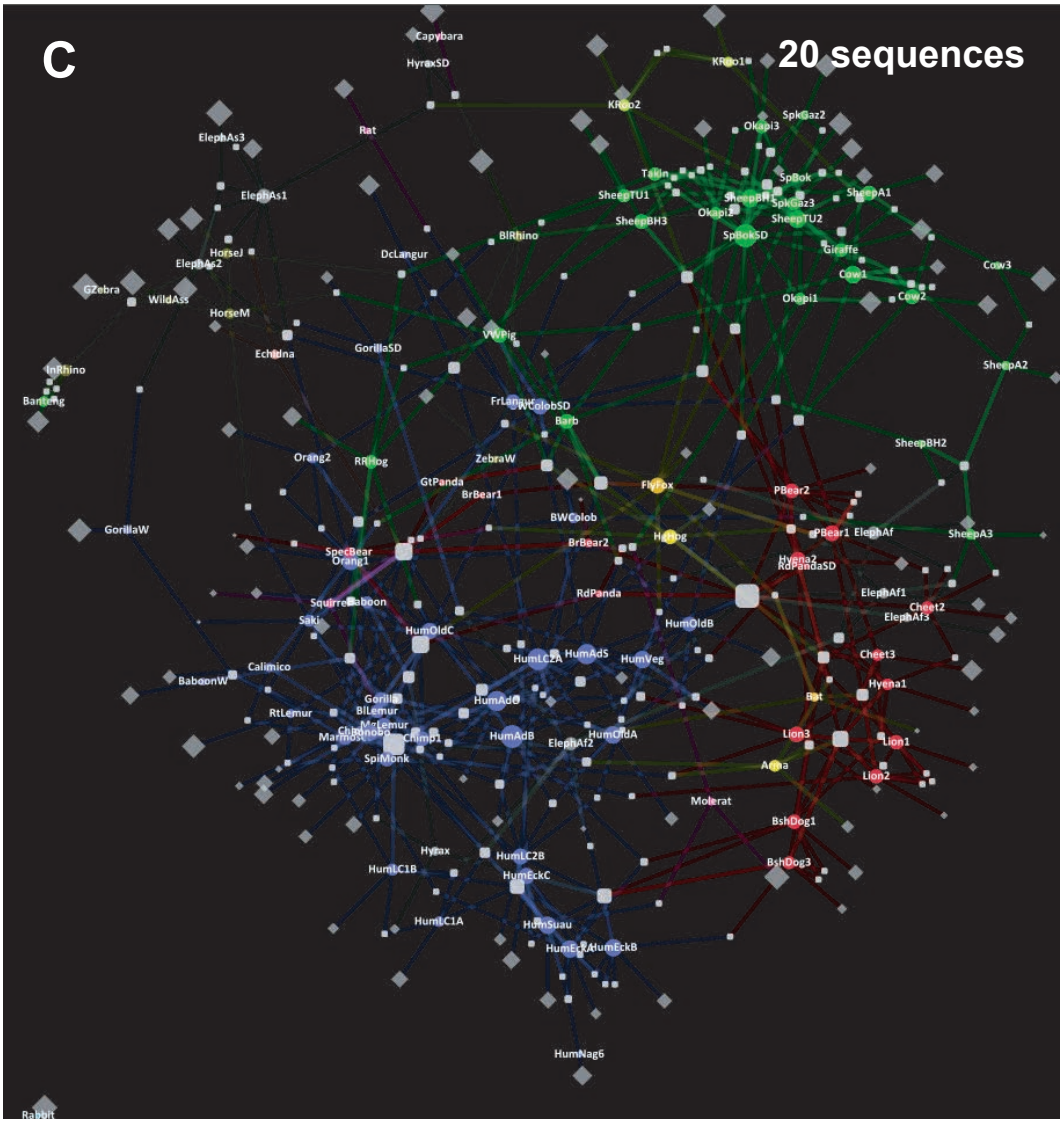
Ley *et al.*, Figure S11



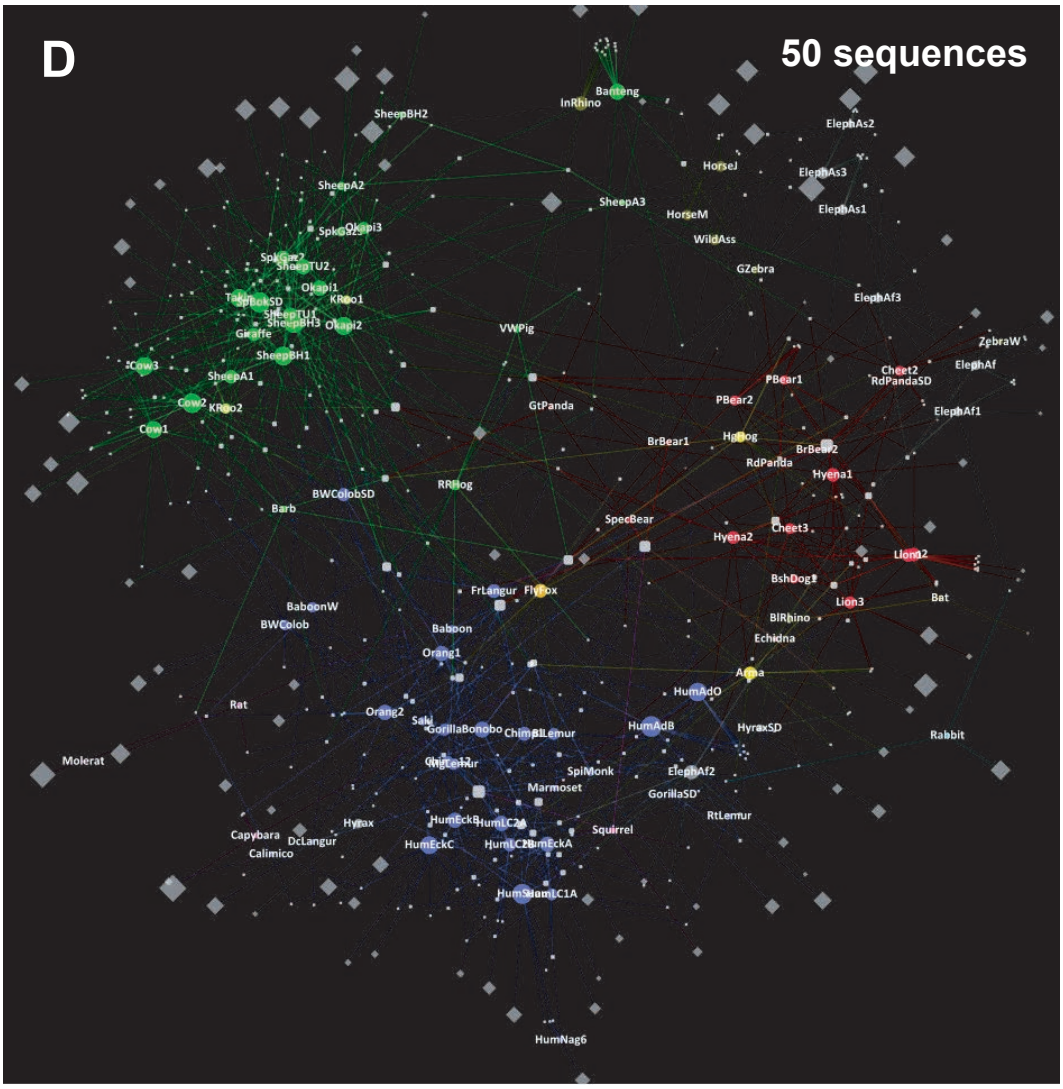
Ley *et al*, Figure S12



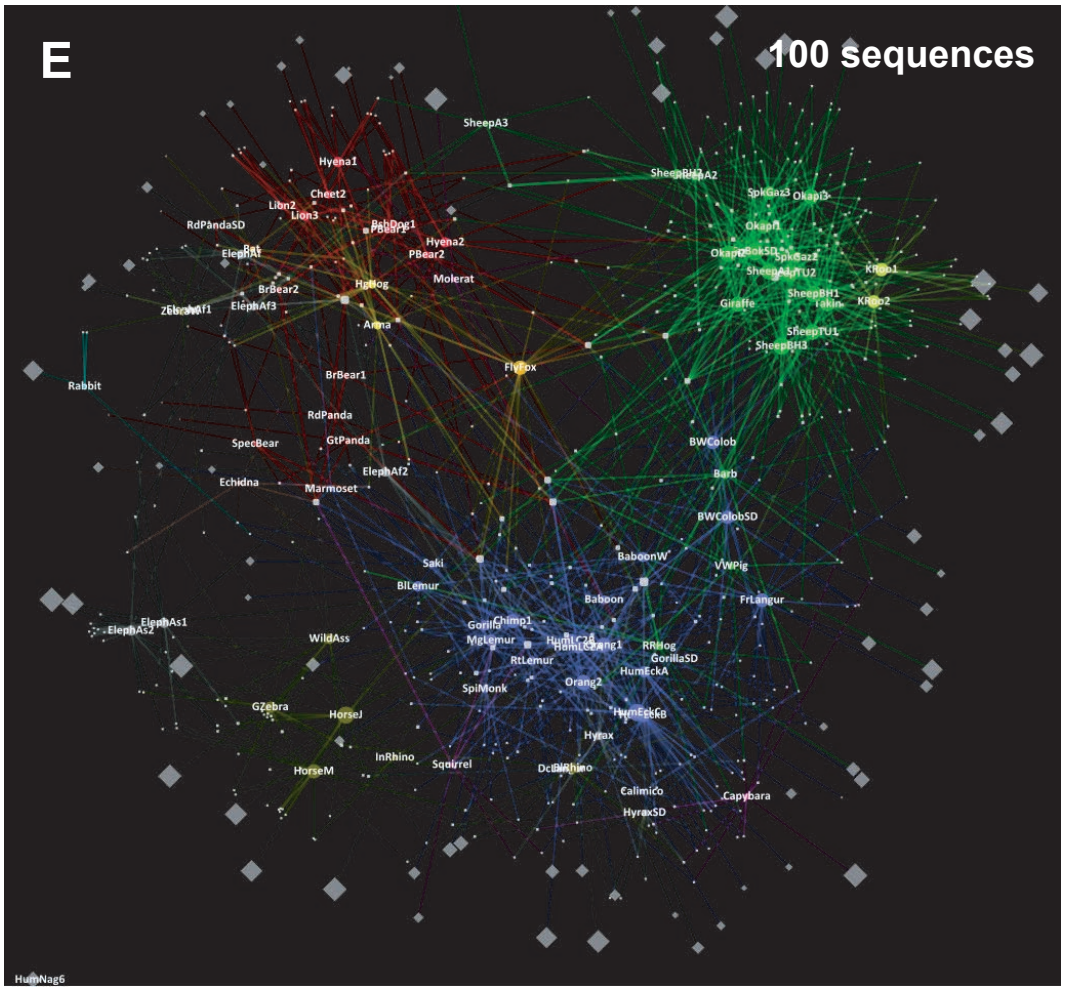
Ley *et al.*, Figure S12



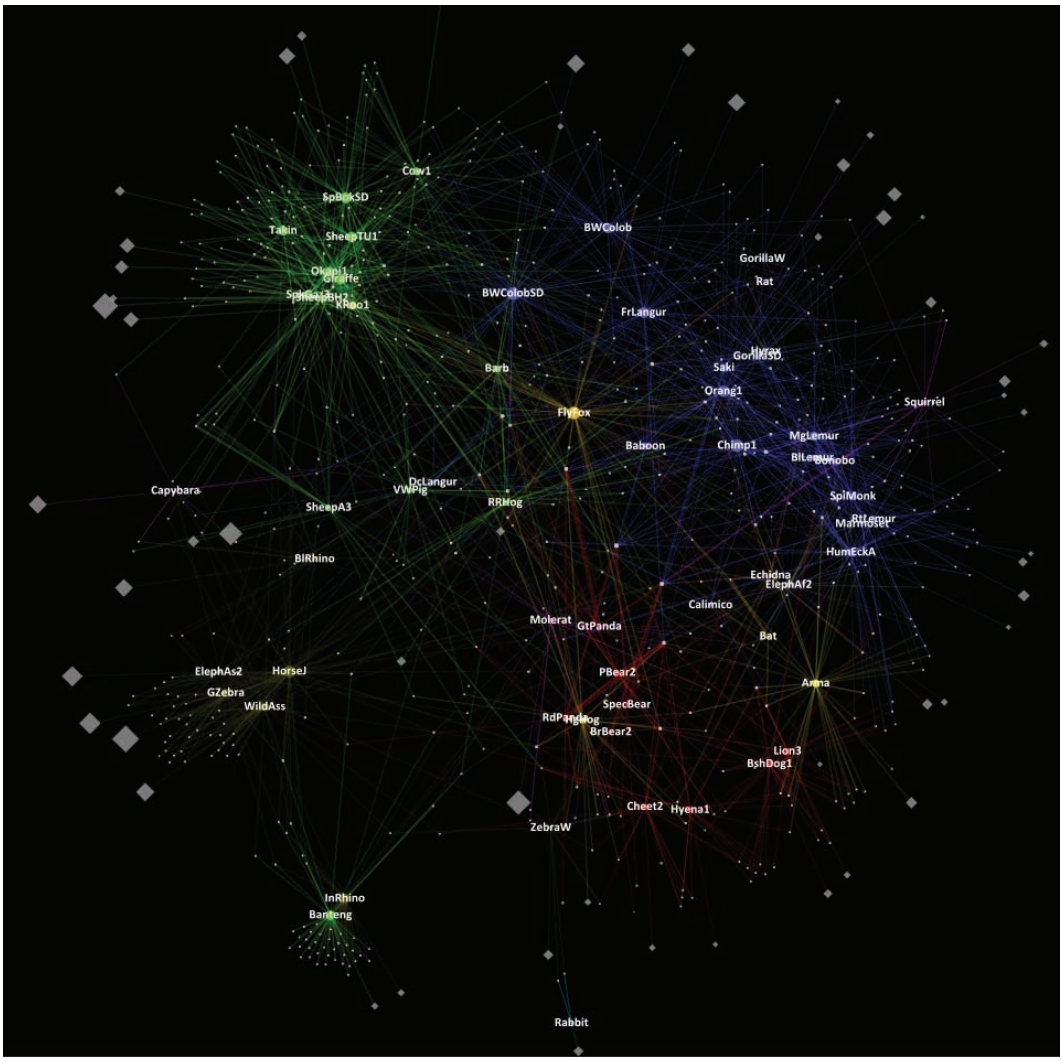
Ley *et al.*, Figure S12



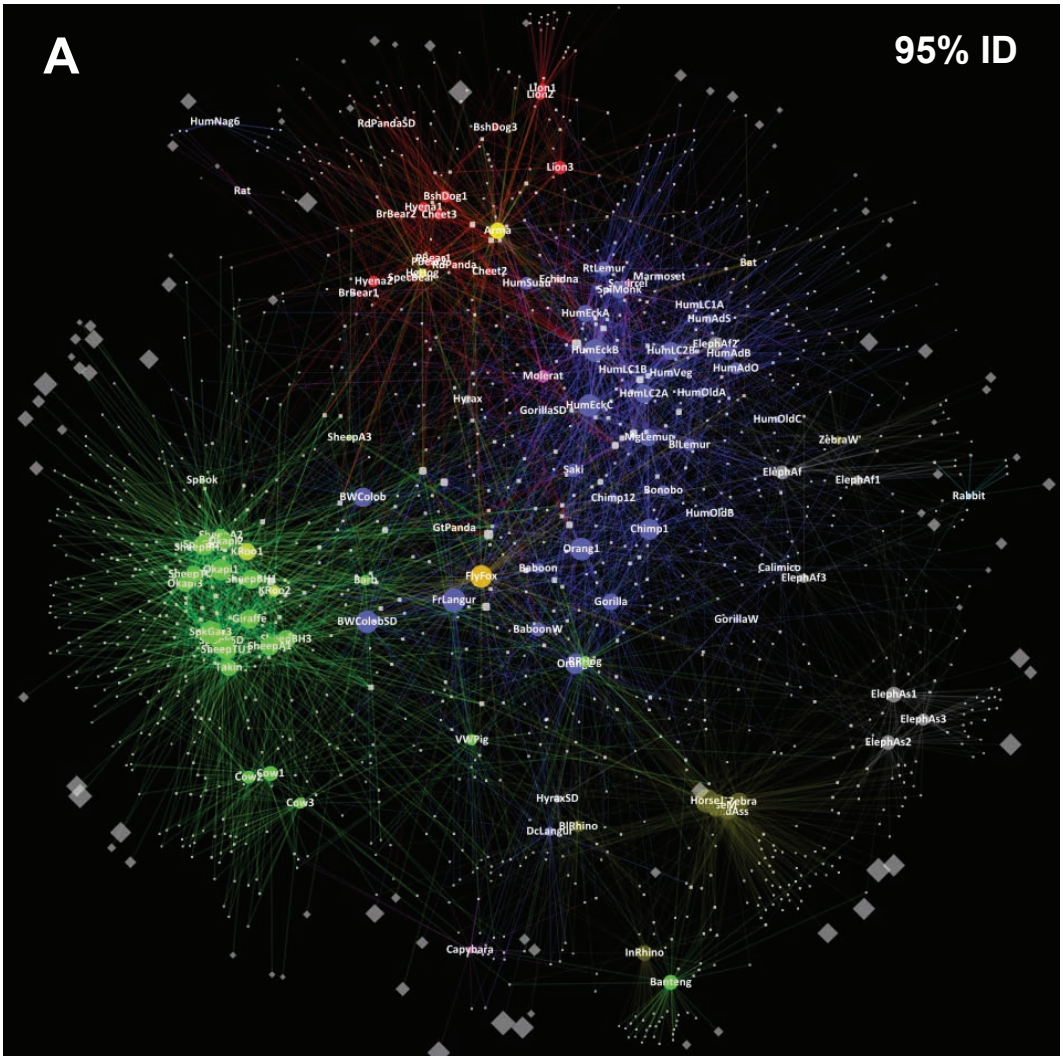
Ley *et al.*, Figure S12



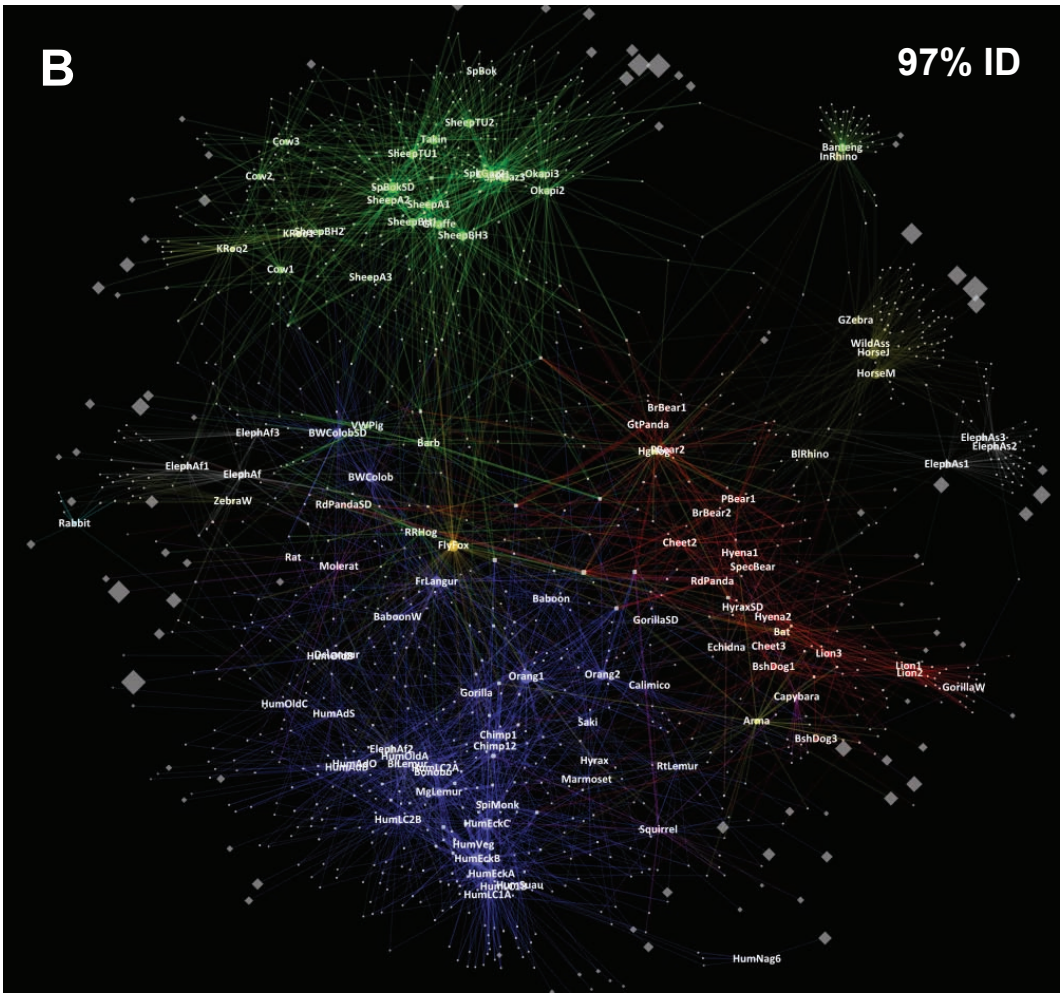
Ley *et al.*, Figure S13



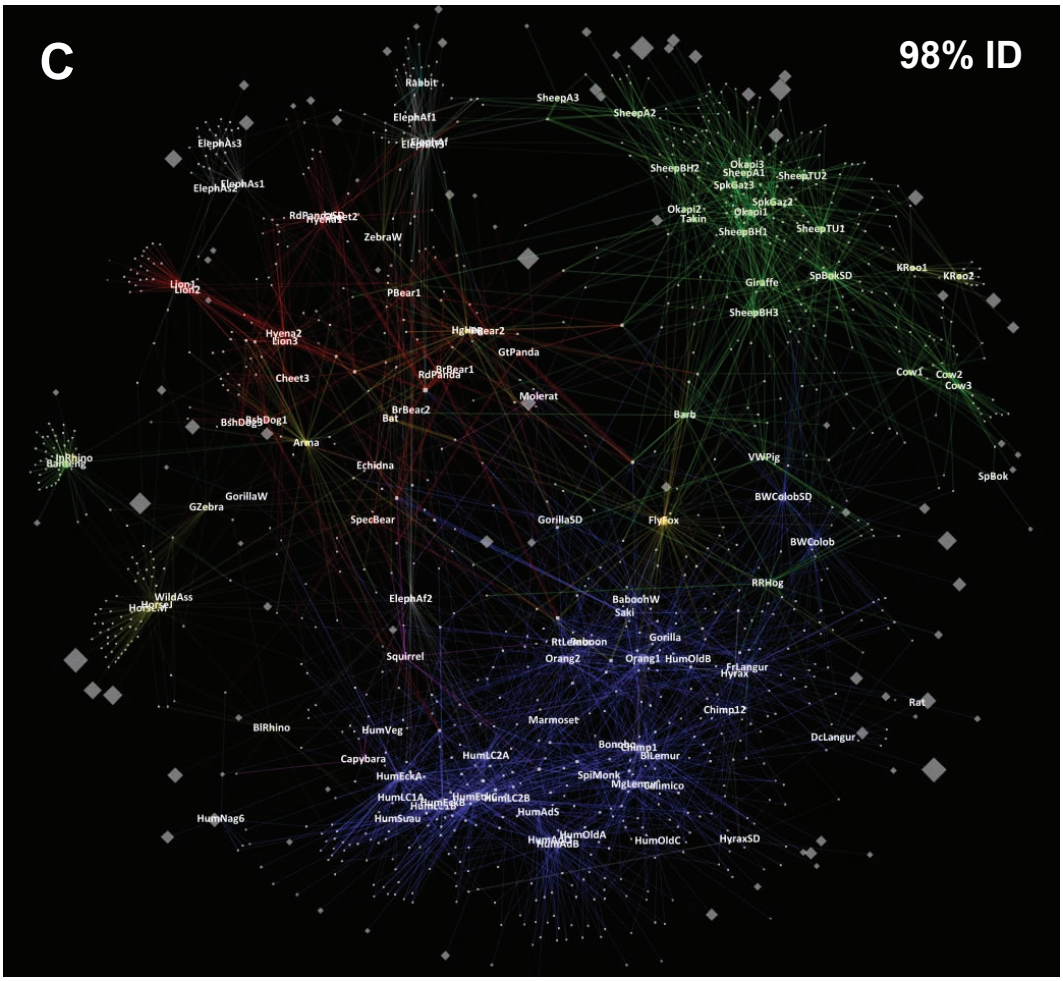
Ley *et al.*, Figure S14



Ley *et al.*, Figure S14



Ley *et al.*, Figure S14



Ley et al, Table S1

Order	Family	Genus/ species	Common name	Sample ID and sequence prefix	Number of sequences/ sample	Total OTUs 96%ID	Estimated number of OTUs unique to sample	Estimated %Unique OTUs	Provenance	Diet	Gut physiology	Order	Stable Isotopes		%C	%N	ADF	NDF	Fiber Index (ADF+1)* (NDF+1)	Fiber index category
													$\partial^{13}C$	$\partial^{15}N$						
Artiodactyla	Bovidae	*Antidorcas marsupialis*	Springbok	SBK	33	30	23	76.7	W	H	FG	AR	-16.5	9.7	36.8	2.5	27.8	50.9	1493.8	500-1500
				SBSD	163	111	41	36.9	SD	H	FG	AR	-27.2	1.2	43.0	2.0	N/A	N/A	N/A	N/A
Artiodactyla	Bovidae	*Bos javanicus*	Banteng	BG	100	83	37	44.6	ST	H	FG	AR	-27.9	2.3	39.5	1.7	28.4	45.3	1361.2	500-1500
Artiodactyla	Bovidae	*Budorcas taxicolor*	Takin	TAK	160	107	51	47.7	SD	H	FG	AR	-23.9	2.4	34.6	2.0	36.7	63.0	2412.8	1500-3000
Artiodactyla	Bovidae	*Gazella spekei*	Speke's Gazelle	SP2	223	196	129	65.8	ST	H	FG	AR	-27.5	0.9	40.8	3.2	27.8	42.0	1238.4	500-1500
				SP3	249	200	136	68.0	ST	H	FG	AR	-28.5	1.1	44.6	2.1	27.8	42.0	1238.4	500-1500
