Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat

M. Lisandra Zepeda Mendoza[a*,δ], Zijun Xiong[b,c,δ], Marina Escalera-Zamudio[d], Anne Kathrine Runge[a], Julien Thézé[e], Daniel Streicker[f], Hannah K. Frank[g], Elizabeth Loza-Rubio[h], Shengmao Liu[b], Oliver A. Ryder[i], Jose Alfredo Samaniego Castruita[a], Aris Katzourakis[e], George Pacheco[a], Blanca Taboada[j], Ulrike Löber[d], Oliver G. Pybus[e], Yang Li[b], Edith Rojas-Anaya[h], Kristine Bohmann[a], Aldo Carmona Baez[a,k], Carlos F. Arias[j], Shiping Liu[b], Alex D. Greenwood[d,l], Mads F. Bertelsen[m], Nicole E. White[n,o], Michael Bunce[n,o], Guojie Zhang[b,c,p], Thomas Sicheritz-Pontén[q*], M. Thomas P. Gilbert[a,o,r*].

a. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark.
b. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China.
c. China National GeneBank, BGI-Shenzhen, Shenzhen, 518083, China.
d. Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany.
e. Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom.
f. Institute of Biodiversity, Animal Health and Comparative Medicine & MRC-University of Glasgow Centre for Virus Research, University of Glasgow, G12 8QQ, Glasgow, United Kingdom.
g. Department of Biology, Stanford University, 371 Serra Mall, 94305-5020 Stanford CA, USA.
h. Centro Nacional de Investigación Disciplinaria en Microbiología Animal-INIFAP, Km 15.5, Blvrd. Reforma, Zedec Sta. Fé, 01219 Ciudad de México, D.F., Mexico.
i. San Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road. Escondido, 92027 CA, USA.
j. Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Chamilpa 62210, Cuernavaca Morelos, Mexico.
k. Undergraduate Program for Genomic Sciences, Center for Genomic Sciences, National Autonomous University of Mexico, Av. Universidad s/n Col. Chamilpa 62210, Cuernavaca, Morelos, Mexico.
l. Department of Veterinary Medicine, Freie Universität Berlin, 14163 Berlin, Germany.
m. Center for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, 2000 Frederiksberg, Denmark.
n. Australian Wildlife Forensic Services, Department of Environment and Agriculture, Curtin University, 6102 Perth, Australia.
o. Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, 6102 Perth, Australia.
p. Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, 2100 Copenhagen, Denmark.
q. Center for Biological Sequence Analysis, Department of Bio and Health Informatics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.
r. Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway.

*Correspondence should be addressed to M.L.Z.M. (lisandracady@gmail.com), T.S.-P. (thomas@cbs.dtu.dk) or M.T.P.G (tgilbert@snm.ku.dk).
[δ]These authors contributed equally to the work.

Supplementary methods

Genomic methodologies

Sequencing

We used a whole genome shotgun strategy to sequence the *Desmodus rotundus* (common vampire bat) genome from a sample collected by the NIH through the Catoctin Wildlife and Zoo in Thurmont, Maryland, USA. The genomic DNA was extracted at the Laboratory of Genomic Diversity (LGD). The extracted DNA was then fragmented to 2-10 kb. Afterwards sequencing libraries were constructed with multiple insert sizes (ranging from 170 bp to 10 kb) according to the Illumina protocol for the final sequencing on Illumina Hiseq2000 following the manufacturer's instructions. For short insert sizes, the DNA was fragmented to the desired insert size, end-repaired and ligated to the Illumina paired-end adaptors. Ligated fragments were then size selected, purified and PCR amplified. The long insert size mate-pair libraries were constructed by circularizing the DNA, then digesting the linear DNA. Subsequently, the circularized DNA was fragmented, purified and biotinylated, to then perform adapter ligation. Sequence reads were generated from the output of Illumina data processing pipeline. We sequenced with a read length of 49 bp for the long insert size libraries (including 2 kb, 5 kb and 10 kb), and 100 bp for the short insert size libraries (including 170 bp, 500 bp and 800 bp).

Processing of raw reads

Before assembly, we corrected the sequencing errors based on *k*-mer frequency methodology and applied the following criteria to filter low-quality reads using SOAPfilter[1]:

1) Sequence reads were removed if >10% of the bases were N's.
2) For short insert-size libraries (<2 kb), reads were removed if the quality score of >60% bases was less than seven. For large insert-size libraries (≥2 kb), reads were removed if the quality score of >80% bases was less than seven.
3) We removed the duplicate reads or identical reads, and removed the adaptor sequences.
4) Reads were removed if Read1 (one end of the paired-end read) and Read2 (the other end of paired-end read) were completely identical.
5) For short insert size paired-end sequences, the reads were removed if the overlapping length was ≥10 bp between the Read1 and Read2. These cleaned reads were used for the subsequent assembly.

Estimation of D. rotundus genome size and genome survey

We used a 17-mer analysis to estimate the genome size of *D. rotundus* using Kmerfreq[2]. Reads were divided into sliding short sequences of 17 bp, overlapping by 16 bp, with the exception of the first base pair. The generated distribution of 17-mer followed a Poisson distribution. The genome size was estimated to be 2.13 Gb by dividing the total number of 17-mer by the peak of the distribution.

D. rotundus genome assembly

High quality reads were then assembled using SOAPdenovo[1]. Firstly, short insert library reads were assembled as initial contigs ignoring the sequence pair information. Secondly, all reads were aligned to the previously generated contig sequences. Scaffolds were constructed step by step, from short insert-size libraries to large insert-size libraries by weighting the paired-end relationships between pairs of contigs, where at least 3 read pairs were required to connect any two contigs. Finally, the gaps in the scaffolds were closed using the gap-filling module under SOAPdenovo (GapCloser)[1]. In order to assess the quality of the assembled genome, we downloaded from NCBI the publicly available *D. rotundus* transcriptome dataset[3]. A total of 9,057 transcripts were aligned to the genome assembly using BLAT[4] with default options (-minScore=30, -minIdentity=90, -tileSize=11). We required 90% of the sequence identity over at least 90% of its length to assign a match. Coverage percentage was calculated by dividing the total bases of transcripts that anchored the genome assembly by the total bases of transcripts, and we classified the data into three levels, >1 kb, >2 kb and >5 kb.

Protein-coding gene annotation

a) Homology based gene prediction

We used a homolog-based method to annotate the protein-coding sequences in the common vampire bat genome by using the Ensembl gene sets of *Myotis lucifugus*, *Pteropus alecto*, *Myotis davidii*, horse and human. We aligned the protein sequences of the reference gene sets to our genome assembly using tbastn[5] with E-value $\leq 1e^{-5}$, and linked the blast hits into candidate gene loci with genBlastA[6]. We filtered out those candidate loci with a query homologous block length <90%. We extracted the genomic sequences of candidate gene loci, including the intronic regions and the 3 kb upstream/downstream sequences. The sequences were passed to GeneWise[7] (wise2-2-0) to search

for accurately spliced alignments. The outcomes of Genewise included the predicted gene models. For single-exon genes, we filtered out pseudogenes containing more than one frame error. Potentially pseudogenized single exons were also removed if they were part of a multi-exon gene. We then aligned protein sequences of these genes against Uniprot databases using blastp (E-value $\leq 1e^{-5}$) and filtered out genes without matches. We also filtered out 9,573 genes that have more than 80% repeat regions, which may be transposon related genes.

b) De novo gene prediction

De novo gene prediction was performed using AUGUSTUS v2.5.5[8]. Prior to gene prediction, all TE-related repeats were masked as 'n'. To filter out TE-derived genes, we aligned the predicted protein-coding sequences to annotated TE sequences, and genes that aligned >50% of their length were filtered out. We trained the AUGUSTUS algorithm with common vampire bat transcripts from a published transcriptomic study[3]. We randomly selected 1000 Vampire bat genes with intact open reading frames (ORFs) and the highest GeneWise score from the homology-based gene set to train the Augustus gene prediction tool with default parameters. We filtered out partial (missing start or stop codon) and short (coding region <150 bp) predicted genes.

c) Building a non-redundant gene set

To build a non-redundant (nr) reference gene set, we integrated both homolog-based and *de novo* evidence, with the homology-based evidence prioritized over the *de novo* evidence. We used a stringent cut off for *de novo* genes. If *de novo* genes were chosen in the reference gene set, we only retained those with >30% of their length aligning when searched in Uniprot[9] and that contained at least 3 exons.

Gene function and repeat annotation

The integrated gene set was translated into amino acid (aa) sequences, which were used to search the InterPro database with iprscan v4.8[10]. We used BLAST to search the metabolic pathway database (release58) in KEGG[11] and homologs in the SwissProt and TrEMBL databases in UniProt (release-2012_03).

We performed repeat annotation on the assembled *D. rotundus* genome, as well as on the genomes of *Pteronotus parnelli*, *Megaderma lyra*, and *Pteropus vampyrus*. For each genome, we first

identified known transposable elements (TEs) using RepeatMasker[12] v3.3.0 and RepeatProteinMask against RepBase v16.10[13] TE library. We used RepeatScount, PILER-DF and RepeatModeler-1.0.5[12,14], to construct a *de novo* TE library. The *de novo* TE library was then used by RepeatMasker (-nolow -no_is -norna -parallel 1) to predict repeats. We predicted tandem repeats using TRF v4.04[15] (Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50). LTR_Finder software[16] was used to detect long terminal repeats (LTRs). The Repbase-based annotations and the *de novo* annotations were merged into a union set.

In order to examine the significance of the presence of MULe-MuDR elements, we identified the genes whose annotation coordinates overlap with the annotation coordinates of the MULe-MuDR elements and performed GO enrichment analyses on those genes using GOrilla[17].

Endogenous viral elements search

Non-retroviral endogenous viral elements (EVEs)

1) Genome screening

An in-house pipeline for *in silico* analyses was developed to search for EVEs in the *D. rotundus* genome. We first constructed a comprehensive library of all non-retroviral virus protein sequences available in public databases (GenBank and EMBL), including proteins from small RNA and DNA viruses, as well as large dsDNA viruses. This library was used as a target to perform blastx-like searches[18] using DIAMOND[19] to screen for *D. rotundus* genome sequences exhibiting similarity to virus sequences. Then, we extracted the matching aa sequences resulting from the blastx-like alignments for each sequence of the subset and performed reciprocal blastp-like searches[18] with DIAMOND[19] against the whole set of nr protein sequences of the National Center for Biotechnology Information (NCBI) database. The *D. rotundus* genome sequences were considered of viral origin if they unambiguously matched viral proteins in the reciprocal best hits (E-value $\leq$0.001).

From these sequences, putative viral open reading frames were inferred through a combination of automated alignments, using the exonerate program[20] and manual editing, based on the most closely related exogenous viral sequences in the nr protein database. For each resulting putative *D. rotundus* viral peptide, we retrieved the function and predicted the taxonomic assignation by comparison to the best reciprocal blastp-like hit.

2) Phylogenetic analyses

Multiple sequence alignments were performed using MAFFT[21] and manual curation using AliView[22]. Multiple alignments were performed for each inferred *D. rotundus* EVE peptide, including closely related exogenous viral proteins resulting from the reciprocal blastp-like analysis (with a blast cut-off of 1.0 e-6), closely related proteins of reference virus species recognized by the International Committee on Taxonomy of Viruses (ICTV[23]), and EVEs found in the genome of *Pteronotus parnellii,* the closest sequenced bat relative to *D. rotundus*. Maximum likelihood (ML) inferences were performed on each multiple sequence aa alignment using RAxML[24] with the substitution model WAG + G + I. The substitution model and parameters were selected using ProtTest3[25]. Support for nodes in the ML trees were obtained from 100 non-parametric bootstrap iterations, and the root of the ML trees was determined by midpoint rooting.

3) Experimental validation

In order to confirm the presence of Avian Bornavirus, Avian Dependoparvorius, Parvovirus, and Polyomavirus, we constructed double indexed Illumina libraries[26] on DNA derived from the spleen tissue of four different common vampire bats. Libraries were pooled and sequenced using 2X150 bp paired-end chemistry using the Illumina platform NextSeq 500 with the v2 kit on high output mode at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv). The raw paired-end reads were quality assessed, merged and filtered by mapping against the bacterial, human and chiropteran reference genome database with SMALT v0.7.5, and a stringency of 60% (https://www.sanger.ac.uk/resources/software/smalt/). Resulting reads with a length >150 pb were then run through a viral assignment pipeline[27] using blastx v2.2.29[28] to search against the viral nr protein database (NCBI 2016). Viral-matching reads were assigned to nodes within a tree structure of the NCBI taxonomy database with MEGAN v5[29] (LCA algorithm=60, max. matches=100, min. core=50, top percent=10, min. support=20, min. complexitiy=0.3). For each library, assigned reads matching viral sequences were manually verified by reciprocal blastx analysis. Positive viral hits were defined by the following criteria: at least two different reads matching two different proteins or two different regions within the same viral protein.

Endogenous retroviral elements (ERVs)

1) RepeatMasker

In order to detect ERV candidates by identifying flanking LTR regions, we used RepeatMasker v.4.0.6[30] with default parameters using RMBLAST (NCBI BLAST v2.3.0+ with RepeatMasker extensions, http://www.repeatmasker.org/RMBlast.html), Tandem Repeats Finder v4.09[15] and RepBase[31] (release 20150807) to the common vampire bat genome assembly.

2) Blast RefSeq Retrovirus

A database search using blastn and tblastx (v2.2.29+) was performed using the 66 retro-viral reference genome sequences from the RefSeq database of NCBI as in May 2016 against the *D. rotundus* genome. To filter for significant retro-viral signals an E-value $\leq 1e^{-5}$ was used and reads matching each of the RefSeq sequences were extracted. The 1,485 hits assigned to retroviral-like sequences were sorted by GI, collapsed by ID and manually verified by reciprocal blastn.

3) Validation of putative ERV identifications

To validate the detected novel ERV sequences, tblastx of the putative ERV sequences was re-run restricting the search to the *Retroviridae* (taxid:11632). Sequences giving true hits were selected under the following criteria: >55% percentage identity, alignment length >100 aa and an E-value <1e-10. The retrieved ERV sequences were further mapped against the *D. rotundus* genome using SMALT v0.7.6 under a stringency of ≥70%. The presence of DrERV and DrgERV within other available bat genomes (*Miniopterus natalensis*, *Eptesicus fuscus*, *Myotis lucifugus*, *Myotis brandtii*, *Myotis davidii*, *Pteropus alecto*, *Rousettus aegyptiacus,* and *Pteronotus parnellii*) was further verified. The full genomic sequences of DrERV (NC_027117.1) and DrgERV were run through tblastx against the Whole Genome Shotgun (WGS) database, using the standard parameters and restricting the search to Chiroptera (taxid:9397). Selected genomic contigs giving an alignment score ≥200 were further analyzed by reciprocal blastx against the non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF database restricting the search to the *Retroviridae* (taxid:11632), or against the DrgERV sequences.

4) Phylogenetic analysis

The DrgERV-like *Pol* and *Gag* sequences were retrieved and aligned to gammaERVs reference sequences using MUSCLE implemented in SeaView[32]. For *Gag*, a total of 49 sequences with length of 151-368 aa were used, including 12 DrgERV-like sequences. For *Pol*, a total of 50 sequences with length of 578-1082 aa were used, including 14 DrgERV sequences. The best-fit aa substitution

model for each alignment was identified using jModelTest2[33] (LG and empirical residue frequencies +F, with among-site rate heterogeneity modeled by the Γ distribution with four rate categories)[34,35]. Trees were inferred under ML criteria with the LG+I+G aa substitution model and node robustness assessed by the Shimodaira–Hasegawa-like approximate likelihood ratio test[36] using RAxML[24].

Dovetail improvement to genome contiguity

We improved the contiguity of the scaffolds using the Dovetail technologies. To this end, we first prepared two high-quality DNA Chicago libraries[37], which are based on *in vitro* reconstitution of chromatin. To obtain the minimum 5 ug of high molecular weight DNA for each of the two Chicago libraries, we used cultured cells from the San Diego Zoo collection, that were originally derived from a skin sample taken from between the shoulder blades of a *D. rotundus* individual.

DNA was extracted with Qiagen Blood and Cell Midi kits according to the manufacturer's instructions. Briefly, the cultured *D. rotundus* cells were lysed and centrifuged to isolate the nuclei. The nuclei were further digested with a combination of Proteinase K and RNAse A. The DNA was bound to a Qiagen genomic column, washed, eluted, precipitated in isopropanol, and pelleted by centrifugation. After drying, the pellet was re-suspended in 200 μL TE (Qiagen). The steps of chromatin assembly, biotinylation and restriction digestion, dNTP fill-in, ligation, exonuclease digestion, shearing, and library prep specific for the Chicago libraries were performed as described in [37]. Briefly, 500 ng of the extracted, naked genomic DNA of each library was reconstituted into chromatin *in vitro*, and fixed with formaldehyde. Fixed chromatin was then digested with *Mbo*I, the 5' overhangs were filled in with biotinylated nucleotides, and free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from the protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was sheared to ~350 bp mean fragment size, and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were then isolated using streptavidin beads before PCR enrichment of the library.

To improve the genome assembly contiguity of our initial *D. rotundus* assembly, the shotgun sequencing data, and the Chicago libraries sequences (916 M read pairs from Illumina HiSeq 2500 2X100 bp rapid run) were used by Dovetail Genomics as input data for HiRise, a software pipeline

designed specifically for using Chicago library sequence data to improve the assembly of genomes[37]. Briefly, this pipeline works by aligning the shotgun and Chicago library sequences to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The distance between the Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model, and the resulting likelihood model was used to identify putative misjoins and score prospective joins. After scaffolding, shotgun sequences were used to close gaps between contigs. Assembly contiguity statistics comparing the different assemblies we generated were obtained with the assemblathon perl script assemblathon_stats.pl with –n=1 and with an in-house script to calculate the N90.

Estimation of genome completeness

The BUSCO[38] software was used to assess the assembly and gene annotation qualities by examination of the coverage of highly conserved genes. In BUSCO pipeline, we chose the vertebrata taxonomic group that contained 3,023 universal single-copy orthologs as query gene sets. The *D. rotundus* genome was assessed and compared to those of *Erinaceus europaeus*, *Rhinolophus ferrumequinum*, *Myotis brandtii*, *Myotis davidii*, *Myotis lucifugus*, and *Pteronotus parnellii.* The result indicated that 65.3% of the excepted vertebrata genes were presented in the common vampire bat assembly (64.21% complete genes and 1.09% fragmented genes), while 12.64% were considered missing.

As another means to measure the correctness of the final genome assembly, the high-quality reads of short and long insert size libraries that satisfied our filtering criteria were aligned onto the assembly using BWA with parameters of "-t 4 -e -1". A total of 97.5% short insert size reads and 83.9% long insert reads could be mapped the assembly (Supplementary Information File 1). Moreover 95.15% of the assembly had more than 30-read coverage (Supplementary Fig. 3). We analyzed the paired-end information in the mapped reads and found that more than 99.46% and 83.19% of paired-end reads were properly mapped to the genome with an insert size according to the built short and long insert size libraries, respectively. When removing the paired-ends affected by the cyclization step during the long insert-size library building, the proportion increased to 91.01%.

Comparative genomics methodologies

We undertook comparative analyses of the *D. rotundus* genome against a range of other mammalian species (Supplementary Table 2). These included publicly available bat reference genomes encompassing 6 different families: *Eidolon helvum* (straw-coloured fruit bat)[39], *Pteropus alecto* (black flying fox)[40], *Pteropus vampyrus* (large flying fox)[41], *Megaderma lyra* (greater false vampire bat)[39], *Rhinolophus ferrumequinum* (greater horseshoe bat)[39], *Pteronotus parnellii* (Parnell's mustached bat)[39], *Myotis brandtii* (Brandt's bat)[42], *Myotis davidii* (David's myotis)[40], and *Myotis lucifugus* (little brown bat)[41]. As outgroups, we used *Erinaceus europaeus* (European hedgehog), *Equus caballus* (horse), *Bos taurus* (cow), and *Homo sapiens* (human).

Repeat annotation comparison

It has been previously shown that insectivorous bats have continuously acquired new DNA elements via horizontal transfer (HT), implying that predation on a large quantity of insects might increase bat exposure to HT[43]. Given the predation of the common vampire bat on mammals, we hypothesized that there may exist an elevated level of mammalian-derived specific TEs in the *D. rotundus* genome in comparison to the genomes of other non-sanguivorous bats species. In order to examine this, we compared the repeat annotations from *D. rotundus*, *P. parnellii* (sanguivorous and insectivorous, respectively; from the Yangochiroptera suborder), *M. lyra*, and *P. vampyrus* (carnivorous and fruigivorous, respectively; from the Yinpterochiroptera suborder). Comparison of their abundance was performed using t and Wilcoxon tests as implemented in R[44].

Orthologous gene families

We performed clustering of orthologous genes using two strategies. The first method identifies single-copy orthologs in the species using the TreeFam method[45,46]. In order to increase the number of clusters, we used a second approach which identifies 1:1 orthologs by building pair-wise orthologs between *D. rotundus* and the other species and uses a reciprocal best hit (RBH) plus synteny approach. In the Treefam method, we used the following ten genomes: *E. helvum, P. alecto, M. lyra, R. ferrumequinum, P. parnellii, D. rotundus, M. lucifugus, M. brandtii, M. davidii*, and the outgroup *E. europaeus*.

As the genome annotations of *E. helvum*, *M. lyra*, *P. parnellii* and *R. ferrumequinum* contain less annotated genes compared to the other bats and the length of the coding sequences of many genes is very short (<100bp), we re-annotated them using a homology-based method. We downloaded the gene sets of *M. davidii* from GigaDB and horse from Ensembl database. We first aligned these homologous protein sequences to the vampire bat genome assembly using tblastn with an E-value cutoff of 1e-5, and linked the blast hits into candidate gene loci with GenBlastA. We then extracted genomic sequences of candidate loci, together with 3 kb flanking sequences, using GeneWise to determine the gene models. Finally, we filtered pseudogenes that had only one exon with frame errors, as these loci were probably derived from retrotransposition. The non-redundant gene set was retained by clustering the overlapping genes. These annotations were used for building the orthologous gene families. The longest transcript of each gene from each species was used in an all-vs-all blastp with an E-value cutoff of $1e^{-7}$.

The 1:1 approach is able to deal with the many-to-many gene orthology relationships, and was performed as previously described in [47]. Briefly, the steps involved are the following:

1) An all-vs-all blastp (E-value $\leq 1e^{-5}$) was performed with the protein catalogues from *E. helvum*, *P. alecto*, *M. lyra*, *R. ferrumequinum*, *P. parnellii*, *D. rotundus*, *M. lucifugus*, *M. brandtii*, *M. davidii*, *E. europaeus* and *E. caballus*. The local alignments were then combined with Solar. Afterwards, we identified RBH orthologs in all aligned gene pairs.
2) We placed best-hit gene pairs on their annotated coordinates from the GFF files of each species according to *D. rotundus*. One best-hit gene pair (A1A2; 1 and 2 denote *D. rotundus* and any other species, respectiveely) and its nearest best-hit gene pair (B1B2) were considered to have syntenic evidence if they met the following criteria: *i*) genes A1 and B1 are on same chromosome or scaffold; *ii*) genes A2 and B2 are on a same chromosome or scaffold; *iii*) the number of genes between A1 and B1<5; iv) the number of genes between A2 and B2<5. We also retained a best-hit gene pair if one of their scaffolds had only one gene.
3) We built pair-wise orthologs between the *D. rotundus* gene set and the gene set of each of the other species by retaining orthologs supported by protein similarity (step 1) and gene synteny (step 2). We then constructed the clusters of orthologous genes with all the species by merging the pair-wise orthologs according to the gene set from *D. rotundus*. The protein and CDS

sequences of each cluster were obtained, aligned with PRANK, and cleaned with Gblocks as in the first method.

dN/dS analyses

In order to analyze the ratio of non-synonymous to synonymous substitutions (*dN/dS*) using PAML codeml[48], we used the cleaned CDS alignments from the two sets of orthologous families built using all the stated genomes, and the following corresponding phylogenetic tree drawn from the Chiroptera phylogenetic relationships as reported in [49,50]. Given that fragmented assembly and incompleteness of the protein-coding genes of the genomes of *E. helvum*, *M. lyra*, *P. parnellii* and *R. ferrumequinum*, we removed them from the built gene families and performed codeml of the remaining species using hedgehog as outgroup.

In order to identify genes with accelerated evolution in the *D. rotundus* lineage, we ran the two-ratio branch model using the next parameters: model=2, NSsites=0, fix_omega=0. As null model, we used the one-ratio model with model=0, NSsites=0, fix_omega=0. Using the results from the two and one-ratio models, we performed likelihood ratio tests (LRT) to identify genes with significant p value (*P*). *P* was computed assuming a null distribution of 50:50 mixture of a chi-square distribution with 1 degree of freedom and a point mass at zero. In order to adjust for multiple testing, we used the False Discovery Rate (FDR) method using the R package "qvalue". On the set from the single-copy orthologous families we used as cutoff values $P<0.01$ and FDR 10%. On the set from the 1:1 orthologous families we used a cutoff value of FDR 10%.

In order to identify genes with positively selected sites (PPS) in *D. rotundus*, we also used a branch-site with PAML codeml using model=2, NSsites=2, fix_kappa=0, fix_omega =0. For the null model we used model=2, NSsites=2, fix_kappa=0, fix_omega=1, omega=1. LRT and FDR were computed as for the branch model tests using the same cutoffs.

Phylogenetic reconstruction

CDS sequences from each single-copy family identified with the Treefam methodology were aligned guided by MUSCLE alignments of protein sequences and concatenated. The sequence of the concatenated genes was used to build a phylogenetic treewith PhyML under HKY85+gamma

for nucleotide sequences and WAG+gamma model for protein sequences. We performed 1,000 rapid bootstraps.

Divergence time estimation

To estimate the divergence time, we extracted the fourfold degenerate synonymous sites of the third codons sites from the single-copy genes and implemented the PAML mcmctree v4.4 package to estimate divergence time with the approximate likelihood calculation method. Calibration-times were obtained from the TimeTree database[51]. The MCMC process of PAML mcmctree was run to sample 1,000,000 times, with sample frequency set to 2 after a burn-in of 10,000 iterations. "Finetune" parameters were set as "0.05, 0.1, 0.12, 0.1, 0.3".