Artiodactyla	Bovidae	*Ovis ammon*	Argali Sheep	AS1	129	111	58	52.3	W	H	FG	AR	-26.1	1.5	43.2	1.5	N/A	N/A	N/A	N/A
				AS2	195	87	44	50.6	W	H	FG	AR	-26.9	5.8	41.1	1.6	N/A	N/A	N/A	N/A
				AS3	232	29	16	55.2	W	H	FG	AR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Artiodactyla	Bovidae	*Ovis canadensis*	Bighorn Sheep	BH1	193	113	57	50.4	W	H	FG	AR	-27.7	0.7	46.2	2.0	N/A	N/A	N/A	N/A
				BH2	263	70	34	48.6	W	H	FG	AR	-26.5	3.6	42.8	2.2	N/A	N/A	N/A	N/A
				BHSD	164	99	44	44.4	SD	H	FG	AR	-25.9	2.7	44.5	3.1	27.3	50.1	1447.3	500-1500
Artiodactyla	Bovidae	*Ovis vignei*	Trancaspian Urial Sheep	TU1	170	140	83	59.3	ST	H	FG	AR	-27.4	3.9	43.3	2.3	28.4	41.8	1258.3	500-1500
				TU2	136	115	62	53.9	ST	H	FG	AR	-26.6	3.1	44.1	2.6	28.4	41.8	1258.3	500-1500
Artiodactyla	Giraffidae	*Giraffa camelopardalis reticulata*	Reticulated Giraffe	gir	176	138	82	59.4	ST	H	FG	AR	-26.1	2.0	44.9	1.0	32.0	44.3	1494.9	500-1500
Artiodactyla	Giraffidae	*Okapia johnstoni*	Okapi	OK1	153	118	63	53.4	ST	H	FG	AR	-27.2	2.6	43.7	1.9	27.0	41.0	1176.0	500-1500
				OK2	120	88	45	51.1	ST	H	FG	AR	-26.2	3.0	44.7	1.8	27.0	41.0	1176.0	500-1500
				OK3	120	93	49	52.7	ST	H	FG	AR	-28.7	1.8	44.2	2.3	27.0	41.0	1176.0	500-1500
Artiodactyla	Suidae	*Babyrousa babyrussa*	Babirusa	BARB	174	58	27	46.6	SD	H	FG	AR	N/A	N/A	N/A	N/A	30.7	53.3	1721.3	1500-3000
Artiodactyla	Suidae	*Potamochoerus porcus*	Red River Hog	RRH	167	73	45	61.6	SD	H	FG	AR	-20.8	2.2	42.1	1.6	27.8	52.0	1526.4	1500-3000
Artiodactyla	Suidae	*Sus cebifons*	Visayam Warty Pig	VWP	144	82	58	70.7	SD	H	FG	AR	-27.4	1.1	26.0	1.2	28.7	46.7	1416.7	500-1500
Carnivora	Canidae	*Speothos venaticus*	Bushdog	bdog1	140	75	42	56.0	ST	C	S	CA	-23.5	5.9	32.7	1.8	0.0	16.0	17.0	0-50
				bdog3	35	19	9	47.4	ST	C	S	CA	-24.0	11.5	17.7	0.9	0.0	16.0	17.0	0-50
Carnivora	Felidae	*Acinonyx jubatus*	Cheetah	CE2	186	57	34	59.6	ST	C	S	CA	-26.3	1.7	43.0	3.6	0.0	41.0	42.0	0-50
				CE3	91	53	32	60.4	ST	C	S	CA	-23.6	7.3	19.6	1.4	0.0	41.0	42.0	0-50
Carnivora	Hyaenidae	*Crocuta crocuta*	Spotted Hyena	HY1	150	49	18	36.7	ST	C	S	CA	-21.4	6.4	9.1	0.9	0.0	32.0	33.0	0-50
				HY2	113	53	28	52.8	ST	C	S	CA	-21.8	8.1	3.5	0.3	0.0	32.0	33.0	0-50
Carnivora	Pantherinae	*Panthera leo*	Lion	LI1	80	29	1	3.4	ST	C	S	CA	-20.4	8.0	27.5	3.9	0.0	41.8	42.8	0-50
				LI2	111	36	5	13.9	ST	C	S	CA	N/A	N/A	N/A	N/A	0.0	41.8	42.8	0-50
				LI3	221	61	30	49.2	ST	C	S	CA	-24.2	7.7	33.9	2.0	0.0	41.8	42.8	0-50
Carnivora	Ursidae	*Ailuropoda melanoleuca*	Giant Panda	GP	565	17	13	76.5	SD	H	S	CA	-24.8	1.8	42.0	2.1	36.9	61.3	2361.2	1500-3000
Carnivora	Ailuridae	*Ailurus fulgens*	Red Panda	RP	850	24	16	66.7	ST	H	S	CA	-23.6	2.1	42.1	1.7	15.8	28.0	487.2	50-500
				RPSD	301	10	5	50.0	SD	H	S	CA	-22.1	8.3	15.4	1.8	14.5	25.8	415.4	50-500
Carnivora	Ursidae	*Tremarctos ornatus*	Spectacled Bear	SB	195	18	9	50.0	ST	O	S	CA	-22.1	1.5	42.6	2.4	12.0	21.5	292.5	50-500
Carnivora	Ursidae	*Ursus americanus*	North American Black Bear	BB1	178	14	6	42.9	ST	O	S	CA	-23.1	6.0	42.0	2.3	1.0	6.0	14.0	0-50
				BB2	196	14	6	42.9	ST	O	S	CA	-24.4	3.1	41.1	2.1	1.0	6.0	14.0	0-50
Carnivora	Ursidae	*Ursus maritimus*	Polar Bear	PB1	221	33	17	51.5	ST	C	S	CA	-22.1	8.3	15.4	1.8	0.0	17.3	18.3	0-50