Gene family expansion/contraction

We identified gene families under expansion or contraction in *D. rotundus* with CAFE[52]. In CAFE, a random birth and death model is proposed to study gene gain and loss in gene families across a dated phylogenetic tree. We used CAFE with the single-copy orthologous gene families and our divergence time estimation. CAFE was run using -p 0.05, -r 10000, and lambda -s. We identified gene families with $P<0.05$ on the *P. parnelli-D. rotundus*, Phyllostomidae-Vespertilionidae, and Microchiroptera-Pteropodidae splits. We further filtered out those significant families in which *D. rotundus* and *P. parnelli* did not contain the gene.

Gene loss

Using the human gene catalogue as reference, and *B. taurus*, *E. caballus*, and *E. europaeus* as outgroups, we identified genes putatively lost in *D. rotundus* based on blast analyses with orthologs as previously described in [53]. The outgroups are used to distinguish the cases of genes that are putatively lost in the target species when they are specific gene gains on the reference species. We chose the human genes as reference becasue the human genes are better annotated than in any other mammalian species, with >90% of the annotated human genes having intact ORF and most of them having functional annotations. We first downloaded the human gene set from Ensembl (release-73) and obtained a nr human gene catalogue for further analysis by collapsing redundant genes and keeping the longest ORF.

In order to detect the putatively lost genes, we first mapped the filtered human proteins to the cow, horse, hedgehog, and *D. rotundus* genomes with tblastn (E-value $\leq 1e^{-2}$). Then, we determined the gene structure of each potential locus using GeneWise v2-2-0. Loci that mapped to multiple human genes were collapsed by keeping the query with the highest GeneWise score. Subsequently, we defined a gene as putatively lost in *D. rotundus* if it had: *i*) no predicted record, *ii*) a predicted record with an aligning rate <30% relative to the query human gene, or *iii*) frameshifts or premature stop codons. Afterwards, the genes identified as missing in *D. rotundus* were searched with blastp (E-value<0.001) against the protein catalogues of the other available bats. They were defined as absent using the same criteria as before.

We further corroborated the loss of genes that we deemed relevant to traits related to adaptation to sanguivory (taste and olfactory genes, PGA5, CTSG, AQP8, CCL2, KLRB1, PMP2, CNGB3, UGT2B17, RNLS, and HBB) with the following analyses.

1) Conservation of synteny

The identification of genomic regions exhibiting conserved synteny between certain species enables bioinformatics prediction of locations containing orthologous genes. Thus, a lack of *in silico* identification of a gene for *D. rotundus*, within a region with conserved synteny for other bat species, provides supporting evidence for the gene loss in *D. rotundus*. Thus, we undertook synteny analysis for each of the above-mentioned genes. To this end, we first performed pairwise whole genome alignments for the horse and the bat genomes used for the comparative genomic analyses with Lastz[54] (setting the parameters T=2, C=2, H=2000, Y=3400, L=6000, K=2200). Next, we joined the alignments using CHAINNET from the UCSC kintUtils (https://github.com/ENCODE-DCC/kentUtils) in order to define the genomic synteny blocks.

2) Transcriptome search

We searched the putatively lost genes against the published common vampire bat transcriptome which we used for homology annotation[3] using blastn with $E = 1e^{-5}$.

3) GC content

It has been shown that in some genomes, a high GC content can lead to missannotation of genes[55,56]. Thus, we analyzed the GC content of the regions surrounding the putatively lost genes. We applied a 500 bp non-overlapping sliding window across the common vampire bat genome and calculated the GC content for each window. We then compared the GC content of syntenic regions

around the putatively lost genes between common vampire bat and other bats to the common vampire bat genome average GC content.

4) PCR identification

We designed PCR primers (Supplementary Information 7) to attempt amplification of fragments of the putatively missing genes by searching for the genes reported from horse, human, and bats in NCBI and then aligning the sequences to design the primers based on them. We performed the PCR in the following species from which tissue samples exist within the collections of the Natural History Museum of Denmark and Copenhagen Zoo: *D. rotundus*, *R. aegyptiacus*, *Rhinolophus euryale* (insectivorous), *Miniopterus schreibersii* (insectivorous), *Nyctalus leisleri* (insectivorous), and *Nyctalus lasiopterus* (insectivorous, seasonal carnivorous). We isolated genomic DNA from the samples using the *KingFisher™ Duo Prime Purification System* (Thermo Scientific™) following the *KingFisher™ Cell and Tissue DNA Kit* (Thermo Scientific™) protocol with no considerable modification. PCR reactions were performed using the *AmpliTaq Gold™ DNA Polymerase* system (Thermo Scientific™) following the manufacturer's protocol. 32 PCR cycles were performed in an ordinary thermocycler with the following setup: Initial denaturation for 10 min at 95°C, denaturation for 15 sec at 95°C, annealing for 30 sec at a temperature depending on specific primer T_m, extension depending on specific fragment size (1 min/kb), final extension for 5 min at 72°C and hold at 4°C indefinitely. PCR products were visualized on 2% agarose gels ruled by the *O'GeneRuler 100 bp DNA Ladder* (ThermoFisher Scientific™). We cleaned up successfully amplified PCR products that were confirmed to have worked through the presence of a band of correct size using the *QIAquick PCR Purification Kit* (QIAGEN) following the standard protocol with a final elution of 21 µl of Elution Buffer. DNA concentrations were estimated for all PCR products after clean-up following the same method as before. Amplicons were then sent for Sanger sequencing for both forward and reverse directions using the commercial service provided by Macrogen (Seoul, Korea).

Functional characterization

In order to characterize the genes identified as significant from the selection, gene family expansion or contraction and gene loss analyses, we performed gene ontology (GO) analysis using GOrilla[17], as well as manual characterization using the GO annotations of the human genes downloaded from Ensembl (release-73) and literature mining. Furthermore, we used PROVEAN[57] to characterize the functional impact of the non-synonymous substitutions identified only in the proteins of *D.*

rotundus, and examined those previously detected to be under selection and with a PROVEAN score ≤-2.5.

Afterwards, the pooled significant genes of all the analyses were characterized for their functional biological relevance. We generated a manually curated list of the genes grouped into traits directly and indirectly related to sanguivory. Directly related traits include genes involved in: bile, lipid, digestion, feeding, taste, vitamin, iron, glucose, osmosis and homeostasis (levels), liver, and urea. The indirectly related traits include genes involved in: response to starvation, platelet, vasoconstriction, pancreas, endocrine system, energy storage, olfaction, infrared sensing, and immunity.

Protein modeling

After examining the biological relevance of the genes with a significant *dN/dS* ratio as well as a significant PROVEAN score in sites annotated as relevant in Uniprot, we performed protein modeling on the free fatty acid receptor 1 protein (FFAR1), regenerating islet-derived protein 4 (REG4), ribonuclease A family member 7 (RNAS7) and taste receptor type 2 member 3 protein (TA2R3). We tested for positive selection and PPSs as follows:

1) Complete CDS alignments for the FFAR1, PLXNA4, REG4, RNAS7 and TA2R3 sequences of over 25 representative mammalian species (Afrotheria, Xenarthra, Primates Rodentia, Chiroptera, Artiodactyla, Perissodactyla, Carnivora, Soricomorpha and Scandentia orders) were downloaded from the OrthoMaM database[58]. In parallel, the previously obtained bat sequence alignments were added to each dataset.
2) All sequences were re-aligned based on their protein translation using MUSCLE[59] implemented in SeaView[32] and phylogenetic analysis based on the nucleotide sequences was performed under ML criteria using PhyML 3.0 [36]. Three tests were performed for each dataset:
    a. M1/M2 branch model: The effect of positive selection acting upon the *D. rotundus* branch compared to the rest of the bat lineage and to the rest of the mammalian tree was tested under the M1/M2 branch model with CODEML in PAML v4 constraining the *D. rotundus* node[48]. Contrasting to the M0/M2 models, the M1/M2 model comparison allows discerning between selection and relaxation of selective pressure in coding sequences. The M1 (model=2, NSsites=0) model assumes for neutral evolution

(fix_omega=1, omega=1) assuming identical ω ratios among all branches, while the M2 model (model=2, NSsites=0) constraints a specific node and allows for positive selection (fix_omega=0, omega=1) to act on the selected branches. Models were evaluated under a LRT using a $\chi 2$ distribution with the number of degrees of freedom (d.f.) obtained from the number of parameters used and testing under a $P<0.05$.

b. M8a/M8 site model: For a conservative detection of PSSs within the *D. rotundus* sequence, datasets were tested under the M8a/M8 site model using CODEML in PAML v4[48]. The M8 model (model=0, NSsites=8) allows all sites to evolve under positive selection (fix_omega=0, omega=1), while the M8a model uses the same parameters (model=0, NSsites=8) but with ω fixed to 1 (fix_omega=1, omega=1)[48]. Models were compared under a LRT as described above and PSS were scored under Naive Empirical Bayes (NEB) and Bayes empirical Bayes (BEB) with a $P \geq 95\%$.

c. Branch-site model A (BSA): To further test for sites unique to the *D. rotundus* node evolving under episodic positive selection, we evaluated the bat sequence alignment datasets constraining the common vampire bat node under the BSA using CODEML in PAML v4. The BSA (model=2, NSsites=2) estimates ω values upon sites and specific branches, classifying sites into four different categories: class 0 (purifying selection on all branches), class 1 (neutral evolution on all branches), class 2a (positive selection in selected branches and purifying selection for the rest of the tree) and class 2b (purifying selection on the selected branch and neutral evolution for the rest of the tree)[48]. Thus, it is sensitive to lineage-specific PSS that evolved under positive selection at some point in evolutionary history[60,61]. The BSA model was evaluated under a LRT against the null hypothesis (BSA with omega=1 and fix_omega=1), while PSSs were scored under NEB and BEB with a $P \geq 95\%$.

Next, we performed protein modeling by mapping the sites identified in the previous positive selection and PSSs analyses. The 3D proteins models of the *D. rotundus* FFAR1, REG4, RNAS7, PLXNA4 and TA2R3 protein sequences were constructed using Phyre2[62] based on profile hidden Markov model analysis to detect distant homologs and build high confidence 3D models (i.e. with a high probability that the match between the query and template is a true homology) based on low sequence identities (<15%). The following templates were used for modeling: the crystal structure of the G-protein-coupled rhodopsin receptor bound to arrestin for FFAR1 and TA2R3, the mouse

plexin a3 intracellular domain for PLXNA4, the Human RegIV Protein for REG4 and the human RNase 7 for RNAS7[63–66]. Sites previously identified as aa substitutions with impact on the biological function of the proteins (using PROVEAN), and detected to be evolving under positive selection (using PAML), were mapped onto the *D. rotundus* 3D models using PyMOL Molecular Graphics System Version 1.5.0.4 (Schrödinger, LLC).

Metagenomics methodologies

Sampling

We used fecal samples from bats of diverse feeding strategies in order to compare them to *D. rotundus*. Specifically, the examined diets are: frugivorous (*Rousettus aegyptiacus*), insectivorous (*Rhinolophus ferrumequinum*), and carnivorous (*Macroderma gigas*).

The *D. rotundus* fecal samples collected in Mexico, including one milk sample from a lactating *D. rotundus* female, were obtained in accordance with Mexican regulation NOM-062-ZOO-1999 under the following sample collection permit and export certificates: Num/SGPA/DGVS: 03173/14; SAGARPA: 241111524599811488A467371. Bats were captured using mist nets by personnel of the Livestock Protection and Promotion Committees in the States of Veracruz. Rectal swabs were obtained from every individual using Copan N°160C swabs, (approx. 2 mm diameter tip). Bats were anesthetized according with Mexican regulations (NOM-033ZOO 1995). Swabs were wet with MEM media and kept into a tube that contained 0.5 mL of RNAlater. Samples were stored at -70 ºC until use. The *D. rotundus* samples from Peru were obtained from individuals captured using mist nets and/or a harp trap placed outside of each roost between 18:00 and 06:00. Bats were stored individually in cloth bags until sampling and were released at the site of capture after application of a 4-digit incoloy wing bands (3.5 mm, Porzana Inc.). Fecal samples were collected opportunistically from cloth bags or during collection of the samples and preserved in RNAlater. The University of Georgia's Institutional Animal Care and Use Committee approved the protocols for the capture and handling of bats (AUP # A2009-10003-0) and the collection and exportation permits were granted by the Peruvian General Directorate of Forestry and Wildlife (RD-222-2009-AG-DGFFS-DGEFFS & RD-273-2012-AG-DGFFS-DGEFFS; 011989-AG-DGFFS). The access to the genetic resources of Peru was given by the permit RD-054-2016-

SERFOR-DGGSPFFS. The blood meal sample was obtained from bats that were captured after feeding using a non-lethal procedure[67]. Extracted blood (ca. 50μL) was expelled onto Whatman FTA cards and desiccated. The *D. rotundus* samples collected in Las Cruces Biological Reserve (8° 47' N, 82° 57' W, 1,100 m above sea level), Coto Brus, Costa Rica were approved by the Costa Rican permit office with permit number RT-044-2015-CONAGEBIO. Bats were caught in 12 m mist nets which were open from 6pm to 10 pm.

Only the fecal samples were used for the comparative analyses. Due to the small sample size of the milk sample (n=1), we excluded it from the results presented in the main manuscript, but present the conclusions from its analysis as Supplementary Note 8.

The *R. ferrumequinum* samples from Woodchester Mansion, UK were collected under license from Natural England (20122272) and the Home Office (PPL 3002513 and PIL 30/3261). Bat droppings were collected from *R. ferrumequinum* maternity roost in the attic of Woodchester Mansion, Gloucestershire, UK (51°2'N, 2°90'W). Droppings were collected during eight nights from mid-June to mid-September 2012. After foraging adult bats had left the roost, collectors entered the roost and laid out five paper collection plates, each with their own piece of urine-absorbing kitchen paper. Collection plates were placed where the largest volume of existing droppings was found. Droppings from all collection plates were combined, and collected in 25 ml tubes the following evening, and remained at ambient temperature, before freezing as soon as possible. The *M. gigas* samples from Pilbarra, Australia were collected by the consulting company Biologic in Pilbarra, Australia from free ranging bats. Samples were collected dry and frozen at -20°C after collection until analysis. Consultant surveys were performed under the regulation 17 from DPAW and the collection with number SF007770. The *R. aegyptiacus* fecal samples were collected from the Copenhagen Zoo. Samples were collected and frozen at -20°C after collection until analysis.

DNA extraction of microbiomes

All fecal samples and stomach contents were extracted using the PowerFecal DNA Isolation Kit (MoBio) with modifications to the manufacturer's instructions. Briefly, 750 μl of Bead Solution was added to each sample in a provided Dry Bead Tube and vortexed before adding 60 μl of Solution C1, preheated to 60°C. Samples were incubated at 65°C for 12 min before they were put in the Beadbeater for 2 minutes at 30 hz to complete lysis. The supernatant was collected, 250 μl of

Solution C2 was added and samples incubated for 5 minutes at 4°C. The supernatant was collected, 200 µl of Solution C3 was added and samples incubated for 5 minutes at 4°C. The supernatant was collected, 1200 µl of Solution C4 was added and the mixture was loaded onto the provided spin filter (650 µl at a time) to bind DNA. 500 µl of Solution C5 was used to wash. Elution was done with 100 µl of Solution C6, and collected after incubating at 37°C for 15 min. We extracted the DNA from the *D. rotundus* milk sample using a standard commercial kit (Qiagen), following manufacturer's instructions. DNA was kept at -20 °C for further use.

Anal swab samples were extracted using the BiOstic® Bacteremia DNA Isolation Kit (MoBio) following the manufacturer's protocol. Briefly, 450 µl Solution CB1 was added to each sample and incubated at 70°C for 15 min. Samples were vortexed for 10 minutes at 30 hz in the bead beater to complete lysis. The supernatant was collected, 100 µl Solution CB2 was added and samples were incubated for 5 min at room temperature. The supernatant was collected, 1 ml of Solution CB3 was added and the mixture loaded onto a provided spin filter (600 µl at a time). The spin filters were washed twice with 500 µl of Solution CB4. Elution was done with 50 µl of Solution CB5 and incubated at room temperature for 5 min before collection. All samples were kept at -20°C for further use. All DNA extracts were quantified using the Qubit dsDNA HS (High Sensitivity) Assay Kit (Life Technologies) following the manufacturer's protocol.

Metagenomic DNA sequencing

All extracts were fragmented on the Bioruptor (Diagenode) at 4°C for 4-18 cycles of 15 sec on 90 sec off prior to the library build. Samples were then run on the 2200 TapeStation (Agilent Technologies) following the manufacturer's protocol to make sure the DNA had been fragmented to the target size. All libraries were built using the NEBnext DNA Library Prep Mast Mix Set for 454 (New England BioLabs) following the manufacturer's instructions with some modifications. Briefly, for all but the common vampire bat samples, libraries were only built on extracts with a Qubit DNA concentration above 0.85 ng/µl. When possible libraries were built on approximately 180 ng of DNA. H_2O was added to the extract to reach 180 ng DNA in 42.5 µl if DNA concentrations were low. 42.5 µl of the extract was used regardless of DNA concentration.

5 µl End Repair Reaction Buffer (10x) and 2.5 µl End Repair Enzyme Mix were added to 42.5 µl of extract, and incubated for 20 min at 12°C and 15 min at 37°C in a Thermocycler. Purification was

done with Qiagen MinElute PCR Purification Kit, 300 µl PB Buffer to bind, 700 µl PE Buffer to wash, and 30.5 µl EB Buffer incubated for 15 min at 37°C to elute. For adapter ligation, 10 µl Quick Ligation Reaction Buffer (5x) and 5 µl Quick T4 DNA Ligase (5 U/µl) was added to 30 uL of the end-repaired DNA. Then, 5 µl of Blunt End Adaptor (20 µM) was added and incubated at 20 min at 20°C. Purification was done with Qiagen QIAquick PCR Purification Kit, 300 µl PB Buffer to bind, 700 µl PE Buffer to wash, and 43 µl EB Buffer incubated for 15 min at 37°C to elute. Adapter fill-in reaction was performed with 5 µl Adapter Fill-in Reaction Buffer and 3 µl Bst DNA polymerase added to 42 µl adapter-ligated DNA. Samples were incubated for 20 min at 65°C and 20 min at 80°C in a thermocycle.

To determine how many cycles each sample needed for amplification, 2 µl of a 1:100 dilution of each sample was added to a master mix containing 2.5 µl Platinum 10x Buffer, 1 µl $MgSO_4$, 0.5 µl dNTPs (10 mM), 0.5 µl inPE1.0 F primer (10 µM), 1 µl SYBR Green, 1 µl BSA (10 mg/ml), 1 µl R primer with index, 16.9 µl H_2O and 0.1 µl Platinum hifi Polymerase, for a 25 µl reaction. Samples were run on a Roche LightCycler 480 Instrument with the following settings: start 94°C for 30 sec, denatured at 94°C for 15 sec, annealed at 60°C for 30 sec and extended at 68°C for 40 sec. Denature, annealing and extension was repeated 45 cycles.

For PCR amplification, a master mix containing 5 µl Platinum 10x Buffer, 2 µl $MgSO_4$, 1 µl dNTPs (10 mM), 1 µl inPE1.0 F primer (10 µM), 1 µl BSA (10 mg/ml), 33.5 µl H_2O and 0.5 µl Platinum hifi Polymerase, was added to 5 µl library and 1 µl R primer with an index individual for each sample. Each library was amplified in 3 separate reactions to maximise library complexity. They were run with the following settings: start 94°C for 30 sec, denature at 94°C for 15 sec, anneal at 60°C for 30 sec and extend at 68°C for 40 sec, and a final extension at 68°C for 7 min. Denature, annealing and extension was repeated for the number of cycles determined by qPCR. Purification was done with Qiagen QIAquick PCR Purification Kit, 1000 µl PB buffer to bind, 700 µl to wash, and 30 µl EB buffer incubated for 15 min at 37°C to elute.

DNA concentration of the amplified libraries was tested using the Qubit dsDNA HS Assay Kit (Life Technologies) following the same procedure as previously mentioned. Samples were diluted to 1-2 ng/µl based on Qubit measurements and run on a High Sensitivity DNA chip on the 2100 Bioanalyzer (Agilent Technologies), following the manufacturer's protocol to determine insert size

and molarity in the 100-1000 bp region. Samples were then pooled based on index compatibility and 100 bp paired-end sequenced on the Illumina 2500 HiSeq platform.

Data processing

The reads were cleaned with Trimmomatic v0.32[68] to remove adapter sequences and cleaned of low quality bases with prinseq-lite v0.20.3[69] (-ns_max_p 10 -trim_qual_left 30 -trim_qual_right 20 -min_qual_mean 25 -min_len 25). Afterwards, in order to filter out non-bacterial reads of bat origin, the datasets were mapped against the closest bat available genome and only the non-mapping reads were kept. *M. gigas* was screened against the *M. lyra* assembly (GCA_000465345.1_ASM46534v1) and *R. aegyptiacus* against the *P. vampyrus* assembly (pteVam1.73).

Taxonomic and functional identification

The second improved implementation of the method "Chain-mapper" described in [70], called MGmapper, was used to map the filtered reads against the next databases in full mode: MetaHitAssembly[71], HumanMicrobiome[72], ResFinder[73], Plasmid, Virulence, GreenGenes[74], and Silva[75]. In order to further remove non-bacterial reads, we also mapped in chain mode to the next whole genome databases downloaded from GenBank in the given order: human, plant (plants and common plants), vertebrates (mammals, others, and common animals), invertebrates, protozoa, fungi, and virus. Lastly, the remaining non-mapping reads were mapped in chain mode to the full genome databases of bacteria and draft bacteria. Next, we kept the species identified with more than the 1st quartile (Qu) coverage from the coverage distribution of the corresponding database and filtered out those found on the corresponding extraction blanks.

The MGmapper results were also used to obtain rarefaction curves from each dataset using an in-house script. Using the reads mapping to the databases of interest and the unmapped reads we performed *de novo* assembly using IDBA_UD[76] and predicted genes using Prodigal[77]. Afterwards, we generated a nr gene catalogue with usearch[78] by clustering the predicted genes with 90% id and keeping the centroid sequences. Next, the nr gene catalogue was searched with ublast[78] against the Uniprot database[9] for functional and taxonomic annotation with the use of a customized python script. Finally, we used DIAMOND[19] v0.6.4 to search the unmapped reads from the *D. rotundus* datasets against the Uniprot database keeping only the best hit for functional and taxonomic

annotation. For assessment at a functional level, we annotated the Uniprot protein identifications and also converted the Uniprot ids to KEGG orthology (KO) and eggNOG ids.

Taxonomic and functional metagenomic comparison

We compared the fungal, protozoan, viral, and bacterial taxonomic assignations from the different bat species as follows:

1) First, we filtered potential misidentifications using different levels of stringency for the identifications obtained with MGmapper and from those identified by mapping the nr gene sets and the non-mapped reads against Uniprot.
    a. The taxa identified from the MGmapper databases were filtered as follows. First, we obtained the genomic breadth of coverage for each taxon identified with MGmapper. Then, for each dataset of each bat species, a threshold was defined. For the virus, Silva, and plasmid, the threshold is the 1st Qu of the distribution of breadths. For bacteria, it was 0.0001, and 0.00001 for the rest of the databases.
    b. As a second level of stringency, we removed the filtered identifications of the extraction blanks from their corresponding filtered samples of step a).
    c. The third type of taxonomic identification confidence comes from those identified by mapping the nr gene sets and the non-mapped reads against Uniprot. To filter these identifications, we first removed any non-microbial hit. Then any taxa in which the read pairs matched different genera or only one read had a hit was removed. Next, we rescaled the counts of each taxon by their percentage in the dataset and obtained a threshold (1st Qu) for each bat species dataset. Another level of stringency was added by subtracting the rescaled filtered counts of the extraction blanks from the corresponding samples.
2) Next, we identified a microbial taxonomic and functional sanguivorous core. To this end, we compared the filtered sets of the different bats and kept as core those taxa and genes identified only in the vampire bat samples. The taxonomic and functional cores were analyzed in two ways:
    a. Manual examination of the following two sources of information: 1) taxa from the strictest type of taxonomic identification filtering, and 2) genes from the annotated filtered nr gene set catalogue using the KEGG and COG annotations. In the manual

curation, we classified the entries of each of the different taxa groups (virus, bacteria, protozoa, fungi, etc.) and functions in the next categories relevant to the blood diet:

i. Directly related to the blood diet: coagulation, immune, iron, nutrient, fat, kidney, relevant, hormone.

ii. Indirectly related: salt, aromatic, colon, sulfur, methano, CO_2H_2, and oral.

iii. Others: Elements with industrial applications, pathogen, related to hematophagous insects.

b. An additional way to compare the core to all the samples consisted on analysing the number of genes of each enzyme type and investigating which have a statistically significant difference in abundance, either due to having been lost (upper tail $P \leq 0.05$) or kept (lower tail $P \leq 0.05$). Those in the upper tail (the genes of the enzyme type that are likely to be lost, that are removed from the core because they are present in the other bats with other diets) are hypothesized to derive from redundant enzymes, that do not have any relevant importance for the blood diet. And those enzyme types that lost the least number of genes are supposed to be more specialized for the blood diet. Thus, with a Wilcoxon rank sum test we examined: *i*) if the genes kept in the core with almost no change in abundance from all the samples together are in low or high abundance compared to all the samples together and *ii*) if the enzymes kept in the core with almost no change in abundance are from more "specific" pathways and the enzymes with high difference in the number from the core and all the samples together are more "generic", meaning that they come from more pathways. Furthermore, in order to investigate if there really are more enzymes of any category in the common vampire bat microbial core, than in all the samples together, not just because that category involves more enzymes, we defined a "pathway coverage" as the number of found unique enzymes/total enzymes in the pathway.

3) We compared the relative taxa and functions abundances between all the *D. rotundus* fecal samples and the non-sanguivorous bats as follows:

a. Comparing the distributions of the different functional categories with a Wilcoxon rank-sum test between the common vampire bat dataset and each of the non-sanguivorous bats.

b. Comparing the broad and fine functional hierarchical levels using the KEGG and COG annotations. Taxonomic comparisons were performed at the genus and the species level.

To compare the taxonomic and functional profiles, we used the entire dataset as well as down-sampled subsets in order to account for sampling and sequencing depth bias besides the scaling of the count values. Sampling values were the minimum, median and 3rd Qu values of the counts distributions. Comparisons were visualized with heat maps using the gplots package from R, PCAs with the prcomp and the shapes package from R to apply the GPA method (10 replicates were performed on each subsampling value), and dendrograms using the Euclidean, Bray-Curtis, and Jaccard distance metrics from the R package vegan[79], and the Ward hierarchical clustering method using UPGMA and ward.D.

c. We also compared the pathway coverage of the vampire microbial core to all the *D. rotundus* fecal samples.

d. In order to identify which taxa and functions contributed the most to the variation between the *D. rotundus* and the other bat species, we examined the rotation matrix from the PCA of the normalized count matrix excluding the four deepest sequenced samples (1 *D. rotundus* and 3 *R. aegyptiacus* samples) of the species and genera level and all the pathways together. We defined as pathways and taxa contributing to most of the variation those with an absolute value higher than the 3rd Qu of the values distribution for the first three principal components. Then, we identified the most significantly abundant *D. rotundus* species from those taxa as those with a significantly higher median normalized counts values ($P \leq 0.05$) in *D. rotundus* and a median and mean normalized count values of 0 in the other 3 bat species.

4) As another method to compare the abundances of between all the *D. rotundus* fecal samples and the non-sanguivorous bats, we first constructed and annotated with KEGG a nr gene set catalogue using all the assembled genes from all the bat samples (sanguivorous and non-sanguivorous) using usearch with 95% identity. Next, the reads of each sample were mapped against this bat nr gene set using bwa and a percentage normalized count matrix was formed. A Fisher test was performed with these counts for each of the functional pathways to test for enrichment in the common vampire bat samples.

Rectal swabs and fecal samples comparison

In order to test for the validity of the use of fecal samples as proxy for the gut microbiome characterization, we compared the microbial profile of the two rectal swab samples to that of the

fecal samples. To this end, we first constructed a *D. rotundus* only non-redundant gene set catalogue by using all the assembled genes from all the *D. rotundus* samples (rectal swabs and fecal) using usearch with 95% identity. Next, the reads of each sample were mapped against this *D. rotundus* nr gene set using bwa and a percentage normalized count matrix was formed. Euclidean and Bray-Curtis distances were calculated and clustered with the ward.D method to generate cladrograms and PCA was performed.

Supplementary notes

Supplementary note 1. Repeat annotation

Similar to the previously reported bat genomes of *P. alecto* and *M. davidii*[40], the annotated total content of TEs in the *D. rotundus* genome is 32% (Supplementary Information 1), with a mean divergence of 23.1%. Class II TEs mobilize through a cut-and-paste mechanism and their contribution to genome size is generally much smaller in model mammals (less than 3%)[80–83]. An elevated Class II TE activity has been shown in Verpertillionidae bats (3-5%)[84,85], although the analyzed phyllostomid bat *Artibeus lituratus* had less than 0.5%[84]. Interestingly, the genome of *D. rotundus* harbors a high content of Class II TE (4.06%), thus providing evidence against the hypothesis that vesper bats are unique within Chiroptera in their ability to tolerate and host Class II TEs[86]. Furthermore, members from families such as the Helitron, Cryptons, and MULE-MuDR were identified in the *D. rotundus* genome (Supplementary Table 3). Previous studies have shown that the Helitron superfamily is prevalent in vesper bats and that it has been active in the common ancestor of Vespertilionidae and Phyllostomidae, however no Helitron had been recovered from a Phyllostomid bat nor from a Minipteridae[84,86,87]. Mammalian genomes are protected from TE integration in the germline by piRNA mediated methylation[88,89]. Crypton TEs have been identified in pathogenic fungi and are ubiquitous in prokaryotes but rare in eukaryotes[90–92]. As those identified in *D. rotundus*, all of the few found in animals lack the YR catalytic site.

Supplementary note 2. Non-retroviral EVEs

In the characterization of non-retroviral EVEs, we obtained 14 *D. rotundus* sequences of unambiguous viral origin, ranging from 45 to 964 aa in length (average=298 aa) and showing 49–76% aa similarity (average=60%) to their most closely related exogenous viral protein sequences. The 14 EVEs were assigned to two different families (*Bornaviridae* and *Parvoviridae*) (Supplementary Fig. 4), representing two of the seven major types of viral genomes (-ssRNA, and ssDNA) (Supplementary Information 3).

The eight *D. rotundus* endogenous bornaviruses fall into two distinct lineages. Two EVEs (found in scaffold125) likely correspond to the same viral integration event and are closely related to the Reptile bornavirus 1. The six *D. rotundus* endogenous parvoviruses fall into two distinct lineages belonging to the *Dependovirus* genus. The first lineage is closely related to adeno-associated

viruses and in particular to the avian parvovirus. The second lineage seems to be related to a clade comprising the seal and chipmunk parvoviruses as well as human parvovirus B19. However, this EVE fragment is very divergent, making its position within the *Parvoviridae* phylogeny tenuous. The results from the experimental valiadation confirmed the presence of Parvo and Bornavirus-related EVEs in four independet individual vampire bats, suggesting that the detected EVEs are fixed in the *D. rotundus* population (Supplementary Information 3).

Supplementary note 3. ERV screening

RepeatMasker and genomic blastn analyses showed that *D. rotundus* has a strikingly low number of ERV-like sequences when compared to other bat species[93–95] (Supplementary Information 4). Although we found evidence for other putative ERV-like sequences by blastn, the only confirmed ERV-like sequences were shown to belong to the previously described low copy *D. rotundus* endogenous betaretrovirus (DrERV)[96] (Supplementary Information 4). Comparative analysis of DrERV within the genome of other seven bat species revealed the presence of sequences related to the DrERV LTRs in all bat species tested (coverage of up to 80% and a sequence identity of up to 81%; E-value=0.0), but no CDSs were found. Thus, DrERV is not present in other bat species and may be unique to the *Desmodontinae* family, particularly to *D. rotundus*. For other ERV-like sequences, only short regions likely corresponding to mammalian genes with a viral counterpart were identified, such as the *c-fos* protooncogene homologous to the Murine osteosarcoma virus gene *v-fos* and the non-receptor tyrosine kinase protooncogene 1 (*abl1*) homologous to the Abelson (P160) murine leukemia virus (Ab-MLV) *abl* gene (Supplementary Information 4). However, these viruses have often been found to be false positives due to over amplification in laboratories and reagent contamination[97]. Therefore, given the length of the sequences detected, we considered them as negative results.

Further tblastx revealed the presence of CDSs within the *D. rotundus* genome that are distantly related to the avian reticuloendotheliosis gammaretrovirus (REV)-like (Supplementary Information 4). Three proviral copies of this new retrovirus, named here DrgERV, were detected in scaffold11, C50515620 and scaffold1070 of the *D. rotundus* genome. The DrgERV copy in scaffold11 comprises partial *gag* and *pol* CDSs, while the copy in scaffold C50515620 has only one partial *pol* CDS. Finally, the copy in scaffold1070 is a full-genome provirus with partial or complete CDS for *gag*, *pol* and *env* genes (Supplementary Information 15). The percentage of identity between the

proviral copies (scaffold11 and scaffold1070) was of 70-75%, both in the *gag* and *pol* regions. Reciprocal blastn and blastp of the full DrgERV provirus sequence gave hits in the folowing bat species: *Miniopterus natalensis* (*gag* position: 1008821909_LDJU01000079_1:5770586-5769765, *pol* position: 1008820783_LDJU01000447_1: 799445-802987), *E. fuscus* (398484927_ALEH01156174_1: 3535-4401, 398531442_ALEH01109659_1:9459-6010), *P. vampyrus* (717420875_ABRP02219245_1: 3105-3971, 717420875_ABRP02219245_1:4014-7541), *M. lucifugus* (306726728 AAPE02020841_1: 42012-42875, 306747371_AAPE02005653_1: 3458-7000), *R. aegyptiacus* (1006975623_LOCP02000478_1: 1216084-1215209, 1006970057_LOCP02001762_1: 9255-11546), *P. alecto* (430300286_ALWS01048498_1: 9-890, 430197481_ALWS01091148_1:24637-22505), *P. parnellii* (540046096_AWGZ01020001_1:3020-2535, 539967890_AWGZ01098207_1: 1886-117), *M. lyra* (*gag* position: 538660299_AWHB01335377_1:2124-1321), *M. davidii* (430529539_ALWT01075223_1:1237-1872, 430712041_ALWT01008756_1:2759-5326), and *M. brandtii* (513250238_ANKR01266518_1: 6181-5366, 513263745_ANKR01253011_1:47025-43459) (up to a coverage of 62% and percentage identity of 57%, E-value=0.0). Hits were confirmed as belonging to the *M. lucifugus* endogenous gammaretrovirus group 2 (MLERV 2) family, suggesting a distant relatedness of DrgERV to this group of viruses. Consitent with these results, phylogenetic analysis (Supplementary Fig. 5) revealed that DrgERV is related with previously described MLERV2 bat gammaERVs[93], diverging from gammaretroviruses of avian origin. MLERV2 has been described as a high-copy EVR family in both *M. lucifugus* and *E. fuscus*[93]. However, in *D. rotundus* it is only present as a divergent and low-copy provirus. While DrERV seems to be species-restricted, DrgERV has several homologs within different bat species. The DrgERV-like ERVs are new family of gammaERVs that are the deepest branch within the gammaretrovirus phylogeny, supporting previous observations on the bat gammaERVs representing the most ancient lineages within the Gammaretrovirus genus[94,98].

Supplementary note 4. Genetic adaptations against retroviral integration in *D. rotundus*

We further analysed the gene TRIM5[99], identified under gene family expansion, as well as two other genes well documented to have antiretroviral functions, APOBEC3G[100] and BST2[101].We first corroborated the CDS of the expanded TRIM5 family. All the seven copies are located within the same scaffold (ScWXtkA_636), thus the paralogs could have derived from a tandem duplication.

Their genewise scores are high (>92) and there is no evidence against these being functional elements.

The LRT for the branch model (M1/M2) was not significant for the *D. rotundus* antiretroviral genes APOBEC3G, TRIM5 and BST2, indicating that these are not evolving under positive selection (Supplementary Information 16 sheet 1). However, the LRT under M8a/M8 site model was significant for all genes, yielding a small proportion of positive selected codons (PSC) in *D. rotundus* evolving under a $\omega > 1$ (p1=0.25, ω=2.18 for APOBEC3G, p1=0.183, ω=1.55 for TRIM5-α and p1=0.197, ω=2.04 for BST2) (Supplementary Information 16 sheet 2). For APOBEC3G, the fourteen PSC scored both by NEB/BEB were located within the CMP/dCMP-type deaminase 1 domain of the protein (Supplementary Information 16 sheet 3). No relevant mutations detected for the *D. rotundus* sequence. However, most sites were not comparable given the putative partial APOBEC3G protein sequence for D. rotundus. For TRIM5-α, nine PSC were located within the zinc finger B box-type, apolipophorin-III, coiled coil and B30.2/SPRY domain regions (Supplementary Information 16 sheet 3). A single mutation unique to *D. rotundus* was detected in position 332 (R/V) (numbering in reference to the complete CDS from *Homo sapiens*), possibly affecting restriction of HIV-1 and SIVmac infection[102]. For BST2, five PSC were located within the extracellular region of the protein (Supplementary Information 16 sheet 3). A mutation unique to *D. rotundus* was detected in position 5 (S/F) (numbering in reference to the complete CDS from *Homo sapiens*), likely affecting resistance to the Vpu protein[103], and in site 18 (K/R) determined to abolish viral redistribution to late endosomes[104]. However, the relevance of such mutation cannot be determined without functional assays.

Analysis under the branch-site model (BSAn/BSA) confirmed our previous result that most of the codon sites in such antiretroviral genes in *D. rotundus* and other mammals are evolving either under neutrality (TRIM5-α) or purifying selection (APOBEC3G). Only the LRT for BST2 was significant. Although most of the sites in BST2 are evolving under neutrality (0.47637), a p=0.06609 of sites unique to *D. rotundus* are evolving under positive selection (ω=999) (Supplementary Information 16 sheet 4). Only one site (L50) was scored under NEB/BEB, and was determine to be located in the extracellular region on the protein (Supplementary Information 16 sheet 5).

Given the identified short alignments of the CDS of APOBEC3G and BST2 we then verified their ORF to find whether they have been truncated. The ORF of APOBEC3G could not be found where expected. This truncated CDS seems to be caused by the assembly, since the scaffold in which it is located (ScWXtkA_1824) is very short (1010 bp), while the complete gene length is 1008 bp. Thus, the analysis performed on this gene was limited by the length of the sequence. However, for the gene BST2 we did identify the corresponding ORF and the prediction has a high genewise score (100).

Supplementary note 5. Putative gene sequence/function loss supporting results

We further evaluated the putative loss the following genes likely related to adaptation to sanguivory: TAS1R1 (sweet taste), OR2D2 (olfactory), OR9A2 (olfactory), PGA5 (protein digestion and absorption), CTSG (killing and digestion of engulfed pathogens), AQP8 (aquaporin, osmosis regulation), CCL2 (immunoregulatory and inflammatory processes), KLRB1 (immunoregulation), PMP2 (putative cholesterol transport), CNGB3 (vision), UGT2B17 (elimination of potentially toxic xenobiotics and endogenous compounds and steroid metabolism), RNLS (modulates cardiac function and systemic blood pressure, related to chronic kidney failure), and HBB (Hemoglobin Subunit Beta) (Supplementary Information 7).

In our synteny analyses (Supplementary Fig. 13A) we found that the gene synteny of TAS1R1 is conserved in the horse and *P. alecto,* but the gene was not identified in the corresponding syntenic region in the common vampire bat, strongly indicative of a true loss of TAS1R1 rather than incomplete genomic sequencing in the common vampire bat. Also, the gene synteny of PMP2 in the genomes of *P. alecto*, *M. davidii*, and *M. brandtii* is conserved but not identified in the common vamire bat, suggesting it is likely a true loss in the common vampire bat that would require further attention in future studies. Regarding PGA5, we found that *P. alecto, M. brandtii* and the common vampire bat have the gene PGA, which is in synteny to the human genes PGA3, PGA4, PGA5. Thus, this identification by the gene loss pipeline is actually an identification of a gene expansion in the human that did not happen in the examined bats. The gene synteny of the other tested genes is not conserved, thus it cannot provide conclusive support or refutal to their putative loss.

It has been shown that in some genomes, a high GC content can lead to missannotation[55,56]. Thus, we also analyzed the GC content of the regions surrounding the previously mentioned set of genes

(Supplementary Fig. 13B). The genome average GC content of the common vampire bat is 42.2%. For the genes KLRB1 and CNGB3, we could not find the syntenic regions between the common vampire bat and other bat genomes in the pairwise whole genome alignments. For the genes AQP8 and PMP2, the GC content of the syntenic region is comparable to that of the common vampire bat genome average. For UGT2B17, the GC content of the syntenic region is slightly lower than the genome average (36.8%). In conclusion, there is no reason to suspect that any putatively identified genes are missing due to high GC content-related sequencing biases that could have thus affected the assembly and annotation of such genes.

We also searched against the common vampire bat transcriptome data[3]. The alignments of CCL2 with 43% and HBB with 20.69% gene coverage are the only ones with a match, suggesting that their sequence is not lost in the common vampire bat. This low identity sequence coverage to the reference set caused these genes to be annotated as putatively lost according to the gene loss pipeline thresholds. Likewise, we were able to identify the genes CCL2, CTSG, HBB, and RNLS in our genome annotation. We also looked for a match for the bitter taste genes TAS2R8, TAS2R10, TAS2R41, and TAS2R50, in order to evaluate the reduction in bitter taste genes in the common vampire bat. We could not identify a hit for them in the vampire bat transcriptome, supporting previous studies arguing for a reduction in bitter taste genes in the common vampire bat[105].

Furthermore, we designed PCR primers to attempt PCR amplification in the next species: *D. rotundus*, *R. aegyptiacus*, *R. euryale* (insectivorous), *M. schreibersii* (insectivorous), *N. leisleri* (insectivorous), and *N. lasiopterus* (insectivorous, seasonal carnivorous). Unfortunately, we were able to obtain amplification product only for TAS1R1 and OR2D2. The pattern observed in the TAS1R1 PCR suggests gene loss in the common vampire bat, as the gene could be amplified for the two samples of *R. aegyptiacus*, but not for *D. rotundus*. We note however that we were also unable to amplify it in the samples from *R. euryale*, *M. schreibersii* and *Nyctalus*, possibly due to the primers not being degenerate enough to cover the phylogenetically distant bats. With regards to the primers targeting the olfactory genes, they did not produce products in the the common vampire bat and the other bats with a pattern clearly showing loss in the common vampire bat, as that for TAS1R1 (Supplementary Fig. 13C). The gene OR2D2 amplified only in *D. rotundus* while OR9A2 did not produce any product. We also tested primers for another olfactory putatively lost gene, OR10A6, and obtained products from *D. rotundus*, *R. aegyptiacus* and *R. euryale*, but not from the

Nyctalus bats, again suggesting the primers are not degenerate enough to cover the phylogenetically distant bats. Both amplified sequences, OR2D2 and OR10A6, translate without frameshifts or stop codons. In conclusion, it is not possible to support the reduction in the olfactory receptors using the PCR primers tested here. However, the individual distinction of the olfactory genes from the various paralogues is a complicated issue that could be disserned in future studies. Ultimately, we advocate that in light of our *in silico* identifications for loss of various genes, future studies based upon improved sequence availability for the design of more specific or more degenerate primers, as well as availability of more quality genome assemblies, will be able to solidify our observations.

Supplementary note 6. Comparative genomic analyses

Single copy orthologs: From the TreeFam pipeline we obtained a total of 21,977 initial clusters, of which 3,384 are single-copy families, with each species having only one gene (Supplementary Fig. 14). We used those clusters to perform branch and branch-site model tests and obtained those with a statistically significant different rate of evolution (Supplementary Information 8).

1:1 orthologs: From the 1:1 ortholog pipeline we obtained a total of 27,667 initial clusters, of which 5608 are 1:1 orthologs in which each species has only one gene. We used those cleaned clusters to perform branch and branch-site model tests and obtain those with a statistically significant different rate of evolution (Supplementary Information 8).

dN/dS analysis: The two sets of built orthologous gene families were employed in the dN/dS analysis with codeml. In single-copy orthologous genes of Treefam method, we detected 17 rapidly evolving genes in branch model test and 33 genes with positively selected sites in branch site model test. In 1:1 orthologous genes of RBH method, we detected 77 rapidly evolving genes in branch model test and 63 genes with positively selected sites in branch site model test (Supplementary Tables 11-14).

Gene family expansion/contraction: We filtered the initial 21,977 gene family clusters and used 21,961 as input for CAFE, obtaining a lambda of 0.00650368. We estimated the split time between the common vampire bat and Mormoopidae is 34.5 Mya (Supplementary Fig. 15, Supplementary Table 15) and obtained Rhinolophidae and Pteropodidae clustering together with high bootstrap (Supplementary Fig. 16). We identified 144 expanded gene families and 103 contracted gene

families in vampire bat (Supplementary Fig. 17). The genes in the expanded gene families were enriched in the GO terms of cellular biosynthetic process (GO:0034645, GO:0044249, GO:0044267) and ion transport (GO:0008199, GO:0006826, GO:0000041, GO:0006879). Interestingly, the MHC class I protein (GO:0042612) was also enriched in expanded gene families (Supplementary Table 16).

Gene loss: Using a total number of 20,919 genes from the human genome, 257 were found as putatively lost in *D. rotundus* in low confidence, and 80 in high confidence. Compared to the rest of the bats, 290 (229 of low confidence and 61 of high confidence) are putatively lost only in *D. rotundus*. No GO enrichment was obtained from the lost genes.

Supplementary note 7. Gene recruitment

Regulatory evolution has been suggested to play a fundamental role in evolution[106], including dietary adaptations. A previous study by Phillips and Baker (2015)[107] identified specific genes recruited as secretory products of salivary glands in the vampire bat and in the leech. Salivary glands produce anticoagulants, vasodilators, anti-inflamatory proteins and other agents relevant to sanguivory adaptation. Among the genes they identified as putative leech and vampire bat salivary recruited related to sanguivory is *Clybl* (Citrate Lyase Beta Like), which we also identified in our analyses as evolving slower in the common vampire bat than in the other bats (ω_{null}=0.209, ω_{Alt}=0.0001, *P*= 0.00073, FDR=0.025). This gene may be involved in the metabolism of vitamin B12, which is important for the formation of red blood cells and is related to blood clotting[108–110]. Regarding the suggestion of hemostasis regulation as a trait under regulatory evolution and with relevance to sanguivory, although we did not identify the genes *Entpd1* and *Lrp1* pinpointed in the discussed transcriptomic study, we identified other hemostasis related genes, such as the plasminogen activator gene (*PLAT*, ω_{null}=0.327, ω_{Alt}=0.677, *P*=1.19E-6, FDR=0.000125)[111] under higher rate of evolution in the common vampire bat than in the other bats.

Supplementary note 8. Metagenomic analyses

Microbial taxonomic identification: The gut microbiomes of sanguivorous species (medicinal leech[112], mosquitoes[113], sea lamprey[114], and common vampire bat[115]) are known to contain *Aeromonas*, which are mostly pathogenic bacteria. So far, the most well studied sanguivorous gut microbiome is that of the medicinal leech, which has been shown to contain *A. hydrophila* and *A.*

veronii in large abundance. It has been suggested that *A. veronii* provides a number of contributions to the symbiotic relationship with the leech. These bacteria have been suggested to help in maintaining the microbiome of the digestive tract, digesting of the blood meal and providing necessary nutrients, such as vitamins from the B complex, which are not found in abundance in blood[116,117]. *A. hydrophila* has previously been found in the *D. rotundus* gut, suggesting that these bacteria are necessary for digesting the blood meal in a similar manner as in leeches[115,118]. Interestingly, the percentage of nr genes annotated as *A. hydrophila* is 0.03% in *M. gigas*, 0.45% in *R. aegyptiacus*, 0.042% in *R. ferrumequinum*, and 0.12% in *D. rotundus*. Similarly, the percentage of nr genes annotated as *A. veronii* in *M. gigas* is 0.01%, 0.025% in *R. ferrumequinum*, but 0.2% in *R. aegyptiacus* and 0.07% in *D. rotundus*. Surprisingly, although these two bacteria species are expected to be in very high abundance in the *D. rotundus* gut, as they are in leech to help it digest the hemoglobin, it is *R. aegyptiacus* the one that has them in highest abundance, with *D. rotundus* in second place.

Common vampire bat taxonomic core: Pooling the results from all the searched databases, the common vampire bat core has 5,755 different bacterial species, which total 3,853 when ignoring the strain, and 1,337 when removing the potentially new genera (Supplementary Table 10). The sub-category with most species is the pathogens, with 281, mostly belonging to *Mycoplasma* and *Staphylococcus*. The second most abundant taxa are those related to salty environments, with 106 species, mostly from *Bacillus*. For example, we identified the plasmid *Bacillus megaterium* NBRC 15308 = ATCC 14581 plasmid pBMV_1, which codes for the ectoine utilization protein EutC, which serves as a protective substance by acting as an osmolyte and thus helps organisms survive extreme osmotic stress. Immunity is the third most abundant category, with 88 species, mostly from *Streptomyces*. For example, we identified *Actinokineospora sp.* EG49 in the DIAMOND search. This bacteria produces an antitrypanosomal compound[119], which is relevant given that the majority of trypanosomal species are transmitted by blood-feeding invertebrates. Also, the identification of the plasmid *Haloferax volcanii* DS2 plasmid pHV4, is relevant since it encodes the Duffy chemokine receptor. The coagulation category is in the fourth place, with 67 species (Supplementary Table 9).

Functional identification: The total of 1,212,453 assembled nr genes for each individual were pooled to produce 415,583 nr genes annotated with the Uniprot database. From those, 1,927

sequences have an assigned KO, and 10,042 and eggNOG id. Looking at the COG categories, *D. rotundus* clusters with *R. aegyptiacus*, and *M. gigas* and *R. ferrumequinum* cluster together (Supplementary Fig. 8A). Looking at the KOs grouped by class, *D. rotundus* also clusters with *R. aegyptiacus* and *M. gigas* clusters with *R. ferrumequinum* (Supplementary Fig. 8B). However, when looking into the metabolic pathways not grouped by classes (Supplementary Fig. 8C), *D. rotundus* does not cluster with *R. aegyptiacus* anymore and stays in the base of the dendrogram, with *R. aegyptiacus* also being alone and *R. ferrumequinum* and *M. gigas* clustering together.

When looking at the statistical significance of the distributions of counts of the different metabolic classes, *D. rotundus* is statistically dissimilar to *R. ferrumequinum* in lipid metabolism (P=0.002) and xenobiotic biodegradation and metabolism (P=0.05), but statistically more similar to it in the glycan biosynthesis and metabolism. Taking into account that the Phyllostomidae evolved from an insectivorous common ancestor[120], we hypothesised that the microbiome of an insectivorous bat that specialized in degrading glycans (such as chitin) provided the suitable starting point that subsequently enabled them to adapt to processing of the polysaccharides present in blood. Additionally, the dissimilarity of *D. rotundus* to the insectivorous lipid metabolism can be explained by the difference in the lipid composition in insects and blood, with most of the reserve fat of insects being triglycerides[121,122], while those in blood are in the form of lipoproteins. Lastly, the dissimilarity in the xenobiotic metabolism could also derive from the difference in the kinds of xenobiotics present in blood compared to those present in an insect-based diet.