				PB2	274	34	15	44.1	ST	C	S	CA	-22.5	7.9	27.8	2.9	0.0	17.3	18.3	0-50
Chiroptera	Phyllostomidae	*Carollia perspicillata*	Seba's Short-tailed Bat	bat	274	17	9	52.9	ST	O	S	CH	-24.2	2.9	40.8	1.0	0.7	1.0	3.3	0-50
	Pterodidae	*Pteropus giganteus*	Flying Fox	FF	228	109	34	31.2	ST	O	S	CH	-25.0	2.6	36.4	1.9	0.8	1.1	3.7	0-50
Hyracoidea	Procaviidae	*Procavia capensis*	Rock Hyrax	HRX	142	41	18	43.9	ST	H	FG	PO	-27.7	3.9	43.5	3.0	18.2	32.0	633.6	500-1500
				RHSD	119	45	28	62.2	SD	H	FG	PO	-25.3	3.6	44.3	3.7	14.8	25.9	425.0	50-500
Insectivora	Erinaceidae	*Atelerix albiventris*	Hedgehog	HH	218	54	30	55.6	ST	C	S	IN	-26.7	2.5	41.4	1.5	0.4	0.0	1.4	0-50

Lagomorpha	Leporidae	*Oryctolagus cuniculus*	European Rabbit	RA	111	54	47	87.0	ST	H	HG	LA	-26.9	3.2	41.6	3.5	1.8	2.1	8.5	0-50
Perissodactyla	Equidae	*Equus asinus*	Somali Wild Ass	WA	181	143	100	69.9	ST	H	HG	PE	-28.5	1.9	45.5	1.0	33.6	54.0	1903.0	1500-3000
Perissodactyla	Equidae	*Equus equus*	Horse	horsej	304	223	140	62.8	W	H	HG	PE	-25.4	1.6	33.1	1.0	N/A	N/A	N/A	N/A
				horsem	214	156	93	59.6	W	H	HG	PE	-25.4	2.2	35.5	0.8	N/A	N/A	N/A	N/A
Perissodactyla	Equidae	*Equus grevyi*	Grevy's Zebra	GZ	222	158	114	72.2	ST	H	HG	PE	-28.1	1.6	42.9	2.7	41.0	60.0	2562.0	1500-3000
Perissodactyla	Equidae	*Equus hartmannae*	Hartmann's Mountain Zebra	AFZEB	229	69	45	65.2	W	H	HG	PE	-14.5	12.0	34.8	1.2	N/A	N/A	N/A	N/A
Perissodactyla	Rhinocerotidae	*Diceros bicornis*	Black Rhinoceros	RH	179	119	92	77.3	ST	H	HG	PE	-26.5	6.3	23.7	1.1	36.0	48.0	1813.0	1500-3000
Perissodactyla	Rhinocerotidae	*Rhinoceros unicornis*	Indian Rhinoceros	IR	113	84	35	41.7	SD	H	HG	PE	-14.5	6.8	44.9	1.4	32.8	58.2	2001.0	1500-3000