Common vampire bat functional core: 386,038 of the nr annotated genes are in the common vampire bat gene core, corresponding to 1,900 KOs, and 9,722 eggNOGs. Although all the pathways are equally covered in the core (Supplementary Information 17), the most abundant functions were related to nutrition, the second most abundant category was fat, and the third was immunity. Most of the enzymes kept in the core are specific to a certain pathway, while the majority of those removed are more generic in their functions (P=1.1e^{-12}). Furthermore, the eggNOG category with most identifications per COG is the lipid metabolism (25.6 genes per COG in average). Among the top most abundant are the amino acid metabolism and transport (18.8 with 245 uniq cogs) and energy production and conversion (14.6 with 213 unique COGs). Among the least abundant are those from more generic functions, such as RNA processing and modification (with only 1 unique COG identified and 3 per COG), chromatin structure and dynamics (with only

5 unique COGs identified and 5.6 per COG), and cell motility (with 77 unique COGs identified and 9.5 genes per COG) (Supplementary Fig. 18). This supports the hypothesis that those functions kept in the common vampire bat microbial core play a role in the specific processing of blood as a diet. The most abundant functional subcategory was nutrient (54 entries), the second most abundant was fat (33 entries), and the third was immunity (18 entries). We also identified 31 functions classified as relevant (Supplementary Table 9). Some of the relevant manual annotations in the functional annotation include those related to medium adaptation, nutrition, and coagulation. For example, we identified enzymes for the biosynthesis of capsaicin. Capsaicin has been shown to reduce cholesterol, triglycerides and platelet aggregation, as well as to dissolve fibrin[123–125], with these last two capabilities potentially having an important impact in the ingested blood.

Common vampire bat enriched genes: From the total of 1,146,171 genes that were part of the bat nr gene set catalogue (constructed using all the predicted genes from all the bats and clustered with 95% identity), 569 genes were identified as significantly enriched in the common vampire bat gut microbiome (Fisher test $P<0.05$). From those, 24 *D. rotundus* genes belong to the biosynthesis of other secondary metabolites pathway (and 0 from *M. gigas*, 9 from *R. eagyptiacus*, and 1 from *R. ferrumequinum*), 144 to the amino acid metabolism (and 0 from *M. gigas*, 103 from *R. eagyptiacus*, and 3 from *R. ferrumequinum*), 115 to the carbohydrate metabolism (and 1 from *M. gigas*, 88 from *R. eagyptiacus*, and 6 from *R. ferrumequinum*), 82 to the energy metabolism (and 0 from *M. gigas*, 46 from *R. eagyptiacus*, and 0 from *R. ferrumequinum*), 37 to the glycan biosynthesis and metabolism (and 0 from *M. gigas*, 16 from *R. eagyptiacus*, and 1 from *R. ferrumequinum*), 32 to the lipid metabolism (and 0 from *M. gigas*, 15 from *R. eagyptiacus*, and 0 from *R. ferrumequinum*), 120 to the metabolism of cofactors and vitamins (and 1 from *M. gigas*, 55 from *R. eagyptiacus*, and 3 from *R. ferrumequinum*), 61 to the metabolism of other amino acids (and 0 from *M. gigas*, 30 from *R. eagyptiacus*, and 0 from *R. ferrumequinum*), 34 to the metabolism of terpenoids and polyketides (and 0 from *M. gigas*, 10 from *R. eagyptiacus*, and 1 from *R. ferrumequinum*), 134 to the nucleotide metabolism (and 1 from *M. gigas*, 66 from *R. eagyptiacus*, and 0 from *R. ferrumequinum*), and 13 to the xenobiotics biodegradation and metabolism (and 0 from *M. gigas*, 14 from *R. eagyptiacus*, and 0 from *R. ferrumequinum*).

Comparison of rectal swabs to fecal samples: We confirmed that the swab samples cluster among the fecal samples, as expected if their microbiomes are similar (Supplementary Fig. 19). The swab

sample rectal_swab_2 is placed far from the other samples in the 2D PC plot, including the other rectal swab sample (Supplementary Fig. 19C). This samples had too low extracted DNA to be measured, which could explain its location. However, when looking at the 3D PC plot (Supplementary Fig. 19D), this sample is in the same plane as the other rectal swab sample, supporting the validity of its usage in the analyses that included it.

Supplementary note 9. Breast milk sample

We also obtained the metagenome of one female *D. rotundus* breast milk sample, which we used to gain an insight into the gut microbial phylogenetic influence. In spite of the small sample size, the results from this sample provide information on the heredity of gut microbiome taxa and functions. We identified taxa related to the following challenges of sanguivory (Supplementary Information 18).

Blood digestion: We identified the plasmid *Sphingomonas* sp. MM-1 plasmid pISP0, which contains genes coding for copper resistance proteins, esterase/lipase and secretion protein HlyD family protein (proteins for the secretion of hemolysins).

Nutrition: We also identified *Butyrivibrio crossotus*, which is also part of the sanguivorous microbial core. These bacteria are able to produce butyrate, biohydrogenate lipids and produce microbial inhibitors. Furthermore, we identified relevant plasmids, such as *Psychrobacter* sp. G plasmid PsyG_26, which includes the thiamine (vitamin B_1) biosynthesis protein ApbE. Importantly, the most abundant bacteria identified from the Human Microbiome database is *Propionibacterium* sp. 5_U_42AFAA (38,184 mapping reads), which is able to synthesise vitamin B_{12}.

Lipids: The 9th most abundant plasmid in the milk was the *Novosphingobium aromaticivorans* DSM 12444 plasmid pNL1 (722 mapping reads). *N. aromaticivorans* is a potential initiator of primary biliary cirrhosis[126], and this plasmid contains genes for fatty acid desaturase, ferredoxin, and L-carnitine dehydratase/bile acid-inducible protein F.

Medium adaptation: We identified *Gemella sanguinis* as the second most abundant bacteria from the Human Microbiome dataset (35,284 mapping reads). *Gemella* bacteria are primarily found in

the mucous membranes of humans and other animals, particularly in the oral cavity and upper digestive tract[127]; and strains of this species were originally isolated from samples of human clinical blood[128]. Even more, the most abundant identified plasmid is the *Tistrella mobilis* KA081020-065 plasmid pTM3 (2,900 mapping reads). Interestingly, *T. mobilis* produces polyhydroxyalkanoates[129], which are linear polyesters produced in nature by bacterial fermentation of sugar or lipids in order to store carbon and energy.

Supplementary note 10. Diet of the proto-vampire bat

An aim of the comparison of the common vampire bat gut microbiome to those of bat species basal to the Phyllostomidae clade was to attempt to identify what the potential diet was of the proto-vampire bat, prior to its switch to obligate sanguivory. Several hypotheses have been proposed with regards to this intermediate diet following the proto-vampire's divergence from its insectivorous ancestor. These hypotheses include that in which the proto-vampire bat was primarily insectivorous but with the additional ability to eat some fruit[130]. Other hypotheses suggest that the proto-vampire was a nectivorore, a frugivore, an arboreal-feeding carnivore, a bat that fed on ectoparasites, on insects and larvae from wounds in mammals, or a more general omnivore[131–136]. It has been suggested that no highly specific feeding strategy evolves from another highly specific feeding strategy[137]. Considering the gut microbiome profile as observed in the taxonomic and functional cladograms (Fig. 3), the following scenarios are possible.

1) The sister clade to the common vampire bat samples contains the frugivorous *R. aegyptiacus*, in which case we could suggest that the proto-diet was a frugivorous one.
2) The sister clade to the common vampire bat samples contains the carnivorous *M. gigas*, in which case the proto-diet could have been a carnivorous one.
3) The sister clade contains the insectivorous *R. ferrumequinum* samples, in which case it is not possible to distinguish the host phylogenetic influence from ancestral dietary state.

A portion of the common vampire bat samples cluster in the same clade as the insectivorous *R. ferrumequinum* bat in the the gut microbiome taxonomic results, instead of being in their separate Phyllostomidae clade together with the rest of the common vampire bat samples (Fig. 3A). Thus, it is possible that the proto-diet was insectivorous. However, considering that in the gut microbiome functional dendrogram the closest clade to the common vampire bat clade contains fruit, insect, and

meat eating bats (Fig. 3B), it is also possible that the blood feeding behavior evolved from an omnivorous proto-vampire. Furthermore, there seems to be an influence of the sampling location in the clustering of the common vampire bat samples.

In order to test whether this non-monophyletic clustering is due to differences in sampling location and prey consumption, we examined the prey of the common vampire bats. It sould be noted that all the Costa Rican individuals were sampled from the same site (Coto Brus), while the Peruvian samples come from Lima, Apurimac, and Huanuco. To this end, we examined the MGmapper results from the mapping of the datasets against the mammalian whole genome database. Based on the normalized abundance of reads mapping to putative preys, we observed that in the PCA analysis the samples separate with a pattern similar to that observed in the gut microbiome taxonomic and functional cladograms, although not as strongly influenced by sampling site (Supplementary Fig. 20AB). The signal is less strong in the cladogram (Supplementary Fig. 20C), and there is no statistically significant difference in the total putative prey DNA abundance based on location (P=0.9661). However, in regards to prey availability by sampling site, we found statistically significant differences in the abundance of reads mapping to specific different prey types. We found a statistically significant difference in goat DNA abundance (identified by mappings to *Capra hircus* and *Ovis aries*), with the samples from Costa Rica containing less goat DNA (P=0.01663, Costa Rica mean abundance=14,579.86, Peru mean abundance=82,835.25). We also found a significant difference in abundance of reads mapping to horse (identified by mappings to *Equus caballus* and *Equus przewalskii*), with the samples from Costa Rica containing less horse DNA (P=0.02624, Costa Rica mean abundance=11,024.43, Peru mean abundance=128, 008.50). The abundance of pig prey (identified by reads mapping to *Sus scrofa*) had a P=0.08949, with Costa Rica mean abundance=27,855.57 with a sd=48,044.88 (likely due to difference in time after feeding), and Peru mean abundance=227.67 with a sd=312.036 (with 3 out of the 6 samples having zero mapping reads). Although the Costa Rican vampire bats contain more reads mapping to cow (identified by mappings to *Bos grunniens*, *Bos taurus*, and *Bubalus bubalis*), it did not result in a statistically significant difference in abundance (P=0.1839, Costa Rica mean abundance=244,639.7, Peru mean abundance=157,057.2).

Although prey availability, determined by the sampling location, could play an important factor in the gut microbiome profile, other physiological factors besides sampling site could also drive the

observed clustering, such as age, weight, time after feeding, and reproductive state[138] (Supplementary Information 19). For instance, the vampire bat individual id 577 was a lactating adult female, its physiological state might explain its unique positioning in the bacterial genera taxa cladrogram (Supplementary Fig. 7). The apparent influence of the sampling location in the clustering of the common vampire bat samples in both taxonomic and functional profiles could also be due to the difference in capture time that could reflect influence in time after or previous to feeding (18:00 to 22:00 in Costa Rica, and 18:00 and 06:00 in Peru). Thus, it is not possible to dissern whether the bi-clustering of the common vampire bat samples is due to horizontal or vertical transmission. We suggest that comparison to other bats with other diets from within the phillostomids and other bat families not basal to the phyllostomids, such as the Vespertilionidae, could in the future aid in determining the intermediary diet of the proto-vampire bat before becoming an obligate sanguivorous.

Supplementary note 11. Microbiome phylogenetic and dietary influence

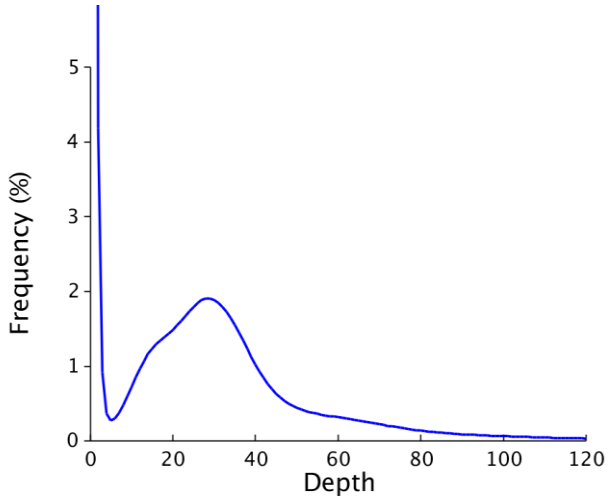
When taking into account the presence/absence of bacteria, the samples cluster by diet type (Fig. 3A). However, when considering the abundance of the present bacteria (Supplementary Fig. 7), both at the genera and species level, the clustering of the samples by dietary type is not as clearly differentiated based on diet. Interestingly, the analysis of the plasmids reveals that both the presence/absence and the abundance data provide a clearer clustering of the samples based on diet (Supplementary Fig. 21).

These microbial taxonomic observations could be explained by the hologenomic proposition in regards to genetic variation which states that, similar to the traditional genetics view, the microbiota's genetic input to the holobiont can vary in terms of *i*) differences in relative abundance of taxa, *ii*) incorporation of new taxa, and *iii*) variation in the genomes of the residing microbes (e.g. through plasmids)[139–141]. Taking this into account, scenarios *i*) and *ii*) explain that the difference between the cladrograms of the bacterial taxonomic presence/absence and abundance could be due to a change in the relative abundances of the gut bacteria. And scenario *iii*) provides an explanation to the particular plasmid cladograms, suggesting that the common vampire bat microbiome has undergone microbial adaptation through aquicition of particular plasmids likely coding for genes required for adaptation to sanguivory. This suggestion is supported by the microbial functional cladrogram, in which the closest clade to the common vampire bat clade

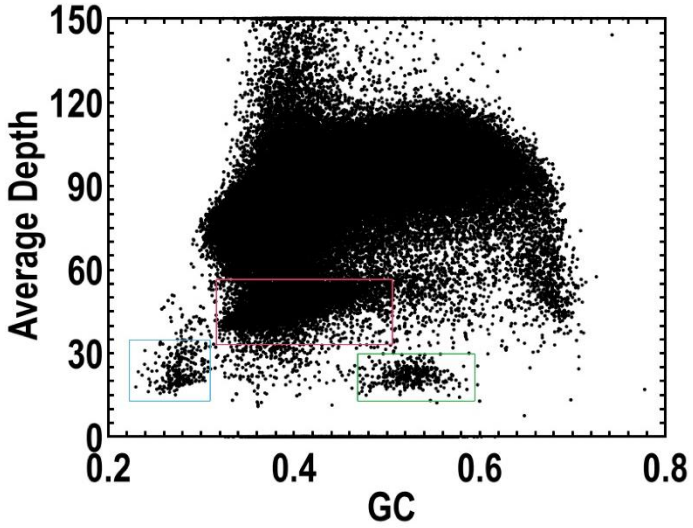
contains fruit, insect, and meat-eating bats (Fig. 3B), while the common vampire bat clade is almost completely separated from the non-sanguivorous bats, highlighting a specialized functional profile for sanguivory.

Notably, some samples deviate from these predictions in Fig. 3. The presence/absence dendrogram shows a *R. aegyptiacus* sample (*R. aegyptiacus* 9) outside of its frugivorous cluster, inside the insectivorous cluster of *R. ferrumequinum* (Fig. 3A). This positioning is likely due to *R. aegyptiacus* 9 being the least sequenced sample from the *R. aegyptiacus* dataset. In order to take depth of sequencing into account for the functional cladrogram (Fig. 3B), where we used the normalized number of assembled proteins with an annotation, we removed the deepest sequenced samples from all the dataset, which were one *D. rotundus* sample (*D. rotundus* 5085) and the three *R. ferrumequinum* samples (*R. ferrumequinum* 6, 8, and 10).

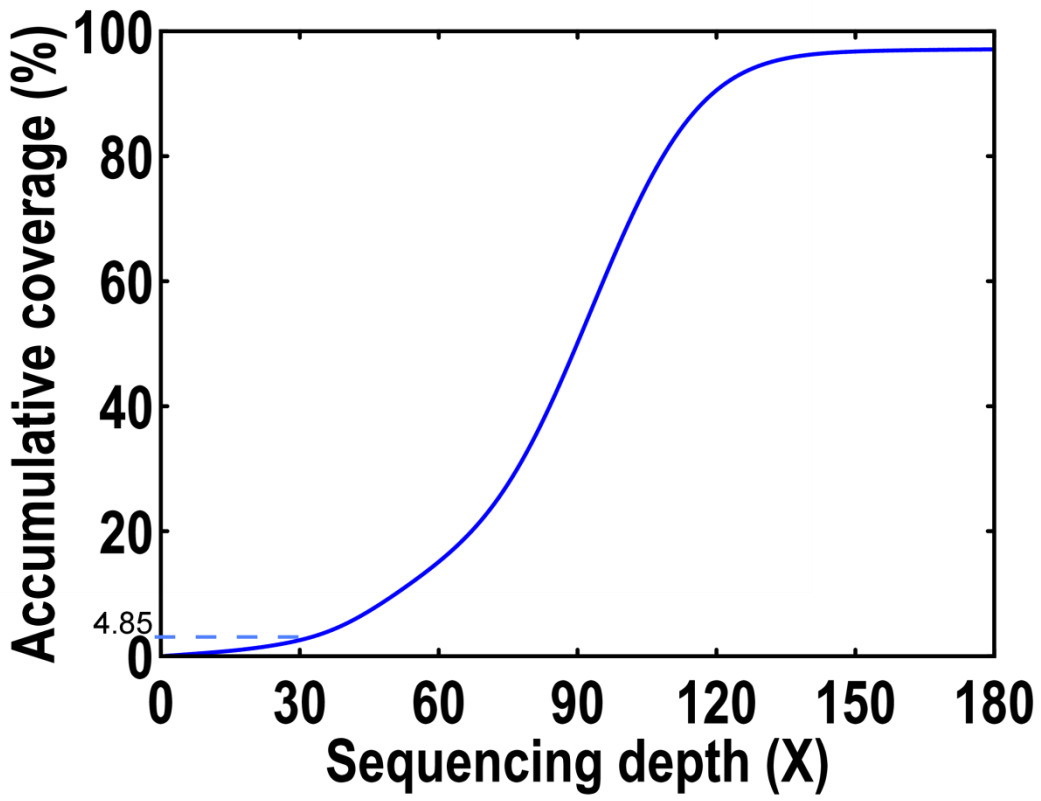
Supplementary Figures



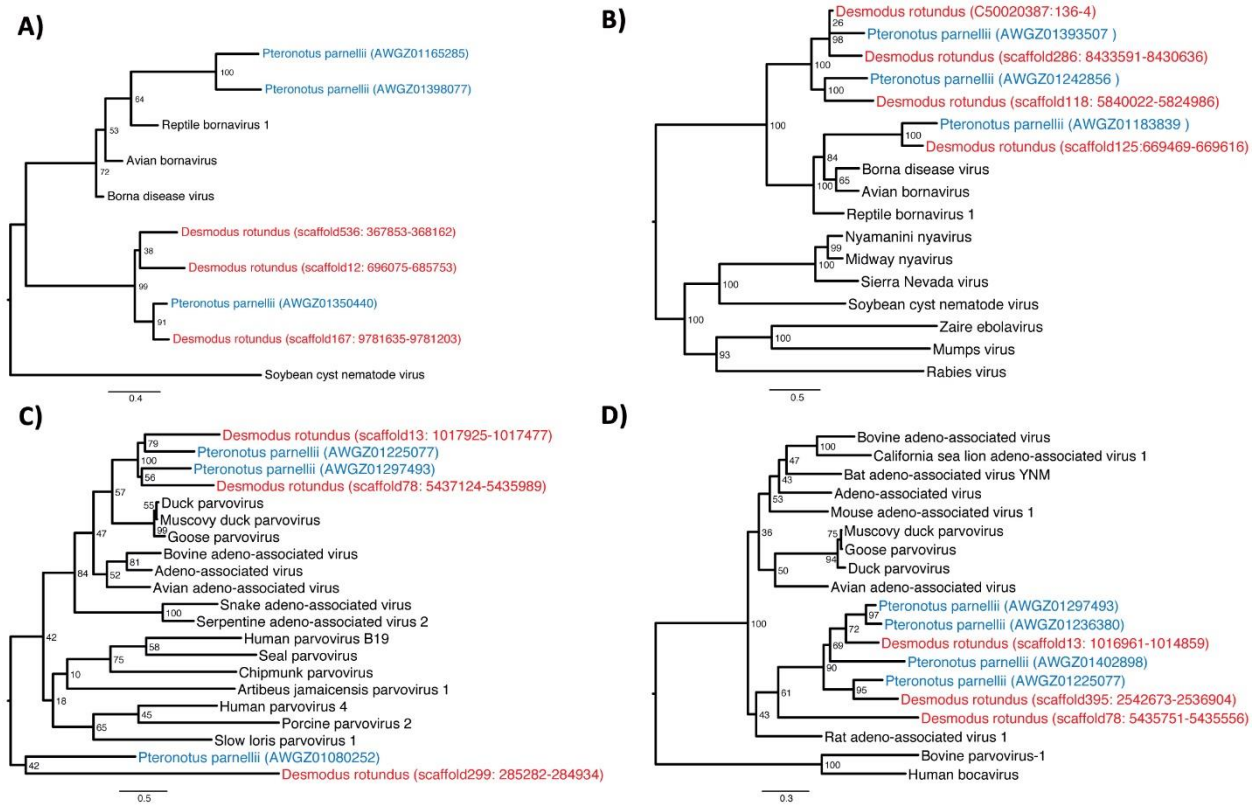
Supplementary Figure 1. Distribution of 17-mer in reads from the *D. rotundus* genome. Genome size of *D. rotundus* was estimated to be 2.13 Gb. The model frequency is located at a depth of 28.



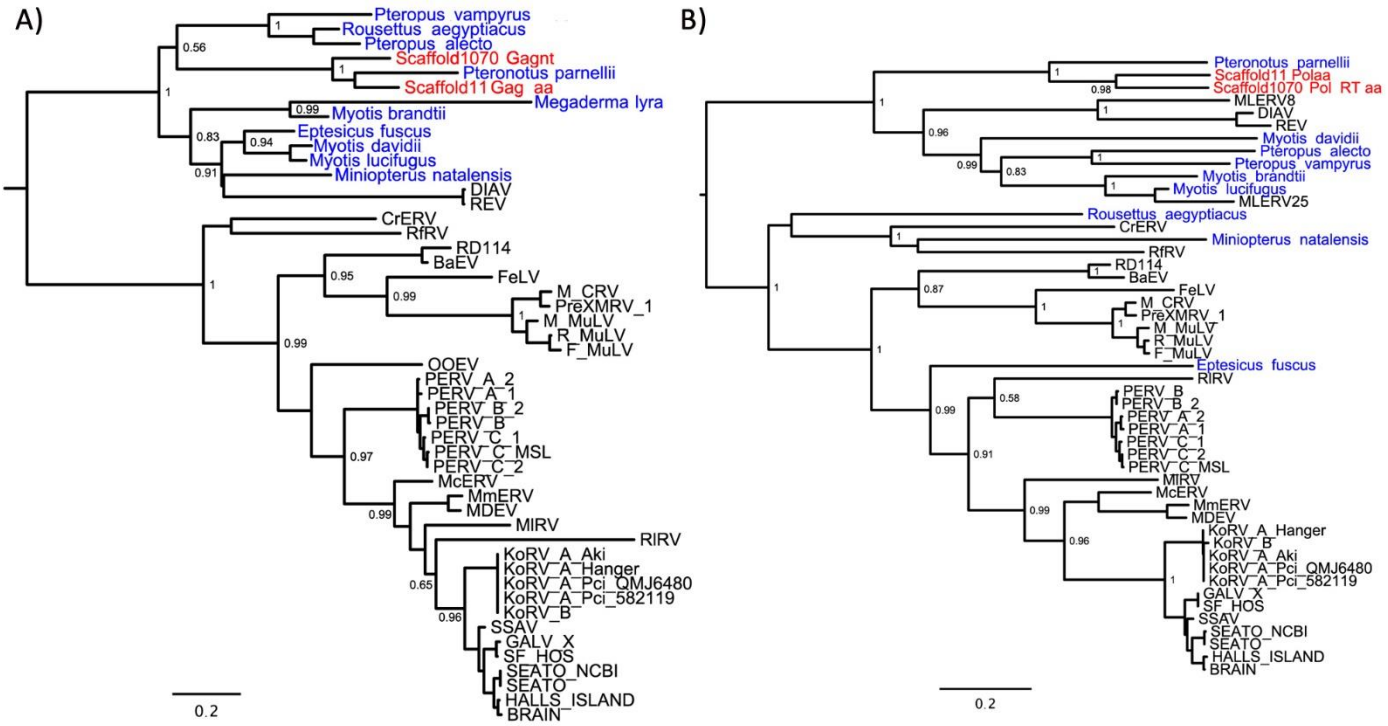
Supplementary Figure 2. Distribution of GC depth. The x-axis represents as GC content; the y-axis represents the average depth. We used 10 kb non-overlapping sliding windows to calculate the GC content and average depth among the windows. The red rectangle region maybe suggested a sex related chromosome as the sample was taken from a male bat. And the blue and green rectangle region maybe suggested bacteria contamination.



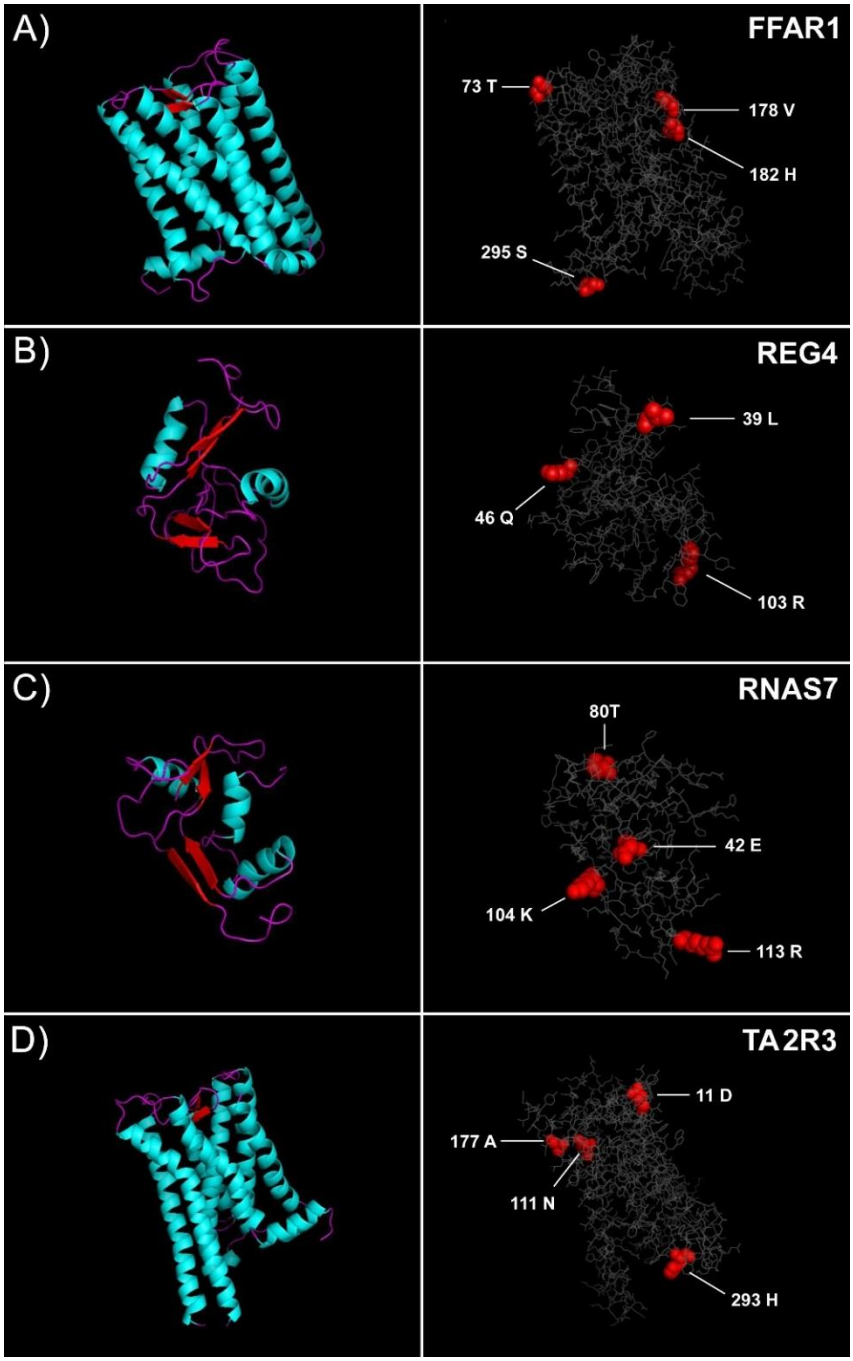
Supplementary Figure 3. Cumulative distribution of single-base depth of vampire bat genome. Approximately 95.15% of the genome is covered by more than 30 reads.



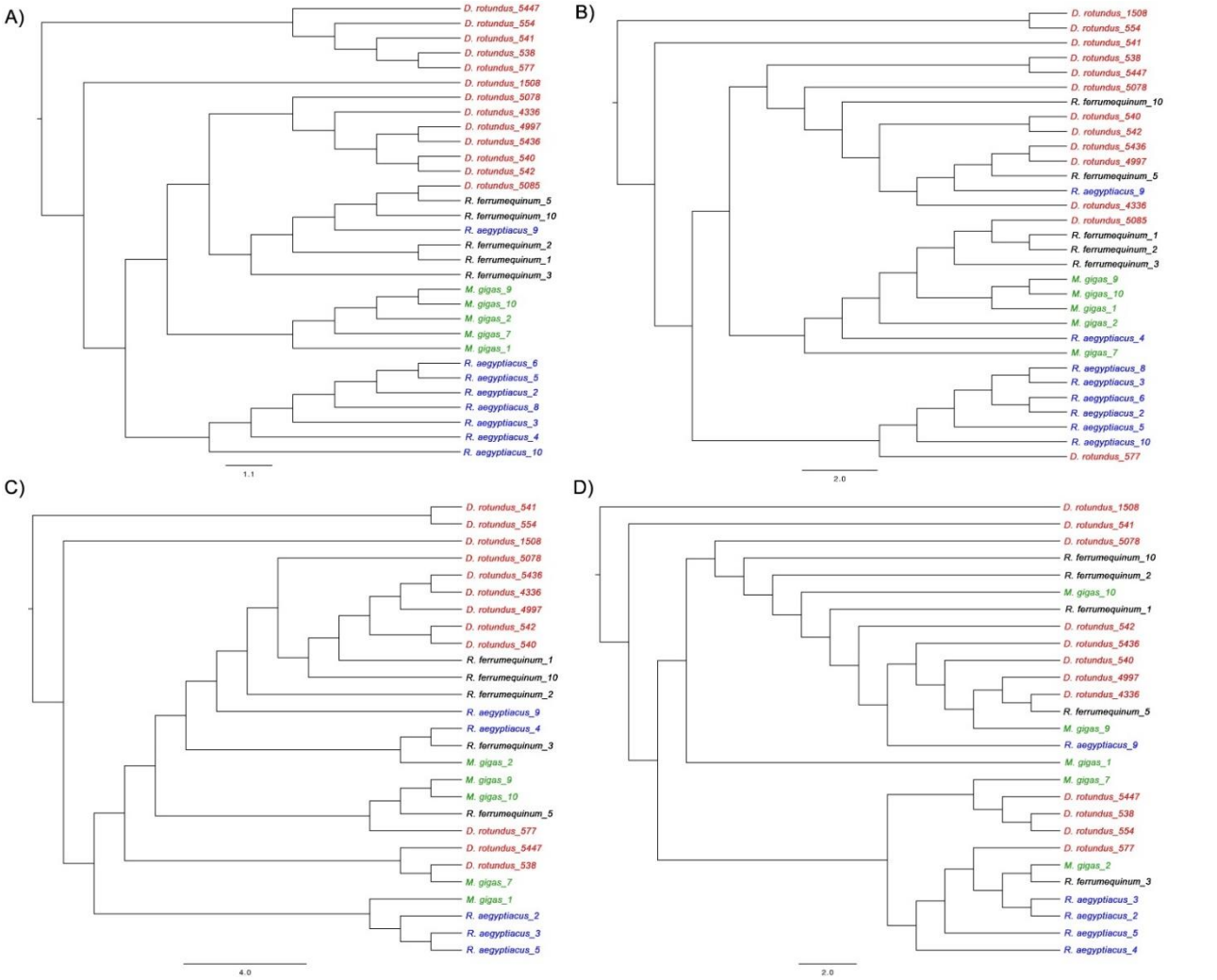
Supplementary Figure 4. Phylogenies of discovered EVEs. (**A**) Bornaviridae nucleocapsid, (**B**) Bornaviridae RNA-dependent RNA polymerase, (**C**) Parvoviridae non-structural protein 1, (**D**) Parvoviridae coat protein.



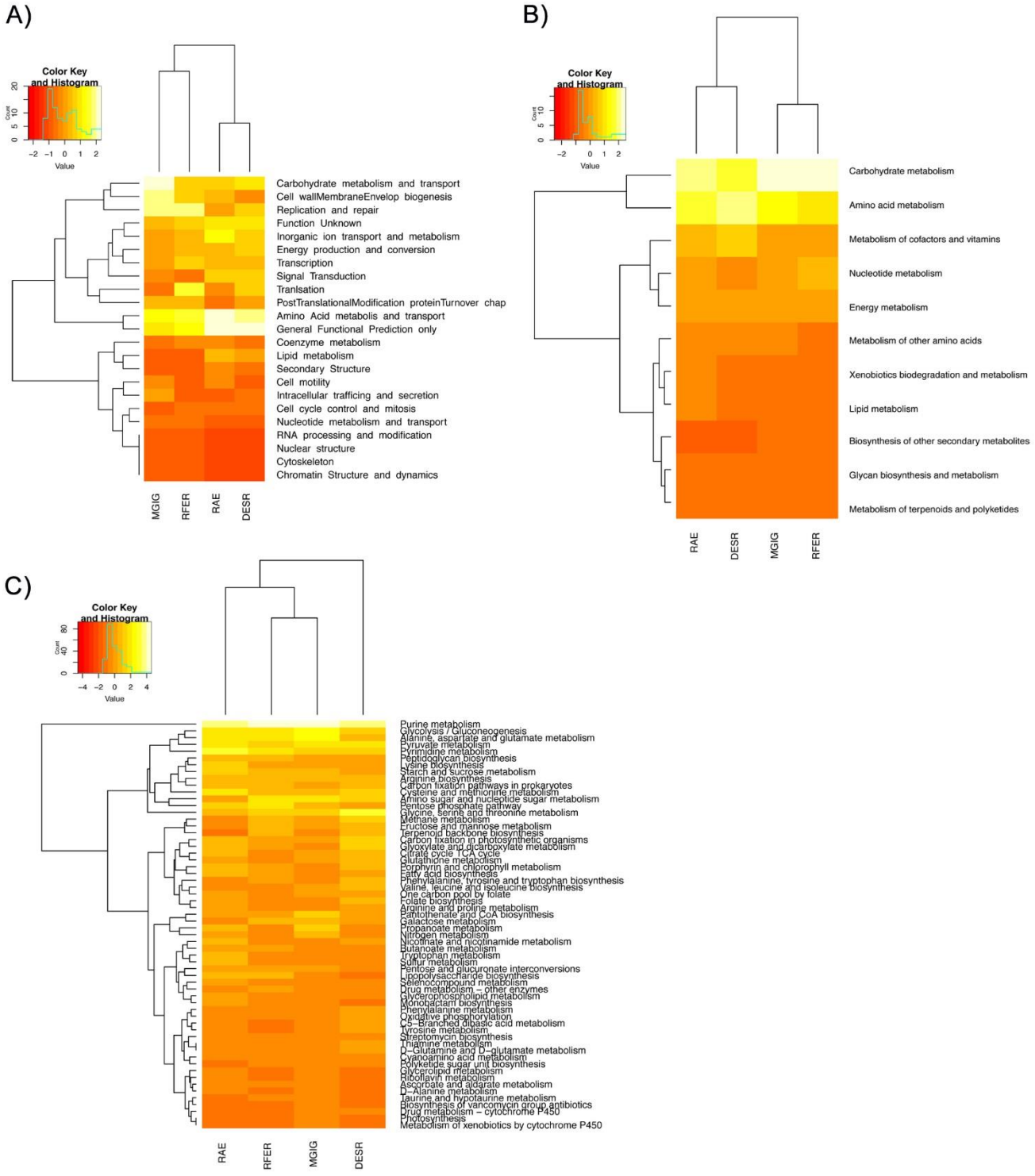
Supplementary Figure 5. DrgERV phylogenetic analysis. ML trees for the (**A**) Gag and (**B**) Pol DrgERV-like protein sequences. DIAV: Duck infectious anemia virus; REV: Reticuloendotheliosis virus; MLERV: Myotis lucifugus endogenous retrovirus; CrERV: Odocoileus hemionus endogenous virus; RfRV: Rhinolophus ferrumequinum retrovirus; RD114: Feline endogenous retrovirus (RD-114 virus); BaEV: Baboon endogenous virus; FeLV: Feline leukemia virus; MuLV sub-cluster: Murine Leukemia Virus-related; OOEV: Orcinus orca endogenous retrovirus; PERV sub-cluster: Porcine endogenous retrovirus; RlRV: Rousettus leschenaultii retrovirus; MIRV: Megaderma lyra retrovirus; McERV sub-cluster: Mus caroli endogenous virus-related; KORV-GALV sub-cluster: Phascolarctos cinereus retrovirus and Gibbon ape leukemia virus-related. The DrgERV sequences found in D. rotundus are highlighted in red, while homologous sequences in other bat species are highlighted in blue. Branch support values in aLRT-SH like are shown for the main nodes. Trees are rooted by midpoint for clarity purposes.



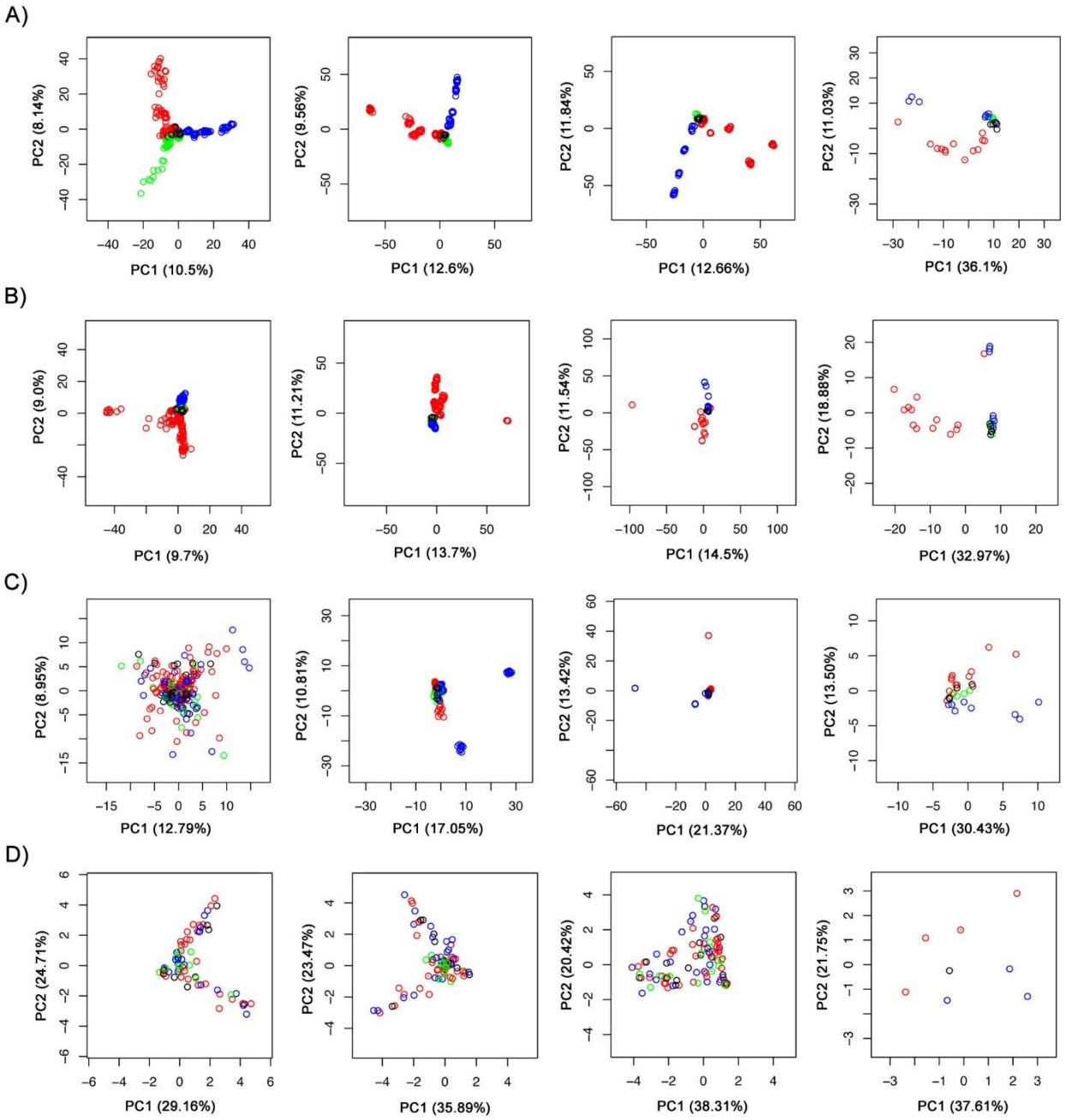
Supplementary Figure 6. Protein 3D models. On the left side the domains of the modeled proteins are illustrated (ribbon diagram) and on the right side the lines diagram of the modeled proteins with the amino acid substitutions specific to *D. rotundus* highlighted as red spheres. (**A**) Free fatty acid receptor 1 protein (FFAR1). (**B**) Regenerating islet-derived protein 4 (REG4). (**C**) Ribonuclease A family member 7 (RNAS7), also called Skin-derived antimicrobial protein 2. (**D**) Taste receptor type 2 member 3 protein (TA2R3).



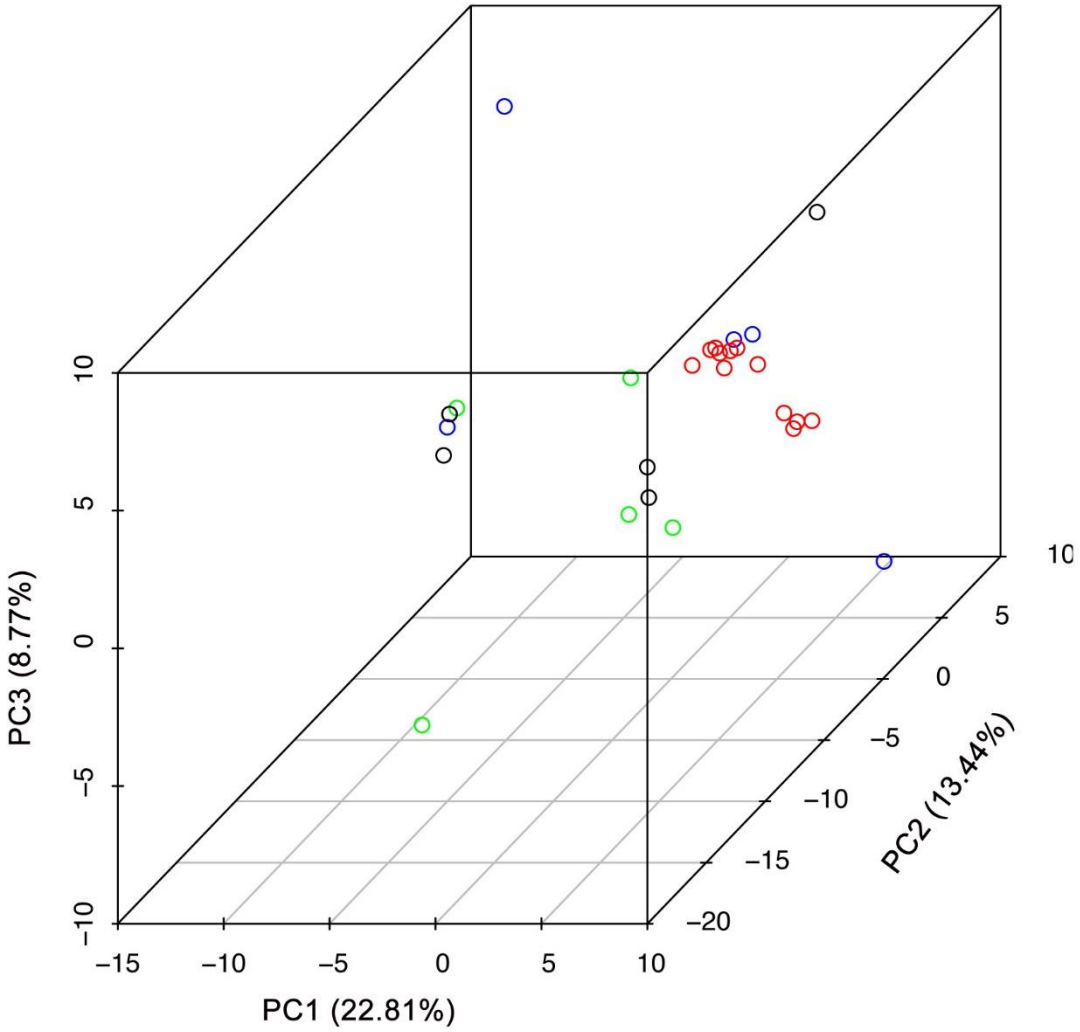
Supplementary Figure 7. Bacterial taxa distance midpoint-rooted dendrogram at species and genera level. (**A**) Using the minimum subsampled for normalization presence/absence at the species level and at the (**B**) genera level. (**C**) Bacterial taxa abundance min subsampled for normalization at the species level and at the (**D**) genera level from *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue).



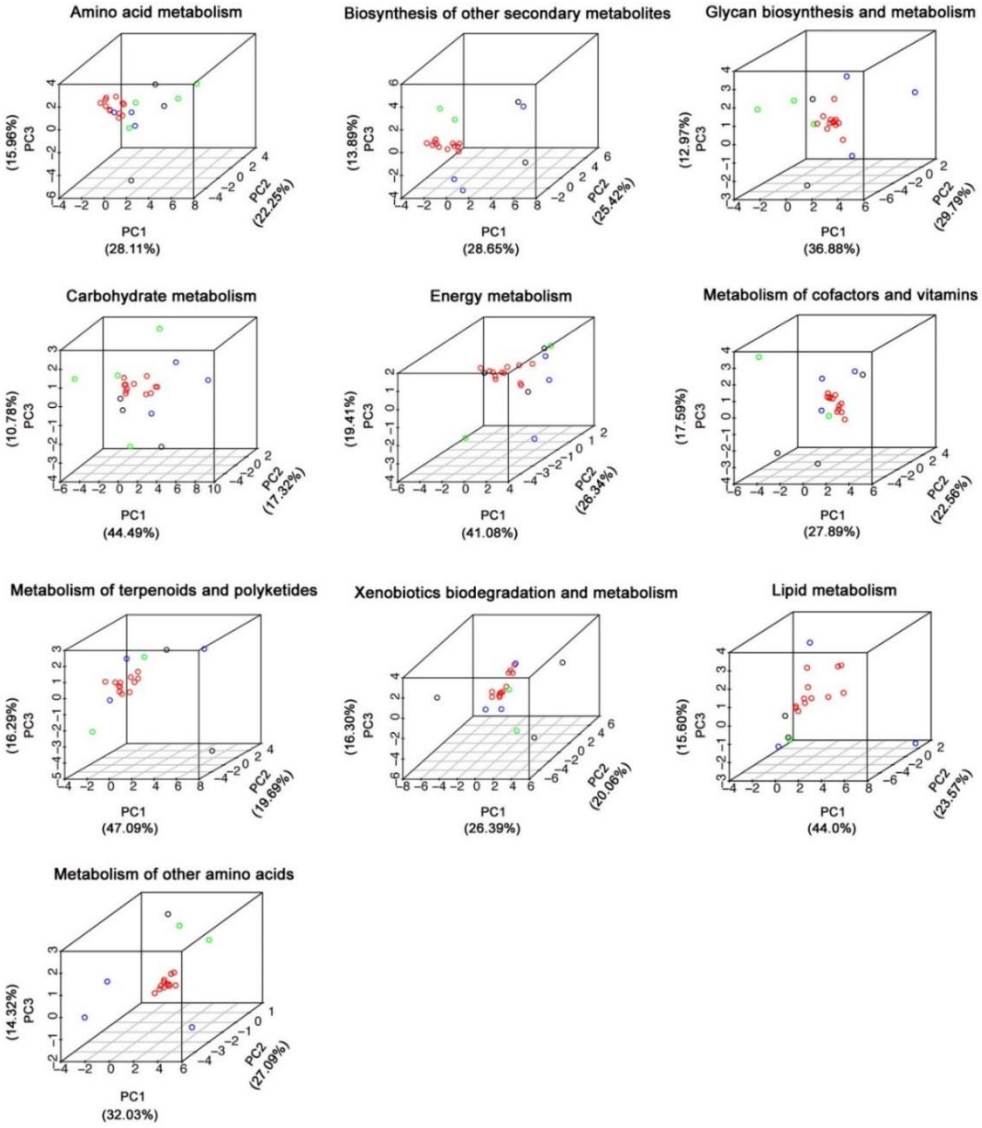
Supplementary Figure 8. Heat maps of minimum down-sampled counts for normalization. Functional abundance comparison of *D. rotundus* (DESR) to *R. aegyptiacus* (RAE), *M. gigas* (MGIG), and *R. ferrumequinum* (RFER)**.** (**A**) COG categories (**B**) Grouped KOs by class (**C**) Grouped by KOs.



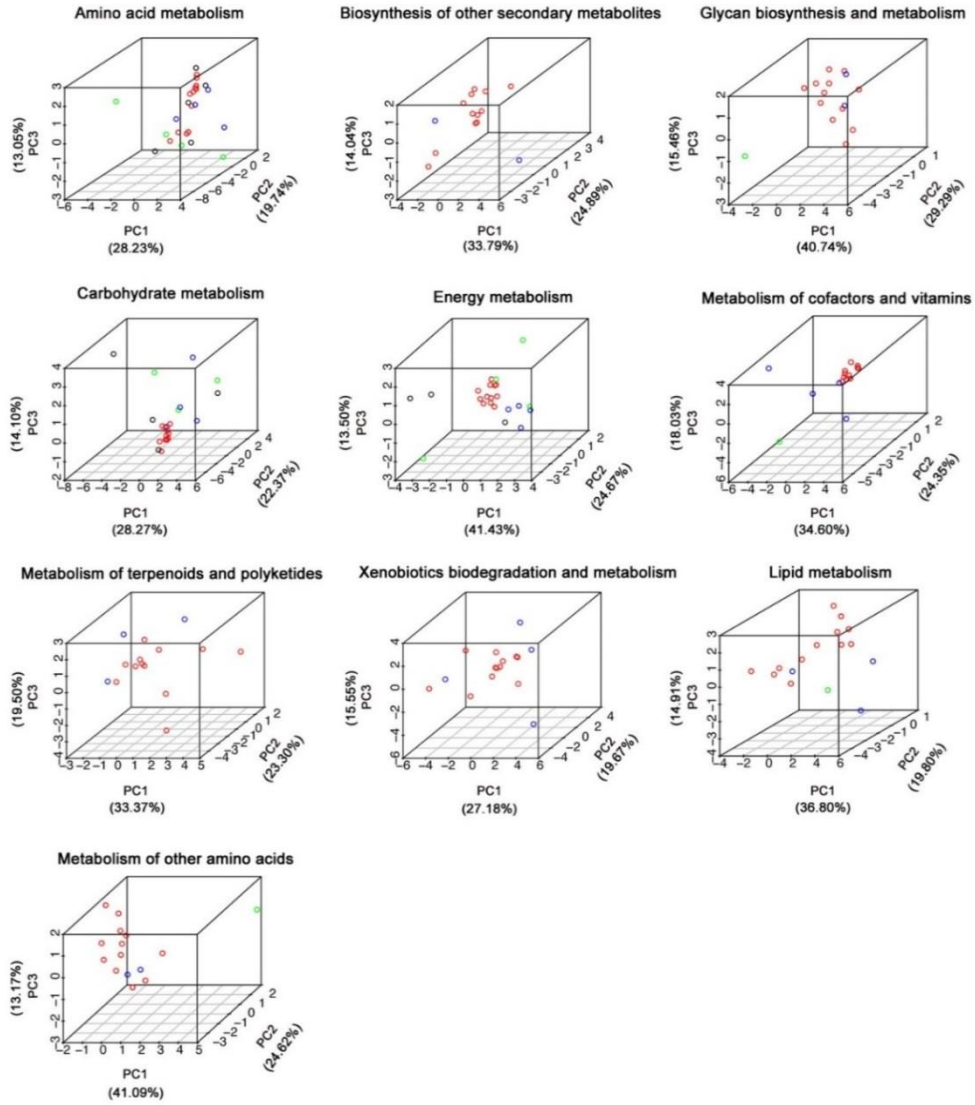
Supplementary Figure 9. PCA analysis of gut microbial taxa. Rotated PCAs of 10 replicates of subsampling the species level taxonomic abundance matrix by minimum, median, 3rd Qu from the abundance distribution, and presence/absence without subsampling for (**A**) bacteria (31 PCs), (**B**) plasmid (31 PCs), (**C**) virus (29 PCs), and (**D**) fungi (7 PCs) from *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue).