Primates	Atelidae	*Ateles geoffroyi*	Black-handed Spider Monkey	SPIM	275	83	44	53.0	ST	O	S	PR	-26.4	4.5	18.0	1.1	11.5	18.4	242.5	50-500
Primates	Callitrichidae	*Callithrix geoffroyi*	Geoffrey's marmoset	MAR	205	36	24	66.7	ST	O	S	PR	-24.6	5.3	24.1	3.1	0.5	1.4	3.6	0-50
Primates	Cebidae	*Callimico goeldii*	Goeldi's Marmoset	CAL	114	32	24	75.0	ST	O	S	PR	-24.6	3.6	26.4	3.7	0.1	1.4	2.6	0-50
Primates	Cercopithecidae	*Colobus angolensis*	East Angolan Colobus	EAC	319	162	95	58.6	SD	H	FG	PR	-26.0	3.1	40.3	3.7	11.8	20.6	276.5	50-500
Primates	Cercopithecidae	*Colobus guereza*	Eastern Black and White Colobus	COL	207	124	66	53.2	ST	H	FG	PR	-26.7	3.2	41.1	2.5	14.5	4.3	81.5	50-500
Primates	Cercopithecidae	*Papio hamadryas*	Hamadryas Baboon	BAZ	198	71	42	59.2	ST	O	S	PR	N/A	N/A	N/A	N/A	12.5	12.0	175.5	50-500
				AFBAB	180	88	51	58.0	W	O	S	PR	-25.5	6.5	35.6	2.4	N/A	N/A	N/A	N/A
Primates	Cercopithecidae	*Presbytis francoisi*	Francois Langur	FL	370	155	84	54.2	SD	H	FG	PR	-26.2	4.4	46.0	3.8	11.0	19.5	246.0	50-500
Primates	Cercopithecidae	*Pygathrix nemaeus*	Douc langur	DL	350	152	133	87.5	SD	H	FG	PR	-22.7	4.4	45.7	6.0	14.9	24.5	405.5	50-500
Primates	Hominidae	*Gorilla gorilla*	Western lowland Gorilla	GOR	177	90	44	48.9	ST	H	HG	PR	-24.4	2.9	42.3	4.3	13.0	22.7	331.8	50-500
				GORSD	296	96	80	83.3	SD	H	HG	PR	-25.9	2.7	44.5	3.1	13.2	20.4	303.9	50-500
Primates	Hominidae	*Pan paniscus*	Bonobo	BNO	91	66	33	50.0	SD	O	S	PR	-26.5	3.5	44.8	3.0	4.1	13.8	75.5	50-500
Primates	Hominidae	*Pan troglodytes*	Chimpanzee	CHIMP1	212	122	60	49.2	ST	O	S	PR	-22.3	3.9	19.0	1.8	12.4	20.0	281.4	50-500