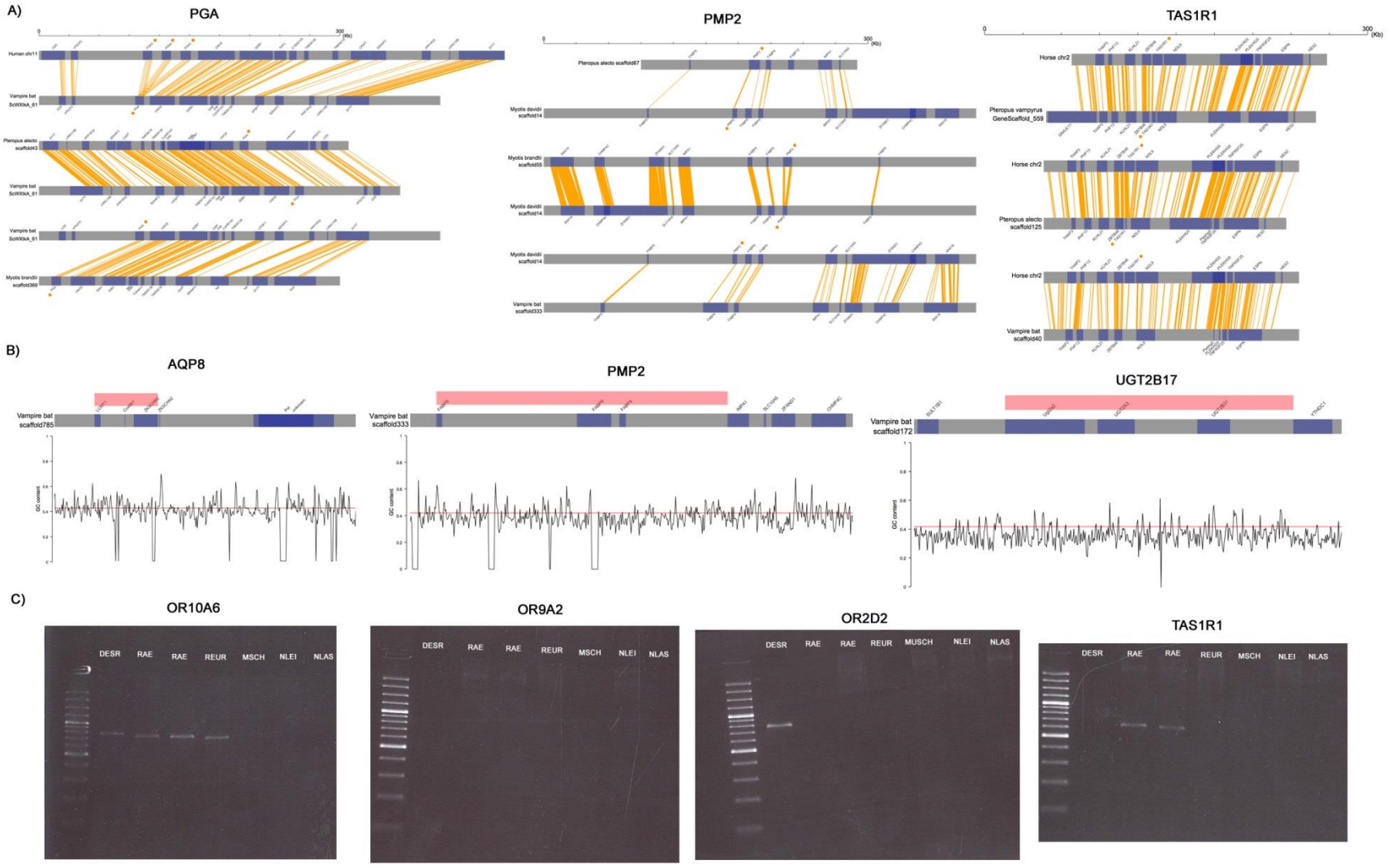
Supplementary Figure 10. PCA analysis of the gut microbial functional profile. The first three PCs, from a total of 27, of the normalized abundance of functions identified from the Uniprot annotated nr genes from *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue).



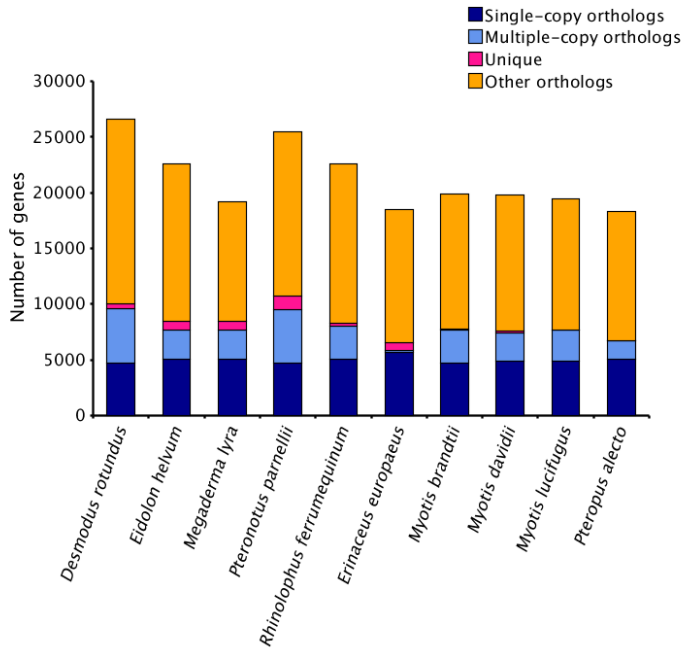
Supplementary Figure 11. Functional abundance of each metabolic category from the *de novo* assembled proteins searched against Uniprot. *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue), excluding the *D. rotundus* sample and the three *R. aegyptiacus* samples with highest depth of sequencing. Normalized values of the copy number of identified genes from each metabolic category were used. Amino acid metabolism: 14 PCs. Biosynthesis of other secondary metabolites: 19 PCs. Glycan biosynthesis and metabolism: 11 PCs. Carbohydrate metabolism: 15 PCs. Energy metabolism: 7 PCs. Metabolism of cofactors and vitamins: 12 PCs. Metabolism of terpenoids and polyketides: 14 PCs. Xenobiotics biodegradation and metabolism: 20 PCs. Lipid metabolism: 15 PCs. Metabolism of other amino acids: 8 PCs.



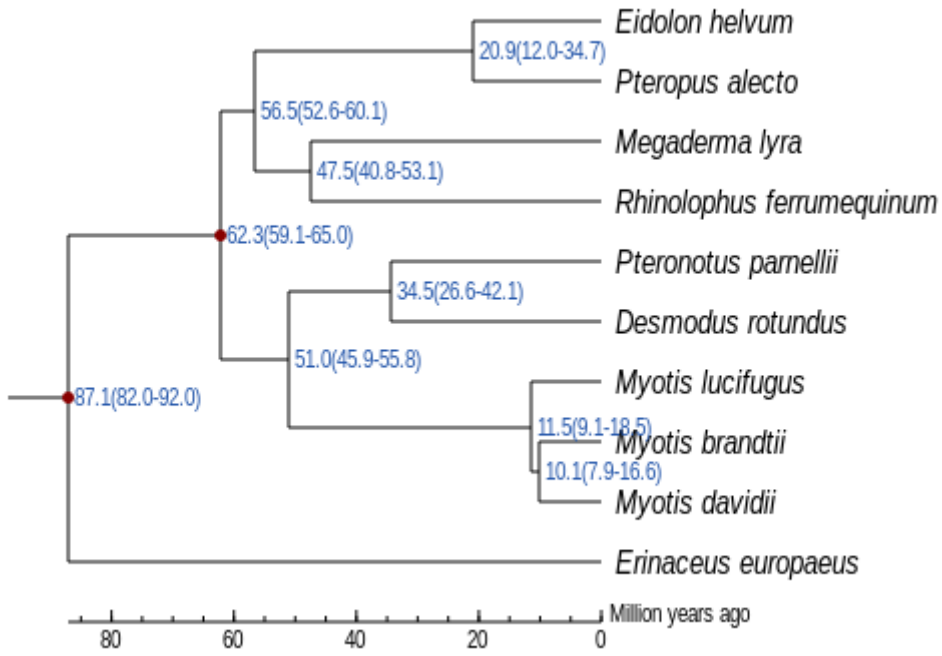
Supplementary Figure 12. **Functional abundance of each metabolic category from the reads not mapping to the whole genome databases searched against Uniprot.** The samples are from *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue), excluding the *D. rotundus* sample and the three *R. aegyptiacus* samples with highest depth of sequencing. Normalized values of the copy number of identified genes from each metabolic category were used. Amino acid metabolism: 14 PCs. Biosynthesis of other secondary metabolites: 10 PCs. Glycan biosynthesis and metabolism: 7 PCs. Carbohydrate metabolism: 15 PCs. Energy metabolism: 7 PCs. Metabolism of cofactors and vitamins: 12 PCs. Metabolism of terpenoids and polyketides: 11 PCs. Xenobiotics biodegradation and metabolism: 16 PCs. Lipid metabolism: 14 PCs. Metabolism of other amino acids: 8 PCs.


A)
PGA
PMP2
TAS1R1
B)
AQP8
PMP2
UGT2B17
C)
OR10A6
OR9A2
OR2D2
TAS1R1
DESR
RAE
RAE
REUR
MSCH
NLEI
NLAS

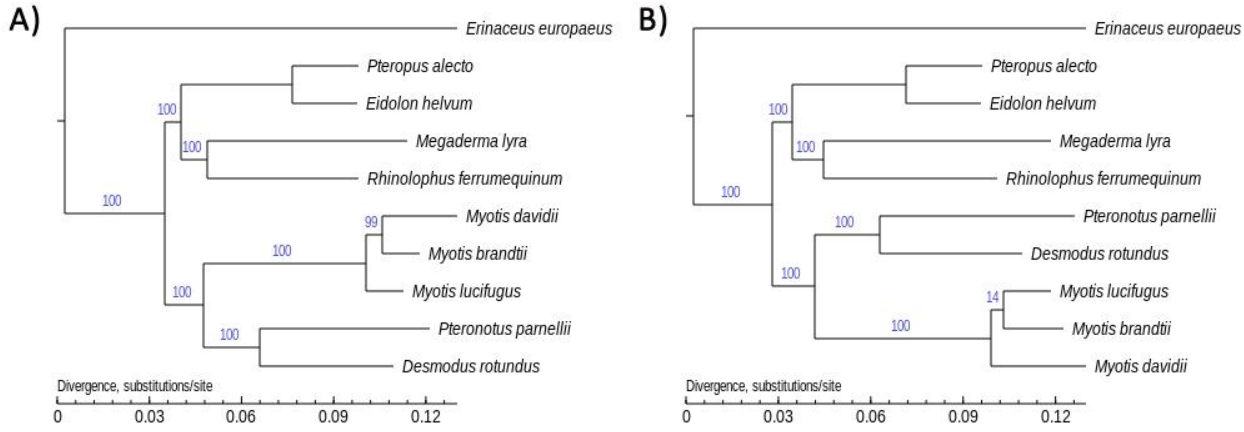
Supplementary 13. Putative loss of sequence/function in the common vampire bat genes related to sanguivorous traits. A) Synteny analyses of the genes PGA, PMP2, and TAS1R1. B) GC content of the syntenic regions of genes AQP8, PMP2, and UGT2B17. C) PCR amplification of the genes OR10A6, OR9A2, OR2D2, and TAS1R1, from the following bat species: *D. rotundus* (DESR), *R. aegyptiacus* (RAE), *Rhinolophus euryale* (REUR), *Miniopterus schreibersii* (MSCH), *Nyctalus leisleri* (NLEI), and *Nyctalus lasiopterus* (NLAS).



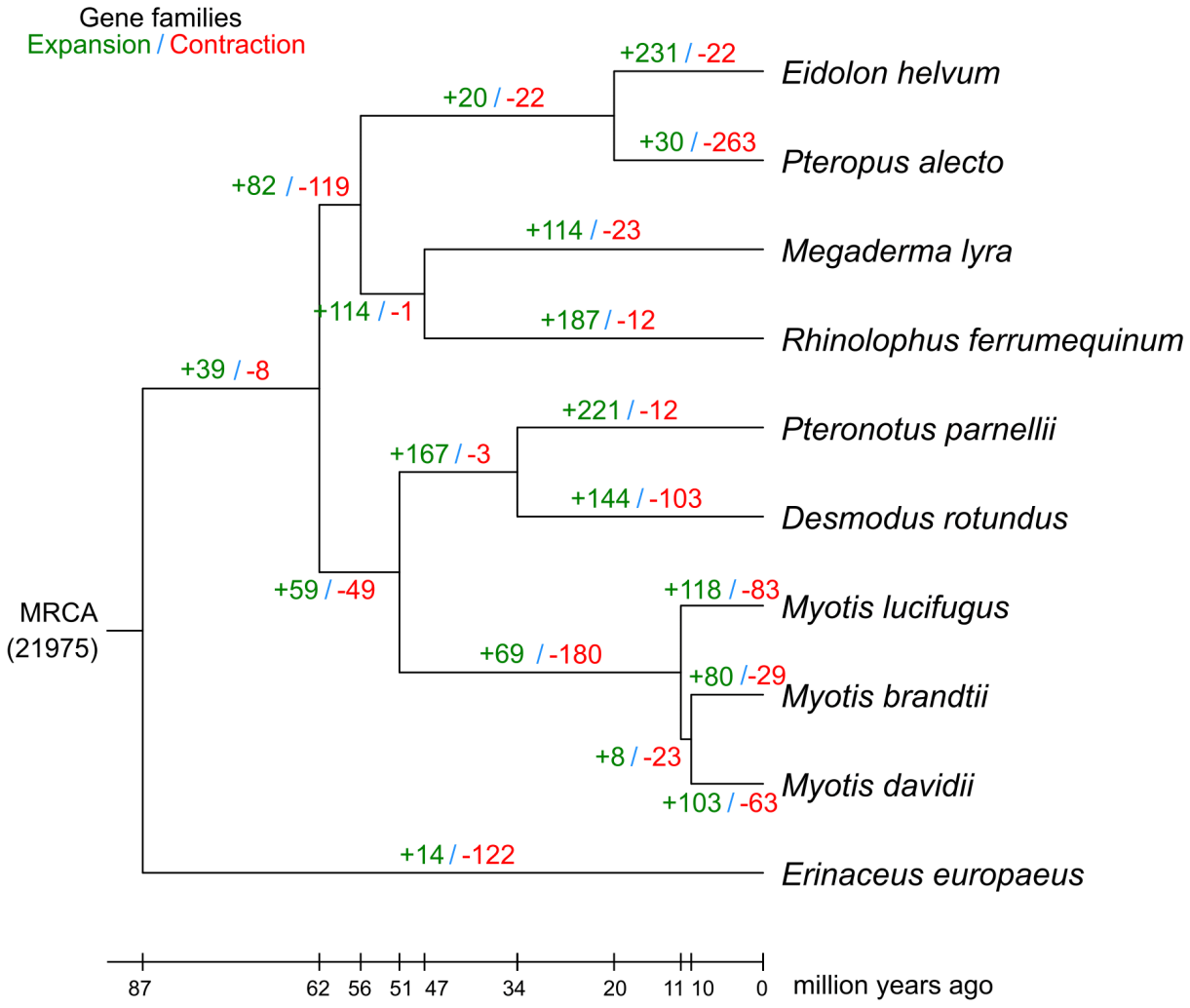
Supplementary Figure 14. Gene family identification. Protein orthology comparisons of predicted genes among *Eidolon helvum, Pteropus alecto, Megaderma lyra, Rhinolophus ferrumequinum, Pteronotus parnellii, Desmodus rotundus, Myotis lucifugus, Myotis brandtii, Myotis davidii*, and *Erinaceus europaeus*.



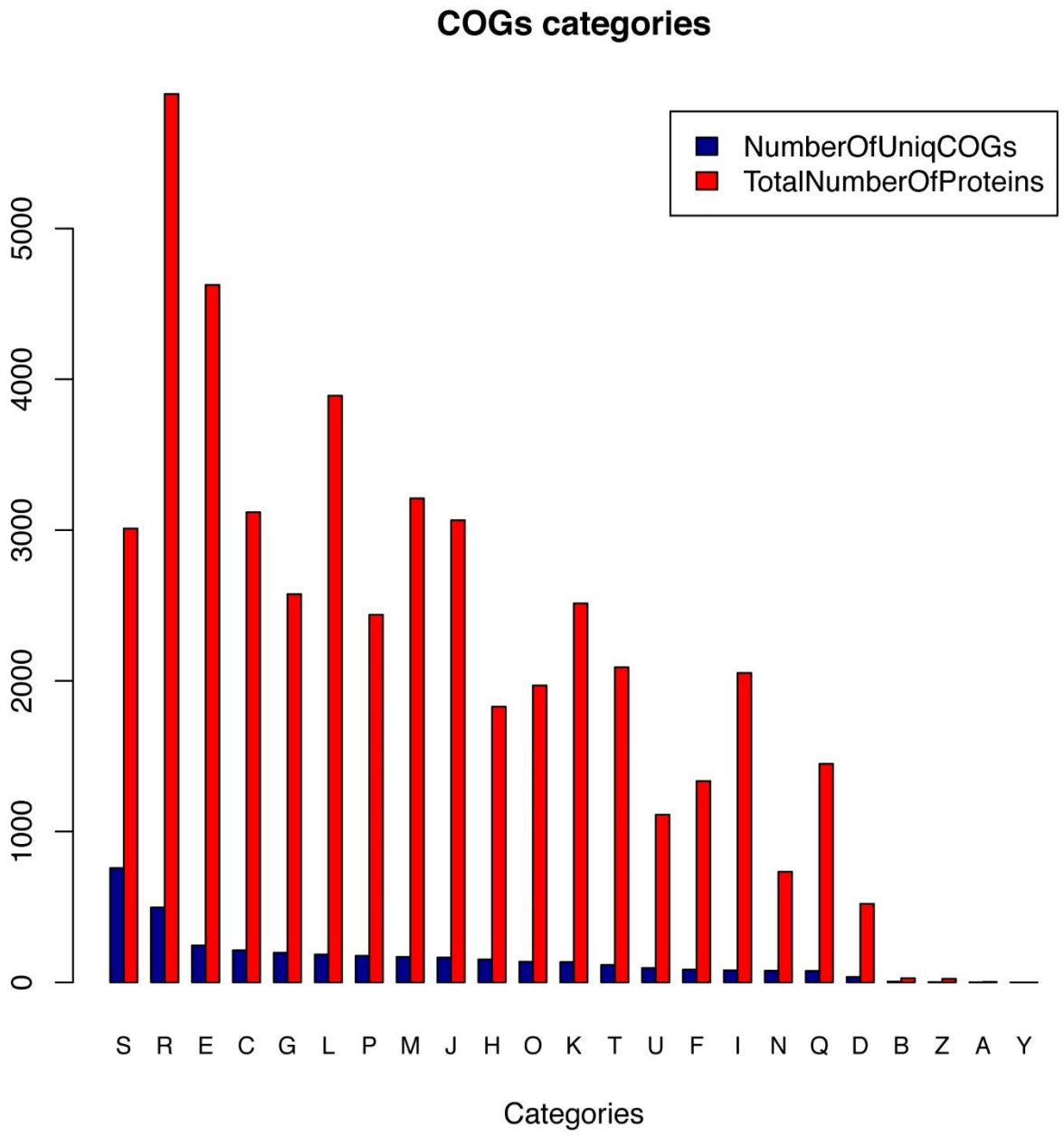
Supplementary Figure 15. Divergence time estimation. Performed for 10 species based on 3,384 single-copy orthologs. The time of divergence was estimated with the error range shown in parentheses. The red points on two of the internal nodes indicate fossil calibration times used in the analysis.



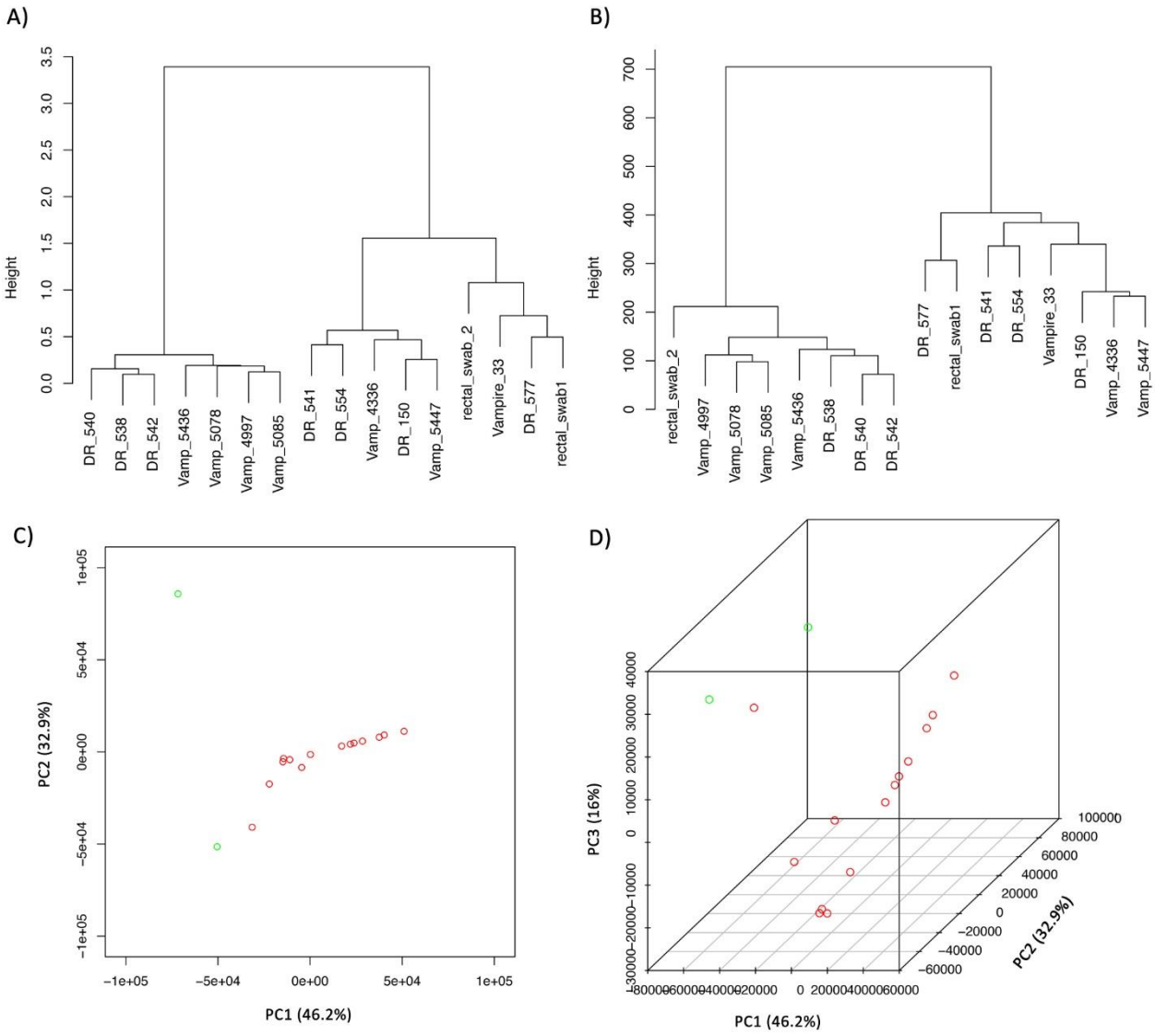
Supplementary Figure 16. Phylogenetic reconstruction. Phylogenetic tree of 10 species constructed with single-copy orthologous genes. (**A**) Nucleotide and (**B**) amino acid sequences of single-copy genes were used to construct the phylogenetic tree under HKY85+gamma model and WAG+gamma model, respectively. The bootstrap values were calculated based on 1,000 replicates.