				CHIMP12	87	56	35	62.5	ST	O	S	PR	-23.9	3.3	23.4	2.2	12.4	20.0	281.4	50-500
Primates	Hominidae	*Pongo pygmaeus abelii*	Sumatran Orangutan	orang1	277	147	71	48.3	ST	H	HG	PR	-27.0	3.3	39.1	2.9	11.0	18.4	232.8	50-500
				orang2	227	107	50	46.7	ST	H	HG	PR	N/A	N/A	N/A	N/A	11.0	18.4	232.8	50-500
Primates	Lemuridae	*Eulemur macaco macaco*	Black Lemur	BKLE	187	56	25	44.6	ST	O	S	PR	-26.0	2.9	42.9	1.8	N/A	N/A	N/A	N/A
Primates	Lemuridae	*Eulemur mongoz*	Mongoose Lemur	ML	283	114	58	50.9	ST	O	S	PR	-23.5	4.1	27.9	2.5	13.2	22.5	332.8	50-500
Primates	Lemuridae	*Lemur catta*	Ring-tailed Lemur	RT	193	91	57	62.6	ST	O	S	PR	-24.4	2.3	40.2	2.8	12.6	22.5	319.6	50-500
Primates	Pitheciidae	*Pithecia pithecia*	White-faced Saki	Saki	420	111	74	66.7	ST	O	S	PR	-23.4	5.2	37.7	5.3	2.7	5.5	24.3	0-50
Proboscidae	Elephantidae	*Elephas maximus*	Asiatic Elephant	AE1	163	123	77	62.6	ST	H	HG	PO	-28.8	1.8	44.4	1.7	35.0	59.0	2160.0	1500-3000
				AE2	168	134	99	73.9	ST	H	HG	PO	-28.6	1.9	44.8	1.5	35.0	59.0	2160.0	1500-3000
				AE3	97	83	45	54.2	ST	H	HG	PO	-29.1	0.7	46.5	1.3	35.0	59.0	2160.0	1500-3000
Proboscidae	Elephantidae	*Loxodonta africana*	African Elephant	AFEL	252	59	27	45.8	W	H	HG	PO	-25.1	7.2	44.3	1.1	N/A	N/A	N/A	N/A
				AFEL2	280	59	20	33.9	W	H	HG	PO	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
				AFEL3	212	72	51	70.8	W	H	HG	PO	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
				AFYEL	258	80	39	48.8	W	H	HG	PO	-25.9	5.7	45.5	1.5	N/A	N/A	N/A	N/A
Rodentia	Bathyergidae	*Heterocephalus glaber*	Naked Molerat	molerat	322	160	135	84.4	ST	H	HG	RO	-27.0	5.2	42.9	2.2	5.0	10.3	67.5	50-500