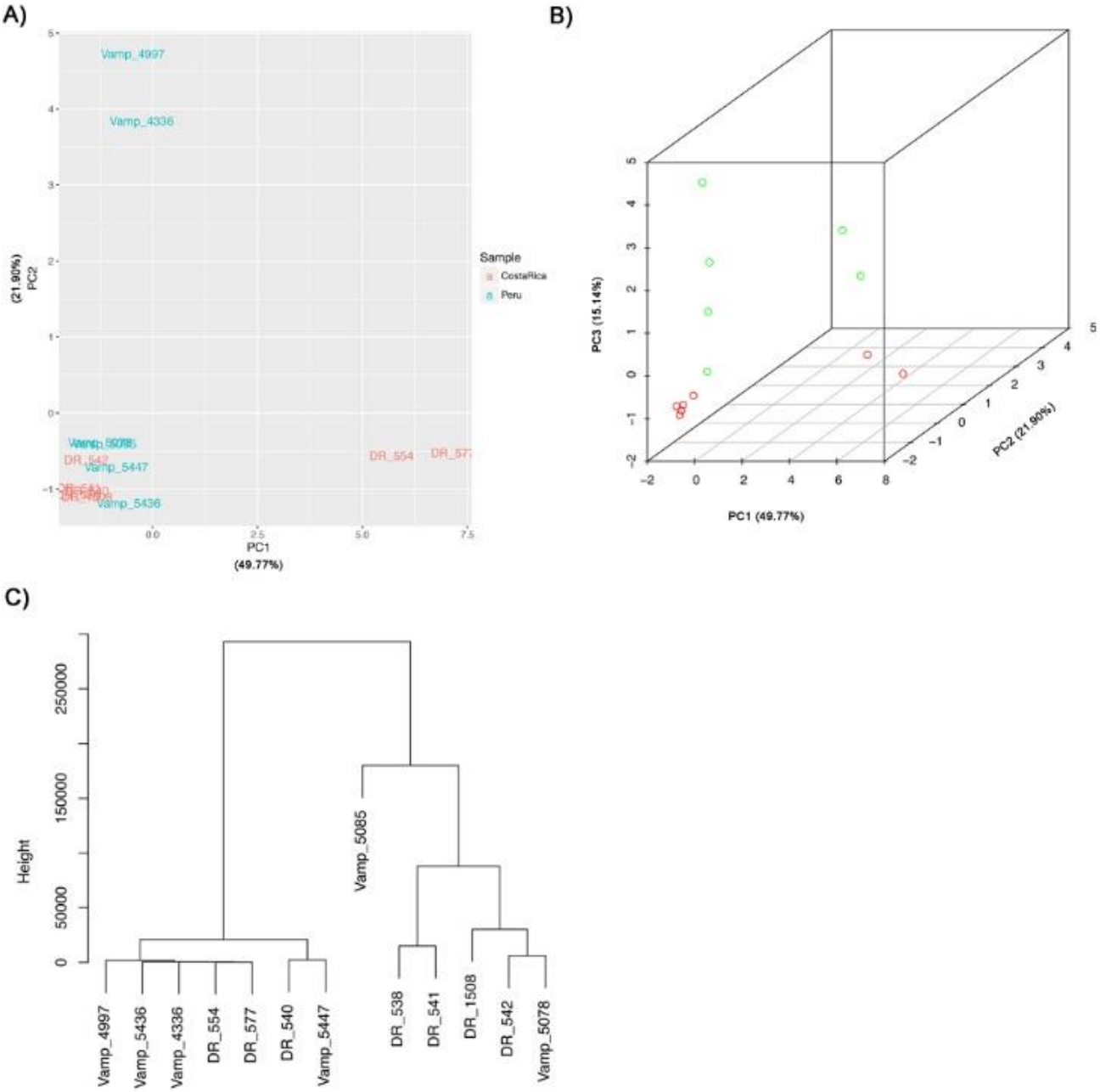
Supplementary Figure 17. Gene family expansion and contraction in the genomes. Gene family expansions are indicated in green, and gene family contractions are indicated in red.



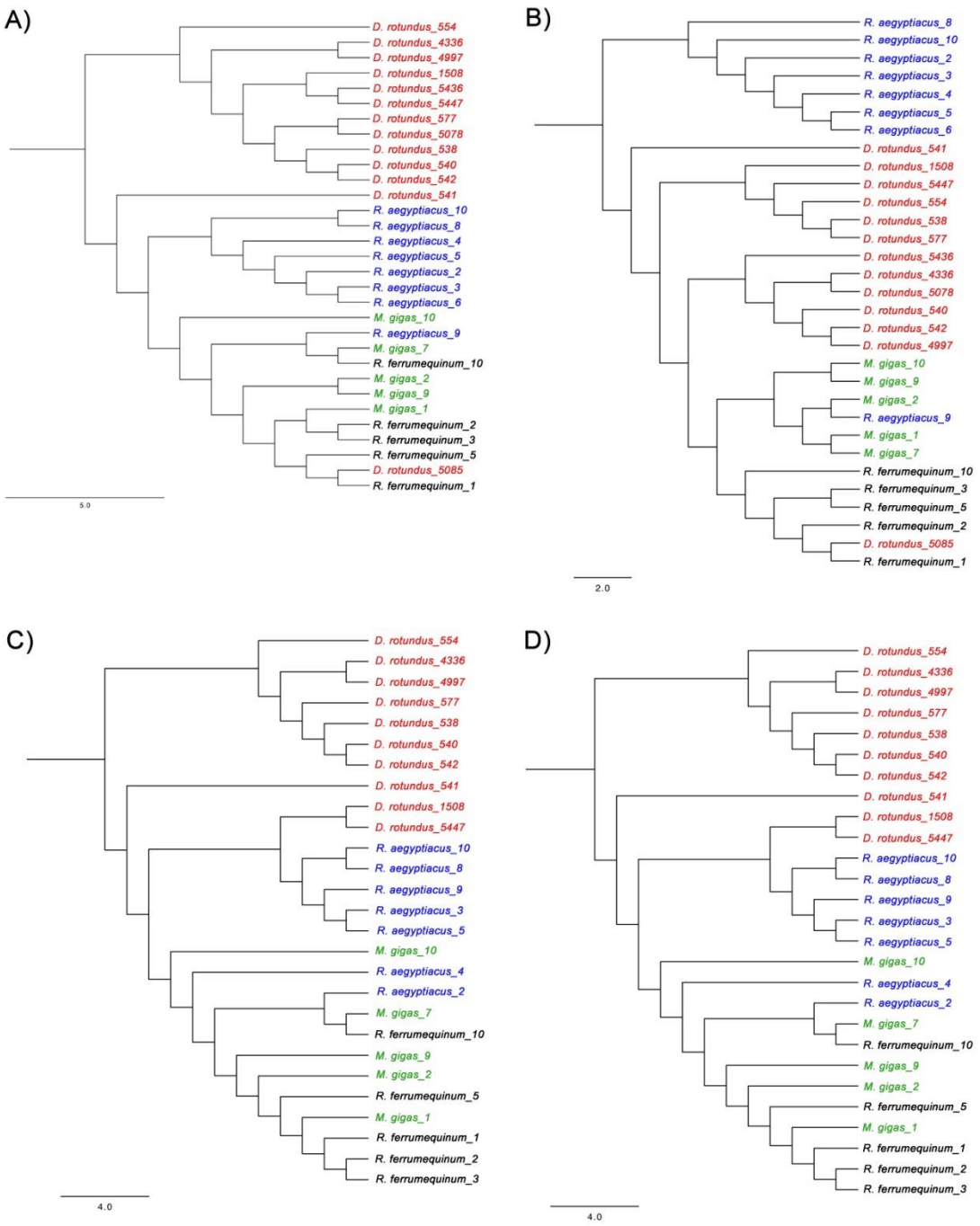
Supplementary Figure 18. Abundance of proteins per COG category in the *D. rotundus* gut microbiome functional core.



Supplementary Figure 19. Comparison of rectal swabs and fecal samples. (**A**) Bray-Curtis distance cladrogram of the *D. rotundus* samples. (**B**) Euclidean distance cladrogram of the *D. rotundus* samples. (**C**) PC1 vs PC2 plot from the PCA of the fecal (red) and rectal swab (green) samples. (**D**) 3D plot from the PCA of the fecal (red) and rectal swab (green) samples. The clustering of the samples confirm the similarity of the microbial compositions from the rectal swabs to the fecal samples.



Supplementary Figure 20. Sampling site and prey abundance influence on the clustering of the common vampire bat samples. Sample ids starting with "DR" come from Costa Rica, and those with "Vamp" come from Peru. (**A**) PCA of the normalized abundance of the putative preys identified from the MGmapper results using A) the first 2 PCs and (**B**) the first 3 PCs out of 13 PCs. (**C**) Hierarchical clustering with the ward.D method of the Euclidean distances of the normalized abundance of the putative preys identified from the MGmapper results.



Supplementary Figure 21. Midpoint-rooted dendrogram of plasmid taxa minimum abundance subsampling for normalization presence/absence. (**A**) Species level (**B**) Genera level (**C**) Bacterial taxa abundance min. subsampled for normalization at the species level (**D**) Genera level. *D. rotundus* (red), *R. ferrumequinum* (black), *M. gigas* (green), and *R. aegyptiacus* (blue).

Supplementary Tables

Supplementary Table 1. Dovetail genome contiguity refinement.

	Starting assembly		First Dovetail refinement		Final Dovetail refinement	
	Scaffold	**Contig**	**Scaffold**	**Contig**	**Scaffold**	**Contig**
Genome size	2.09 Gb	2.04 Gb	2.09 Gb	2.04 Gb	2.09 Gb	2.04 Gb
N50	5.5 Mb	36.6 Kb	10.17 Mb	36.6 Kb	26.9 Mb	36.6 Kb
N90	933 Kb	8.8 Kb	2.21 Mb	8.8 Kb	9.46 Mb	8.8 Kb
Longest length	21.45 Mb	382.1 Kb	43.18 Mb	382.1 Kb	73.32 Mb	382.1 Kb
Shortest length	100 nt	29 nt	100 nt	22 nt	100 nt	22 nt
Total	96,339	212,614	95,269	212,601	94,787	212,627

Supplementary Table 2. Genomes used for the genomic and comparative genomic analyses.

Species	**Abbreviation**	**Family**	**Accession ID**	**Scaffold N50**	**Contig N50**
Eidolon helvum	EIDH	Pteropodidae	ASM46528v1	27.68 Kb	12.67 Kb
Pteropus alecto	PTEL	Pteropodidae	ASM32557v1	15.95 Mb	31.8 Kb
Pteropus vampyrus	PTEV	Pteropodidae	pteVam1.73	182.95 Kb	8.52 Kb
Megaderma lyra	MEGL	Megadermatidae	ASM46534v1	16.8 Kb	7.04 Kb
Rhinolophus ferrumequinum	RHIF	Rhinolophidae	ASM46549v1	21.15 Kb	11.66 Kb
Pteronotus parnellii	PTEP	Mormoopidae	ASM46540v1	22.67 Kb	9.50 Kb
Desmodus rotundus	DESR	Phyllostomidae	NA	26.9 Mb	36.6 Kb
Myotis brandtii	MYOB	Vespertilionidae	ASM41265v1	3.22 Mb	23.29 Kb
Myotis davidii	MYOD	Vespertilionidae	ASM32734v1	3.45 Mb	15.18 Kb
Myotis lucifugus	MYOL	Vespertilionidae	Myoluc2.0	4.29 Mb	64.33 Kb
Equus caballus	HORS	Equidae	EquCab2.75	87.72 Mb	112.38 Kb
Bos taurus	COW	Bovidae	UMD3.1.75	105.7 Mb	96.01 Kb
Erinaceus europaeus	HEDG	Erinaceidae	EriEur2.0	3.26 Mb	21.36 Kb
Homo sapiens	HUMAN	Hominidae	GRCh37.73	146.3 Mb	3.68 Mb

Supplementary Table 3. Number of identified repeat elements in *D. rotundus* compared to other bat species.

Repeat element	***Desmodus rotundus***	***Pteronotus parnelli***	***Megaderma lyra***	***Pteropus vampyrus***
DNA/hAT-Tip100	30,818 (6,463,282 nts)	25,786 (5,379,387 nts)	35,484 (6,827,113 nts)	27,428 (6,078,449 nts)
DNA/CMC-EnSpm	24,908 (2,063,585 nts)	13,083 (990,636 nts)	22,008 (1,614,509 nts)	33,073 (2,371,562 nts)
DNA/MULE-MuDR	7,634 (599,700 nts)	3,949 (358,072 nts)	3,374 (247,954 nts)	4,777 (424,707 nts)
DNA/PiggyBac	1,055 (293,710 nts)	1,803 (462,702 nts)	14,864 (2,371,246 nts)	4,180 (1,054,926 nts)
RC/Helitron	4,508 (514,969 nts)	6,323 (639,870 nts)	24,049 (3,003,815 nts)	12,818 (1,196,930 nts)
LTR/Lenti	21 (7,131 nts)	9 (3,558 nts)	13 (4,757 nts)	13 (3,548 nts)
DNA/Maverick	7,283 (571,625 nts)	5,650 (420,978 nts)	7,895 (613,353 nts)	12,476 (882,408 nts)
LTR/Gypsy-Cigr	4 (1,157 nts)	2 (355 nts)	4 (1,292 nts)	4 (1,142 nts)
LTR/DIRS	62 (4,897 nts)	764 (46,762 nts)	675 (38,512 nts)	502 (31,179 nts)
LINE/Tad1	17 (1,693 nts)	576 (219,824 nts)	132 (8,115 nts)	37 (2,243 nts)
LINE/RTE	517 (29,226 nts)	292 (30,331 nts)	324 (32,616 nts)	770 (87,000 nts)
LINE/RTE-BovB	4,595 (1,409,222 nts)	3,755 (1,204,492 nts)	3,323 (1,216,035 nts)	3,127 (1,186,047 nts)
LINE/RTE-X	5,537 (1,365,789 nts)	6,964 (1,643,376 nts)	6,900 (1,623,750 nts)	7,514 (1,864,055 nts)
LINE/RTE-RTE	32 (9,299 nts)	13 (1,131 nts)	25 (5,532 nts)	9 (2,772 nts)
LINE/L1	1,965,656 (842,009,437 nts)	1,943,737 (729,416,451 nts)	1,419,305 (531,741,641 nts)	1,797,654 (703,681,578 nts)

Supplementary Table 4. Sites unique to *D. rotundus* node mapped onto the protein 3D models.

Protein	Site	*Desmodontinae*	Chiroptera	Score	Importance
FFAR1	73	T	P	-0.606 (Provean)	Topological domain
	178	V	A	-0.573 (Provean)	Topological domain
	182	H	R	-4.885 (Provean)	Binding site
	295	S	G	-0.075 (Provean)	Topological domain
REG4	39	L	S/P/G/K	0.982 (BEB)	PSS
	46	Q	R	-2.839 (Provean)	Topologcal domain
	103	R	Q	-3.244 (Provean)	Carbohydrate-binding region
RNASE7	42	E	H	-7.871 (Provean)	Disulfide bond
	80	T	A	-0.139 (Provean)	Disulfide bond
	104	K	T/N	-0.945 (Provean)	Disulfide bond
	113	R	P	-8.294 (Provean)	Disulfide bond
TA2R3	11	D	Q	0.933 (BEB)	PSS
	111	N	G	0.988 (BEB)	PSS
	177	A	L/T/S/V	0.969 (BEB)	PSS
	293	Q	H	-4.415 (Provean)	Topological domain

Supplementary Table 5. Median and mean values of the normalized read counts of the *D. rotundus* samples from the significant most abundant bacterial genera and bacterial species taxa (*P*<0.05) that contribute to most of the variation between *D. rotundus* and the other three bat species.

Taxonomic level	Taxa	Median	Mean
Genera	*Thermomonospora*	25.0	76.17
	Actinosynnema	29.0	73.33
	Frankia	34.0	175.90
	Amycolatopsis	38.5	252.20
	Bartonella	38.5	88.50
	Singulisphaera	43.0	129.90
	Hyphomicrobium	320.0	1,981.00
	Wolinella	468.0	2,215.00
Species	*Lactobacillus sanfranciscensis* TMW 1.1304	26.5	299.20
	Salmonella enterica subsp. Enterica	26.5	195.50
	Actinoplanes friuliensis DSM 7358	29.0	73.33
	Listeria monocytogenes str. ATCC 19117	29.5	73.58
	Clostridium acetobutylicum EA 2018	29.5	1,443.00
	Escherichia albertii KF1	30.5	92.08
	Legionella oakridgensis ATCC 33761	32.5	363.20
	Listeria monocytogenes str. 4b F2365	33.0	77.50
	Clostridium acetobutylicum DSM 1731	34.0	1,706.00
	Oceanobacillus iheyensis HTE831	37.0	244.50
	Alteromonas sp. SN2	38.5	1,892.00
	Lactobacillus salivarius CECT 5713	39.0	410.40
	Pseudovibrio sp. FO-BEG1	43.0	129.90
	Lactobacillus reuteri SD2112	48.0	823.80
	Listeria monocytogenes 07PF0776	51.0	156.30
	Francisella tularensis TI0902	53.0	346.20
	Listeria monocytogenes 08-5923	56.5	88.67
	Francisella tularensis TIGB03	58.5	893.60
	Listeria monocytogenes serotype 4b	64.0	96.92
	Francisella tularensis	64.0	591.90
	Amphibacillus xylanus NBRC 15112	65.5	2,058.00
	Leuconostoc kimchii IMSNU 11154	83.0	94.00
	Lactococcus garvieae Lg2	85.5	618.80
	Listeria monocytogenes serotype 7	88.5	126.50
	Amycolatopsis mediterranei RB	92.0	2,194.00
	Lactococcus lactis subsp. cremoris UC509.9	98.5	528.40
	Lactococcus lactis subsp. lactis Il1403	101.0	542.50
	Lactobacillus rhamnosus LOCK908	122.5	831.70

	Methylococcus capsulatus str. Bath	171.0	443.10
	Methylocystis sp. SC2	174.0	339.10
	Listeria ivanovii subsp. ivanovii PAM 55	180.5	272.40
	Vibrio anguillarum M3	190.0	263.60
	Stenotrophomonas maltophilia R551-3	468.0	2,215.00
	Listeria monocytogenes str. SLCC2376	628.0	833.30
	Melissococcus plutonius DAT561	646.5	6,166.00

Supplementary Table 6. Median and mean values of the normalized counts of the *D. rotundus* samples from the pathways that contribute to most of the variation between *D. rotundus* and the other three bat species.

Pathway	**Median**	**Mean**
Primary bile acid biosynthesis	0.5	1.667
Isoflavonoid biosynthesis	1.5	1.500
Caffeine metabolism	2.0	1.917
Fatty acid elongation	2.5	7.917
Cutin suberine and wax biosynthesis	2.5	3.750
Arachidonic acid metabolism	3.0	2.417
Styrene degradation	3.0	5.000
Linoleic acid metabolism	4.0	4.917
Steroid hormone biosynthesis	4.5	5.583
Geraniol degradation	6.0	17.750
Fluorobenzoate degradation	16.5	17.670
Biosynthesis of unsaturated fatty acids	17.5	21.080
Indole diterpene alkaloid biosynthesis	18.5	18.750
Aflatoxin biosynthesis	20.0	20.420
Glucosinolate biosynthesis	20.5	21.420
Caprolactam degradation	22.5	32.580
Biosynthesis of siderophore group nonribosomal peptides	23.5	24.420
Ubiquinone and other terpenoid-quinone biosynthesis	41.5	42.250

Supplementary Table 7. Euclidean distances of the gut bacterial taxonomic and functional profiles with all the samples excluding the 4 with the highest depth of sequencing (1 *D. rotundus* and 3 *R. aegyptiacus*).

Functional profiles normalized				
	D. rotundus	*M. gigas*	*R. aegyptiacus*	*R. ferrumequinum*
D. rotundus	7.363	16.146	16.540	17.324
M. gigas	16.146	15.406	20.513	19.399
R. aegyptiacus	16.540	20.513	17.194	19.399
R. ferrumequinum	17.324	19.399	20.952	16.765
Bacteria profiles normalized minimum subsampling				
	D. rotundus	*M. gigas*	*R. aegyptiacus*	*R. ferrumequinum*
D. rotundus	44.199	51.977	52.890	43.171
M. gigas	51.977	41.892	54.237	45.516
R. aegyptiacus	52.890	54.237	41.391	46.149
R. ferrumequinum	43.171	45.516	46.149	28.755
Plasmid species profiles normalized minimum subsampling				
	D. rotundus	*M. gigas*	*R. aegyptiacus*	*R. ferrumequinum*
D. rotundus	54.441	52.587	56.234	49.588
M. gigas	52.587	27.839	41.129	30.887
R. aegyptiacus	56.234	41.129	36.452	37.988
R. ferrumequinum	49.588	30.887	37.988	20.513

Supplementary Table 8. Phyla abundance on the pooled genes from the samples of the 4 bat species.

	M. gigas		***R. aegyptiacus***		***R. ferrumequinum***		***D. rotundus***	
Phyla	**Uniprot**	**Diamond**	**Uniprot**	**Diamond**	**Uniprot**	**Diamond**	**Uniprot**	**Diamond**
Proteobacteria	0.436	0.265	0.837	0.845	0.650	0.333	0.442	0.522
Firmicutes	0.405	0.264	0.035	0.033	0.315	0.545	0.177	0.155
Bacteroidetes	0.112	0.387	0.082	0.062	0.0004	0.038	0.170	0.082
Actinobacteria	0.0006	0.028	0.021	0.035	0.0009	0.022	0.103	0.139

Supplementary Table 9. Top most abundant genera summary from the *D. rotundus* gut microbiome core manual annotation.

Category	Trait	Functions	Taxa	Top taxa
Directly	Coagulation	8	67	7 *Clostridium*; 4 *Streptococcus* and *Staphylococcus*
	Immune	18	88	23 *Streptomyces*; 9 *Staphylococcus*; 7 *Enterobacteria*
	Iron	7	59	4 *Cronobacter*; 3 *Geobacter*
	Nutrient	54	65	5 *Streptomyces* and *Bacillus*
	Fat	33	22	3 *Bacteroides*
	Kidney	10	36	3 *Bacillus*
	Hormone	8	1	
	Relevant	31	68	5 *Borrelia*; 4 *Bacillus* and *Aeromonas*
	TOTAL	169	406	
Indirectly	Salt	2	106	14 *Bacillus*; 8 *Nocardiopsis*
	Aromatic	9	8	
	Colon	5	54	13 *Enterococcus*; 6 *Bacteroides*
	Sulfur	14	64	7 *Desulfovibrio*; 5 *Desulfotomaculum*
	Methanol	10	51	4 *Methanobrevibacter* and *Clostridium*; 3 *Methanosarcina* and *Methanococcus*
	CO_2H_2	16	52	
	Other medium	3	59	4 *Geobacter*
	Oral	0	31	4 *Prevotella*
	TOTAL	59	425	
General	Industry	3	23	4 *Bacillus*
	Pathogen	7	281	22 *Mycoplasma*; 20 *Staphylococcus* and *Nocardia*; 10 *Clostridium* and *Bartonella*
	Hematophagous insects	0	25	8 *Rickettsia*; 4 *Borrelia*
	TOTAL	10	329	

Supplementary Table 10. Taxonomic and functional nr gene catalogue profiling for the *D. rotundus* gut microbiome core.

Database	**Total *D. rotundus***	**Excluding *M. gigas***	**Excluding *R. aegiptiacus***	**Excluding *R. Ferrumequinum***	**Excluding all non vampires**
Protozoa	6	6	4	6	4
Invertebrates	5	5	3	4	3
Unmapped reads genera	684	556	184	587	168
Nr gene set	456	408	104	422	101
Virus	342	292	201	297	197
Fungi	4	4	3	4	3
Plasmid	2,111	1,958	1,574	1,996	1,501
Silva	4,289	4,196	3,391	4,150	3,338
HumanMicrobiome	218	202	112	216	62
MetaHitAssembly	189	187	104	189	92
Bacteria	4,800	4,526	2,257	4,625	2,055
Nr gene set catalogue	415,583	415,296	386,835	414,412	386,038
Nr gene set catalogue KEGG	1,927	1,927	1,902	1,926	1,900
Nr gene set catalogue eggNOG	10,042	10,042	9,726	10,037	9,722

Supplementary Table 11. Detected genes from the single-copy orthologs evolving with a higher dN/dS ratio in *D. rotundus* compared to the other bats with 10% FDR.

Gene ID	Gene name	ω0	ω1	ω2	P-value	Q-value
Vampire_bat_11671	SMARCD1	0.0001	0.00543	0.08226	1.31E-12	3.90E-09
Vampire_bat_22236	MYRFL	0.03931	0.01695	0.89548	6.35E-09	9.42E-06
Vampire_bat_24614	CCDC64B	0.07315	0.06103	0.53558	1.94E-05	0.014377191
Vampire_bat_05733	ITGA10	0.08181	0.06825	0.8264	1.47E-05	0.014529293
Vampire_bat_26545	PLAT	0.28208	0.24233	0.77945	3.67E-05	0.021787091
Vampire_bat_04093	ZNF407	0.25601	0.24348	0.42548	0.0002234	0.051048792
Vampire_bat_15670	PGAP1	0.20771	0.16894	2.14822	0.0002093	0.051801656
Vampire_bat_19358	OR1F12	0.20486	0.17394	0.68404	0.0001764	0.05240221
Vampire_bat_03280	MOS	0.15963	0.14553	1.06528	0.0002039	0.055049605
Vampire_bat_23907	Pigs	0.16073	0.13516	0.37568	0.0001707	0.056335093
Vampire_bat_15560	PLEKHM3	0.12875	0.11531	0.60017	0.0003807	0.059502939
Vampire_bat_17497	VPS11	0.04755	0.04143	0.16511	0.0003805	0.062789173
Vampire_bat_11511	REP15	0.24208	0.20802	0.82159	0.0003667	0.068074894
Vampire_bat_21132	OR4F6	0.22891	0.18904	1.62375	0.0004842	0.068472841
Vampire_bat_11097	RNF6	0.24755	0.2322	0.63193	0.0006594	0.085142168
Vampire_bat_08451	CDCA7L	0.16176	0.13768	0.60428	0.0007492	0.092715616
Vampire_bat_22893	EME1	0.36586	0.33084	0.99053	0.0008097	0.096190546

Supplementary Table 12. Positively selected genes (PSGs) in single-copy orthologs in *D. rotundus* with 10% FDR.