Rodentia	Caviidae	*Hydrochaeris hydrochaeris*	Capybara	CAP	160	108	92	85.2	ST	H	HG	RO	-27.2	4.5	24.7	1.6	27.5	41.0	1197.0	500-1500

Rodentia	Sciuridae	*Callosciurus prevosti*	Prevost's squirrel	SQ	227	55	40	72.7	ST	O	S	RO	-23.1	4.1	36.5	1.8	1.4	2.3	8.0	0-50
Xenarthra	Dasypodidae	*Tolypeutes matacus*	Southern three-banded Armadillo	arma	365	94	57	60.6	ST	C	S	CI	-19.0	6.4	28.5	3.1	3.8	14.1	73.0	50-500
Diprotodontia	Macropidae	*Macropus rufus*	Red Kangaroo	KO1	186	126	78	61.9	ST	H	FG	MA	-27.2	4.5	24.7	1.6	23.7	40.0	1012.7	500-1500
				KO2	103	96	68	70.8	ST	H	FG	MA	-26.3	1.4	28.1	0.9	23.7	40.0	1012.7	500-1500
Monotremata	Tachyglossidae	*Tachyglossidae aculeatus*	Short-beaked Echidna	ECH	394	68	56	82.4	ST	C	S	MO	-23.1	5.7	11.8	0.9	2.4	0.0	3.4	0-50
				TOTAL	**17760**	**7285**	**4289**													

PUBLISHED DATA

Order	Family	Genus/ species	Common name	Sample ID and sequence prefix	Number of sequences/ sample	Total OTUs 96%ID	Estimated number of OTUs unique to sample	Estimated %Unique OTUs	Provenance	Diet	Gut physiology	Order	Stable Isotopes $\partial^{13}C$	Stable Isotopes $\partial^{15}N$	%C	%N	Author. Year	Pubmed ID
Artiodactyla	Bovidae	*Bos taurus*	Dairy Cow (Holstein)	Cow1	77	68	28	41.2	W	H	FG	AR	N/A	N/A	N/A	N/A	Ozutsumi, 2003	16195605
				Cow2	67	59	22	37.3	W	H	FG	AR	N/A	N/A	N/A	N/A	Ozutsumi, 2003	16195605
				Cow3	67	54	23	42.6	W	H	FG	AR	N/A	N/A	N/A	N/A	Ozutsumi, 2003	16195605
Rodentia	Muridae	*Rattus norvegicus*	Norway Rat (Wistar)	Rat	69	44	39	88.6	W	O	HG	RO	N/A	N/A	N/A	N/A	Brooks, 2003	14663493
Primates	Hominidae	*Homo sapiens*	Human	RL116	57	30	14	46.7	W	O	S	PR	-25.7	4.9	47.8	4.4	Ley, 2006	17183309
				RL117	46	25	10	40.0	W	O	S	PR	-25.2	5.5	43.1	4.4	Ley, 2006	17183309
				RL387	173	83	25	30.1	W	O	S	PR	N/A	N/A	N/A	N/A	Ley, 2006	17183309
				RL388	163	92	39	42.4	W	O	S	PR	N/A	N/A	N/A	N/A	Ley, 2006	17183309
				HumAdB	59	51	9	17.6	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2002	12363017
				HumAdO	54	44	6	13.6	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2002	12363017
				HumSuau	55	52	11	21.2	W	O	S	PR	N/A	N/A	N/A	N/A	Suau, 1999	10543789
				HumAdS	45	40	13	32.5	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2002	12363017
				HumVeg	40	37	18	48.6	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2002	12597356
				HumOldA	36	37	7	18.9	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2003	14524616
				HumOldC	24	21	4	19.0	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2003	14524616
				HumOldB	21	21	8	38.1	W	O	S	PR	N/A	N/A	N/A	N/A	Hayashi, 2003	14524616
				HumEckA	1060	66	8	12.1	W	O	S	PR	N/A	N/A	N/A	N/A	Eckburg, 2005	15831718
				HumEckB	617	97	20	20.6	W	O	S	PR	N/A	N/A	N/A	N/A	Eckburg, 2005	15831718
				HumEckC	662	89	18	20.2	W	O	S	PR	N/A	N/A	N/A	N/A	Eckburg, 2005	15831718
				HumNag6	407	71	63	88.7	W	O	S	PR	N/A	N/A	N/A	N/A	Nagashima *et al.*, Biosci. Microflora. 25, 99-107 2006	
Primates	Hominidae	*Gorilla beringei*	Bwindi Gorilla	AFG	38	33	22	66.7	W	H	HG	PR	N/A	N/A	N/A	N/A	Frey, 2006	16672537
				TOTAL	**3837**	**1114**	**407**											
				GRAND TOTAL	**21597**	**8399**	**4696**											