Gene_ID	Gene Name	P-value	Q-value
Vampire_bat_13220	PPP1R3C	0	0
Vampire_bat_26039	RABGGTA	0	0
Vampire_bat_21299	RIOK1	0	0
Vampire_bat_26191	SACS	0	0
Vampire_bat_15645	N/A	0	0
Vampire_bat_03528	ZHX2	0	0
Vampire_bat_22798	PNMT	6.17E-14	2.62E-11
Vampire_bat_13394	SV2B	3.95E-13	1.46E-10
Vampire_bat_20887	TRPM7	6.31E-12	2.08E-09
Vampire_bat_13576	RAG2	1.05E-11	2.83E-09
Vampire_bat_00412	OR51G2	9.69E-12	2.88E-09
Vampire_bat_11750	Map3k12	2.67E-11	6.60E-09
Vampire_bat_03919	XPO7	3.17E-11	7.24E-09
Vampire_bat_10736	ZBTB5	6.43E-11	1.36E-08
Vampire_bat_10696	TEX10	2.02E-10	3.99E-08
Vampire_bat_24161	Gucy2e	1.45E-09	2.69E-07
Vampire_bat_23969	SCARF1	3.65E-09	6.37E-07
Vampire_bat_17120	Nup85	4.18E-09	6.89E-07
Vampire_bat_23049	DDX42	9.99E-09	1.56E-06
Vampire_bat_21870	VPRBP	1.51E-08	2.24E-06
Vampire_bat_25679	RASGRP4	9.31E-08	1.32E-05
Vampire_bat_24930	Fst	3.91E-07	5.28E-05
Vampire_bat_06316	N/A	1.94E-06	0.000250814
Vampire_bat_17938	ZDHHC16	4.32E-06	0.000534441
Vampire_bat_08611	GAB3	5.22E-06	0.00062006
Vampire_bat_24037	CYB5D2	8.41E-06	0.000924252
Vampire_bat_00558	SLC6A1	8.31E-06	0.000948329
Vampire_bat_17475	TNFAIP6	1.10E-05	0.001163268
Vampire_bat_18124	MCU	2.72E-05	0.00278369

Vampire_bat_26089	PRMT5	3.82E-05	0.003776925
Vampire_bat_14714	FAM69A	4.87E-05	0.004659567
Vampire_bat_18761	BRF2	8.06E-05	0.007479232
Vampire_bat_03059	PEX3	9.61E-05	0.008643142
Vampire_bat_06447	YTHDF2	0.000161763	0.014121
Vampire_bat_25211	APOBR	0.000207019	0.017555181
Vampire_bat_05795	LAX1	0.00047671	0.038239876
Vampire_bat_27133	N/A	0.000475689	0.039217875
Vampire_bat_10767	RGP1	0.000512638	0.040039713
Vampire_bat_20490	HEATR4	0.000663599	0.0505016
Vampire_bat_08988	B3GNT2	0.000747052	0.055431294
Vampire_bat_06904	FBXO4	0.000940489	0.066461245
Vampire_bat_23033	TLK2	0.00091841	0.066483912
Vampire_bat_05340	INPP5J	0.000967225	0.066761035
Vampire_bat_26696	TPP2	0.001001536	0.067558171
Vampire_bat_10797	KIAA1045	0.001270689	0.083809015
Vampire_bat_08438	TWISTNB	0.001382559	0.089205107
Vampire_bat_06713	Trim23	0.001496763	0.094518998
Vampire_bat_20521	SUSD6	0.001568581	0.096990608

*N/A indicates not available gene name.

Supplementary Table 13. Detected genes from the 1:1 orthologs evolving with a higher dN/dS ratio in *D. rotundus* compared to the other bats with 10% FDR

Gene ID	Gene name	ω0	ω1	ω2	P-value	Q-value
Vampire_bat_08827	SNX17	0.0001	0.02807	0.03032	0	0
Vampire_bat_07711	ANKRA2	0.08017	0.02182	0.49861	0	0
Vampire_bat_17285	ILF3	0.15509	0.10807	1.10499	0	0
Vampire_bat_20421	Chmp1b2	0.02702	0.01379	2.32151	0	0
Vampire_bat_02545	Snrpd1	0.12038	0.0158	3.98895	1.49E-10	1.05E-07
Vampire_bat_00485	N/A	0.15709	0.08175	1.70575	2.35E-10	1.45E-07
Vampire_bat_26631	RBM39	0.11376	0.07028	0.77079	1.80E-09	9.91E-07
Vampire_bat_08608	OR10V1	0.18763	0.14767	1.16836	1.70E-08	7.62E-06
Vampire_bat_23919	FAM222B	0.14733	0.10322	0.88817	5.55E-08	2.29E-05
Vampire_bat_05505	Rbm22	0.21155	0.17257	0.99596	3.49E-07	0.000123108
Vampire_bat_24065	C1QBP	0.16281	0.10694	0.72856	5.65E-07	0.000174647
Vampire_bat_10391	Nup54	0.3198	0.23459	2.03908	1.98E-06	0.000465071
Vampire_bat_24614	CCDC64B	0.06996	0.05943	0.55778	7.91E-06	0.001449117
Vampire_bat_15017	EXOSC2	0.27023	0.2008	1.02932	1.29E-05	0.002284975
Vampire_bat_22430	Napg	0.09961	0.06168	0.57464	2.08E-05	0.003540426
Vampire_bat_00775	PSMA2	0.12176	0.08791	0.49351	2.15E-05	0.003546725
Vampire_bat_26545	PLAT	0.28468	0.24793	0.78475	2.92E-05	0.004655046
Vampire_bat_18941	RLN3	0.26573	0.21021	1.1588	3.86E-05	0.005616617
Vampire_bat_05733	ITGA10	0.08517	0.07208	0.63909	3.80E-05	0.00568642
Vampire_bat_23907	Pigs	0.15288	0.12847	0.38644	5.21E-05	0.0062861
Vampire_bat_03302	NOL10	0.17154	0.11641	0.80738	7.76E-05	0.008717738
Vampire_bat_23732	Ybx2	0.49801	0.32324	1.28938	8.10E-05	0.008899223
Vampire_bat_15670	PGAP1	0.18454	0.15013	2.01103	8.80E-05	0.009461884
Vampire_bat_04093	ZNF407	0.26097	0.24945	0.44732	0.000110915	0.011191061
Vampire_bat_19112	EIF3K	0.09469	0.06063	0.85742	0.000133685	0.012710319
Vampire_bat_10470	Gstz1	0.26287	0.21642	1.80661	0.00016188	0.014551501
Vampire_bat_17497	VPS11	0.04788	0.04206	0.17329	0.000249476	0.020556819
Vampire_bat_23442	Csnk2b	0.00522	0.0001	0.05554	0.000258613	0.020960383
Vampire_bat_13943	LAPTM4A	0.45755	0.36648	2.51442	0.000301429	0.022579789

Vampire_bat_10601	Emd	0.23873	0.19323	3.78333	0.000287953	0.022597423
Vampire_bat_25578	Erf	0.0195	0.01115	0.16429	0.000331444	0.024097921
Vampire_bat_07729	Exosc1	0.3518	0.28491	1.33161	0.000392511	0.027722506
Vampire_bat_11511	REP15	0.25369	0.22305	0.86331	0.000450889	0.030536926
Vampire_bat_09925	Rab1b	0.07179	0.05784	0.47301	0.000502934	0.033153394
Vampire_bat_26455	CYSLTR2	0.50615	0.45309	1.97679	0.000548104	0.034741367
Vampire_bat_03328	MEST	0.43754	0.38364	1.43809	0.000547404	0.035147589
Vampire_bat_05155	Isy1	0.16122	0.09195	0.83485	0.000562247	0.035186678
Vampire_bat_26494	KCNRG	0.43637	0.33721	1.30573	0.000652823	0.03752971
Vampire_bat_22893	EME1	0.36772	0.33669	1.03887	0.00068204	0.037887717
Vampire_bat_15560	PLEKHM3	0.14798	0.13571	0.6457	0.000860726	0.045270545
Vampire_bat_03439	Sat1	0.36113	0.27751	1.03968	0.000907713	0.047239284
Vampire_bat_07661	ZFAND2B	0.28679	0.21664	1.00619	0.001071352	0.05142489
Vampire_bat_21057	THBS1	0.06461	0.05566	0.13682	0.001120401	0.051768803
Vampire_bat_08451	CDCA7L	0.17097	0.14906	0.58975	0.001037811	0.051827635
Vampire_bat_09053	FSHR	0.20499	0.18626	0.57273	0.001058973	0.051837268
Vampire_bat_03546	RNF139	0.09046	0.07375	0.21551	0.001033854	0.052694569
Vampire_bat_13627	PSAT1	0.32069	0.29294	0.90402	0.001197854	0.053353053
Vampire_bat_19692	COA7	0.1963	0.16032	0.70514	0.001299829	0.056870414
Vampire_bat_20474	VSX2	0.04851	0.0277	0.25706	0.001427587	0.059813466
Vampire_bat_26170	FREM2	0.13133	0.12209	0.26366	0.001534242	0.060682342
Vampire_bat_23453	Nfkbil1	0.32627	0.25366	1.77889	0.001527235	0.060892349
Vampire_bat_14740	PRPF38B	0.26816	0.2225	0.53538	0.001563279	0.061340076
Vampire_bat_11376	TAB3	0.10932	0.1004	0.50652	0.001666911	0.062909995
Vampire_bat_16301	PCDH12	0.19142	0.18187	0.381	0.001742219	0.064280087
Vampire_bat_26219	CCDC70	0.52145	0.43937	2.12073	0.00182644	0.065434206
Vampire_bat_08070	Hsd11b2	0.2239	0.2032	0.67462	0.001805777	0.065645313
Vampire_bat_10747	GLIPR2	0.11907	0.0901	0.97752	0.001862228	0.065763249
Vampire_bat_15992	Tmem230	0.25406	0.17296	0.61903	0.001795863	0.065768491
Vampire_bat_23729	TAF1D	0.21454	0.18334	2.03911	0.001908953	0.065999035
Vampire_bat_16303	PCDH1	0.04274	0.03832	0.12212	0.002046431	0.069298333
Vampire_bat_16904	Unc50	0.05991	0.04532	0.38472	0.002160716	0.071217206

Vampire_bat_03260	CDC123	0.034	0.00728	0.34215	0.002587744	0.082540691
Vampire_bat_25728	COMMD9	0.1641	0.13539	0.52195	0.002663658	0.083348905
Vampire_bat_18369	MARCH8	0.13011	0.11408	0.59485	0.002721178	0.083562139
Vampire_bat_03242	rffG	0.04795	0.06185	3.58448	0.00278732	0.084542997
Vampire_bat_11671	SMARCD1	0.00946	0.005	0.08366	0.00319331	0.091258537
Vampire_bat_12627	LAMTOR5	0.43483	0.3569	1.39965	0.0033163	0.093157869
Vampire_bat_08068	AGRP	0.31154	0.27438	2.3171	0.003373757	0.093183552
Vampire_bat_14913	EEF2	0.0064	0.0054	0.04457	0.003364215	0.093442023
Vampire_bat_20781	FBXL22	0.04731	0.04075	0.65973	0.003453769	0.093821056
Vampire_bat_21172	OR6S1	0.18863	0.16885	0.47586	0.003525183	0.095237726

Supplementary Table 14. PSGs in 1:1 orthologs in D. rotundus using the branch-site model with 10% FDR.

Gene_ID	Gene Name	P-value	Q-value
Vampire_bat_24065	C1QBP	0	0
Vampire_bat_23758	CREBRF	0	0
Vampire_bat_22638	PIH1D1	0	0
Vampire_bat_16482	RAD50	0	0
Vampire_bat_23985	Tsr1	9.99E-16	8.55E-13
Vampire_bat_08716	ODC1	1.33E-15	9.49E-13
Vampire_bat_22792	GRB7	1.15E-14	7.05E-12
Vampire_bat_05365	MTMR3	7.46E-14	3.99E-11
Vampire_bat_21197	METTL3	4.78E-13	2.27E-10
Vampire_bat_00944	HNF4G	2.46E-12	1.05E-09
Vampire_bat_21870	VPRBP	9.11E-12	3.54E-09
Vampire_bat_09919	CATSPER1	3.24E-11	1.16E-08
Vampire_bat_03919	XPO7	9.08E-11	2.77E-08
Vampire_bat_24147	WRAP53	8.46E-11	2.78E-08
Vampire_bat_27232	PIGT	4.69E-10	1.34E-07
Vampire_bat_13110	Fam131a	6.57E-10	1.76E-07
Vampire_bat_13292	GGPS1	8.40E-10	2.11E-07
Vampire_bat_05124	MAPKAPK5	2.88E-09	6.85E-07
Vampire_bat_23219	TJAP1	8.15E-09	1.83E-06
Vampire_bat_06527	Tfap2e	9.96E-09	2.13E-06
Vampire_bat_25147	FBXL19	1.91E-08	3.89E-06
Vampire_bat_05359	SF3A1	3.68E-08	6.84E-06
Vampire_bat_02580	RLIM	3.54E-08	6.89E-06
Vampire_bat_10165	TRMT1L	7.32E-08	1.30E-05
Vampire_bat_08540	Dnajb9	8.03E-08	1.37E-05
Vampire_bat_20565	ZBTB1	1.07E-07	1.77E-05
Vampire_bat_27646	GLDC	1.64E-07	2.59E-05
Vampire_bat_27255	PCIF1	4.51E-07	6.88E-05
Vampire_bat_00588	MTMR14	5.85E-07	8.63E-05

Vampire_bat_24505	RHOT2	4.28E-06	0.000610444
Vampire_bat_22591	TMEM143	8.14E-06	0.001123462
Vampire_bat_03815	CERS4	1.31E-05	0.001694769
Vampire_bat_07445	HOXD10	1.27E-05	0.001696695
Vampire_bat_17475	TNFAIP6	1.47E-05	0.001847017
Vampire_bat_09925	Rab1b	1.52E-05	0.00185616
Vampire_bat_00309	CCNT2	1.87E-05	0.002220276
Vampire_bat_09404	Snap23	2.16E-05	0.002494978
Vampire_bat_13750	USP47	4.94E-05	0.005555428
Vampire_bat_18103	UNC5B	5.40E-05	0.005925069
Vampire_bat_11043	ERCC1	7.17E-05	0.007662273
Vampire_bat_09283	PLEKHB1	7.71E-05	0.008041797
Vampire_bat_12343	GOLPH3L	9.45E-05	0.009616755
Vampire_bat_26041	TINF2	9.71E-05	0.009655629
Vampire_bat_03242	rffG	0.000112521	0.010692041
Vampire_bat_00438	OR52H1	0.000110166	0.0107061
Vampire_bat_05155	Isy1	0.000118501	0.011015464
Vampire_bat_08611	GAB3	0.000155246	0.014124077
Vampire_bat_25211	APOBR	0.000166803	0.014859373
Vampire_bat_03059	PEX3	0.000223266	0.019483377
Vampire_bat_19806	ATP6V0B	0.000242281	0.020313639
Vampire_bat_00570	BRK1	0.000238693	0.020413061
Vampire_bat_23720	CCNG1	0.000430578	0.03473872
Vampire_bat_19925	STOM	0.00042413	0.034876555
Vampire_bat_24563	RNF151	0.000480698	0.038064182
Vampire_bat_13610	Ndufb4	0.000491289	0.038195477
Vampire_bat_10747	GLIPR2	0.000574396	0.043859269
Vampire_bat_00664	HESX1	0.000792266	0.05553656
Vampire_bat_20490	HEATR4	0.000767459	0.055621244
Vampire_bat_05795	LAX1	0.000758999	0.055956578
Vampire_bat_05340	INPP5J	0.000788317	0.056180728
Vampire_bat_06713	Trim23	0.000757549	0.056829472

Vampire_bat_06316	N/A	0.000858394	0.05920151
Vampire_bat_06607	SLFNL1	0.000894411	0.059757806
Vampire_bat_11332	FAM133A	0.000882141	0.059873597
Vampire_bat_15976	DTX3L	0.001172267	0.077117128
Vampire_bat_23934	TP53I13	0.00137133	0.087519525
Vampire_bat_13532	N/A	0.001352722	0.087640009
Vampire_bat_08056	CENPT	0.001510609	0.09499065
Vampire_bat_11671	SMARCD1	0.001549655	0.096033714

Supplementary Table 15. Calibration times used in the divergence time estimation.

Species 1	**Species 2**	**Lower bound (Ma)**	**Upper bound (Ma)**
Eidolon helvum	*Myotis lucifugus*	59	65
Eidolon helvum+Myotis lucifugus	*Erinaceus europaeus*	82	92

Supplementary Table 16. GO enrichment of expanded gene families in *D. rotundus*.

GO_ID	GO_Term	GO_Class	Adjusted p-value	Gene NO.
GO:0005622	intracellular	CC	0	1259
GO:0003735	structural constituent of ribosome	MF	0	935
GO:0005840	ribosome	CC	0	935
GO:0006412	translation	BP	0	935
GO:0005198	structural molecule activity	MF	0	1048
GO:0005737	cytoplasm	CC	0	981
GO:0032991	macromolecular complex	CC	0	1191
GO:0030529	ribonucleoprotein complex	CC	0	946
GO:0043229	intracellular organelle	CC	0	1172
GO:0043232	intracellular non-membrane-bounded organelle	CC	0	1135
GO:0044424	intracellular part	CC	0	1203
GO:0044444	cytoplasmic part	CC	0	973
GO:0034645	cellular macromolecule biosynthetic process	BP	0	1012
GO:0044249	cellular biosynthetic process	BP	0	1032
GO:0044267	cellular protein metabolic process	BP	0	967
GO:1901576	organic substance biosynthetic process	BP	0	1032
GO:0044464	cell part	CC	2.74e-315	1304
GO:0010467	gene expression	BP	1.12E-277	1012
GO:0008199	ferric iron binding	MF	2.22E-268	172
GO:0006826	iron ion transport	BP	6.33E-263	172
GO:0000041	transition metal ion transport	BP	1.32E-257	173
GO:0006879	cellular iron ion homeostasis	BP	5.80E-256	172
GO:0044391	ribosomal subunit	CC	5.42E-243	183
GO:0044260	cellular macromolecule metabolic process	BP	4.72E-185	1053
GO:0015934	large ribosomal subunit	CC	7.74E-173	134
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as	MF	5.42E-139	100

	acceptor			
GO:0044237	cellular metabolic process	BP	1.96E-114	1071
GO:0044446	intracellular organelle part	CC	8.72E-107	400
GO:0044238	primary metabolic process	BP	7.48E-88	1071
GO:0071704	organic substance metabolic process	BP	1.07E-79	1071
GO:0030001	metal ion transport	BP	1.48E-76	174
GO:0005882	intermediate filament	CC	2.00E-76	102
GO:0009987	cellular process	BP	2.56E-71	1415
GO:0008152	metabolic process	BP	5.44E-71	1187
GO:0000786	nucleosome	CC	4.12E-69	78
GO:0015935	small ribosomal subunit	CC	5.32E-69	49
GO:0006812	cation transport	BP	5.98E-60	191
GO:0006811	ion transport	BP	3.29E-25	193
GO:0044430	cytoskeletal part	CC	5.79E-23	122
GO:0045095	keratin filament	CC	6.24E-18	29
GO:0043234	protein complex	CC	2.45E-13	245
GO:0003723	RNA binding	MF	4.03E-12	101
GO:0008097	5S rRNA binding	MF	1.33E-10	11
GO:0000160	phosphorelay signal transduction system	BP	2.00E-10	21
GO:0003924	GTPase activity	MF	4.80E-10	61
GO:0019843	rRNA binding	MF	2.05E-09	24
GO:0000276	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	CC	3.64E-09	17
GO:0005550	pheromone binding	MF	7.53E-08	8
GO:0042605	peptide antigen binding	MF	2.81E-07	12
GO:0003785	actin monomer binding	MF	5.66E-07	8
GO:0042612	MHC class I protein complex	CC	6.13E-07	12
GO:0005200	structural constituent of cytoskeleton	MF	6.96E-07	18
GO:0005852	eukaryotic translation initiation factor 3 complex	CC	1.25E-06	12

GO:0004984	olfactory receptor activity	MF	1.41E-06	92
GO:0032040	small-subunit processome	CC	1.45E-06	11
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	BP	2.37E-06	12
GO:0007156	homophilic cell adhesion	BP	5.86E-06	28
GO:0015986	ATP synthesis coupled proton transport	BP	3.18E-05	17
GO:0000413	protein peptidyl-prolyl isomerization	BP	6.92E-05	22
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	MF	6.92E-05	22
GO:0051920	peroxiredoxin activity	MF	7.93E-05	10
GO:0005874	microtubule	CC	0.000145733	20
GO:0046914	transition metal ion binding	MF	0.00138183	180
GO:0044765	single-organism transport	BP	0.002816891	194
GO:0055114	oxidation-reduction process	BP	0.010110334	116
GO:0016209	antioxidant activity	MF	0.011083248	13
GO:0007283	spermatogenesis	BP	0.012668871	9
GO:0003743	translation initiation factor activity	MF	0.028543578	12
GO:0005525	GTP binding	MF	0.028695668	68
GO:0015078	hydrogen ion transmembrane transporter activity	MF	0.0391379	26

Supplementary references

1. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1,** 18 (2012).
2. Liu, B. *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. 47 (2013).
3. Francischetti, I. M. B. *et al.* The 'Vampirome': Transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat Desmodus rotundus, a vector of human rabies. *J. Proteomics* **82,** 288–319 (2013).
4. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12,** 656–64 (2002).
5. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).
6. She, R., Chu, J. S.-C., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19,** 143–9 (2009).
7. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14,** 988–95 (2004).
8. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34,** W435-9 (2006).
9. The universal protein resource (UniProt). *Nucleic Acids Res.* **36,** D190-5 (2008).
10. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37,** D211–D215 (2009).
11. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
12. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-3.0.
13. Jurka, J. Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* **8,** 333–7 (1998).
14. Smit, A. & Hubley, R. RepeatModeler Open-1.0.
15. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–80 (1999).
16. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35,** W265-8 (2007).
17. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).
18. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–402 (1997).
19. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12,** 59–60 (2014).
20. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).
21. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–80 (2013).
22. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30,** 3276–8 (2014).
23. King, A., Lefkowitz, E., Adams, M. & Carstens, E. *Virus Taxonomy: Ninth Report of the International Committee of Taxonomy of Viruses*. (Elsevier, 2011).
24. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–90 (2006).

25. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).
26. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010,** pdb.prot5448 (2010).
27. Taboada, B. *et al.* Is There Still Room for Novel Viral Pathogens in Pediatric Respiratory Tract Infections? *PLoS One* **9,** e113570 (2014).
28. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
29. Huson, D. H. & Weber, N. Microbial community analysis using MEGAN. *Methods Enzymol.* **531,** 465–85 (2013).
30. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4,** Unit 4.10 (2009).
31. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16,** 418–20 (2000).
32. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–4 (2010).
33. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).
34. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25,** 1307–20 (2008).
35. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39,** 306–14 (1994).
36. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–21 (2010).
37. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 10.1101/gr.193474.115 (2016).
38. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–2 (2015).
39. Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502,** 228–31 (2013).
40. Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339,** 456–60 (2013).
41. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–82 (2011).
42. Seim, I. *et al.* Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. *Nat. Commun.* **4,** 2212 (2013).
43. Tang, Z. *et al.* Repeated horizontal transfers of four DNA transposons in invertebrates and bats. *Mob. DNA* **6,** 3 (2015).
44. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput. Viena, Austria* (2013).
45. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34,** D572-80 (2006).
46. Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M. & Bateman, A. TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic Acids Res.* **42,** D922-5 (2014).
47. Jarvis, E. D. *et al.* Whole Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science.* (2014).

48. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–91 (2007).
49. Teeling, E. C. *et al.* A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record. *Science.* **307,** (2005).
50. Tsagkogeorga, G., Parker, J., Stupka, E., Cotton, J. A. & Rossiter, S. J. *Phylogenomic Analyses Elucidate the Evolutionary Relationships of Bats*. *Current Biology* **23,** (2013).
51. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32,** 835–45 (2015).
52. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22,** 1269–71 (2006).
53. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* **346,** 1311–1320 (2014).
54. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA.* (ProQuest, 2007).
55. Elsik, C. G. *et al.* Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15,** 86 (2014).
56. Hargreaves, A. D. *et al.* Genome sequence of a diabetes-prone desert rodent reveals a mutation hotspot around the ParaHox gene cluster. *bioRxiv* (2016).
57. Choi, Y. & P. Chan, A. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31,** 2745–7 (2015).
58. Ranwez, V. *et al.* OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* **7,** 241 (2007).
59. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–7 (2004).
60. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8,** e1002764 (2012).
61. Yang, Z. & dos Reis, M. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* **28,** 1217–28 (2011).
62. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).
63. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523,** 561–7 (2015).
64. He, H., Yang, T., Terman, J. R. & Zhang, X. Crystal structure of the plexin A3 intracellular region reveals an autoinhibited conformation through active site sequestration. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15610–5 (2009).
65. Ho, M.-R. *et al.* Human RegIV Protein Adopts a Typical C-Type Lectin Fold but Binds Mannan with Two Calcium-Independent Sites. *J. Mol. Biol.* **402,** 682–695 (2010).
66. Huang, Y.-C. *et al.* The flexible and clustered lysine residues of human ribonuclease 7 are critical for membrane permeability and antimicrobial activity. *J. Biol. Chem.* **282,** 4626–33 (2007).
67. Streicker, D. G. & Allgeier, J. E. Foraging choices of vampire bats in diverse landscapes: potential implications for land-use change and disease transmission. *J. Appl. Ecol.* **53,** 1280–1288 (2016).
68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–20 (2014).
69. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–4 (2011).
70. Hasman, H. *et al.* Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *J. Clin. Microbiol.* **52,** 3136–3136 (2014).

71. Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32,** (2014).
72. Peterson, J. *et al.* The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–23 (2009).
73. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67,** 2640–4 (2012).
74. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–72 (2006).
75. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590-6 (2013).
76. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28,** 1420–8 (2012).
77. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11,** 119 (2010).
78. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–1 (2010).
79. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.3-5. (2016).
80. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–62 (2002).
81. Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428,** 493–521 (2004).
82. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–19 (2005).
83. Mikkelsen, T. S. *et al.* Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447,** 167–77 (2007).
84. Pagán, H. J. T. *et al.* Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. *Genome Biol. Evol.* **4,** 575–85 (2012).
85. Ray, D. A. *et al.* Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. *Genome Res.* **18,** 717–28 (2008).
86. Thomas, J., Sorourian, M., Ray, D., Baker, R. J. & Pritham, E. J. The limited distribution of Helitrons to vesper bats supports horizontal transfer. *Gene* **474,** 52–8 (2011).
87. Pritham, E. J. & Feschotte, C. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1895–900 (2007).
88. O'Donnell, K. A. & Boeke, J. D. Mighty Piwis Defend the Germline against Genome Intruders. *Cell* **129,** 37–44 (2007).
89. Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364,** 99–115 (2009).
90. Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26,** 391–406 (1998).
91. Goodwin, T. J. D., Butler, M. I. & Poulter, R. T. M. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* **149,** 3099–109 (2003).
92. Kojima, K. K. & Jurka, J. Crypton transposons: identification of new diverse families and ancient domestication events. *Mob. DNA* **2,** 12 (2011).

93. Zhuo, X., Rho, M. & Feschotte, C. Genome-Wide Characterization of Endogenous Retroviruses in the Bat Myotis lucifugus Reveals Recent and Diverse Infections. *J. Virol.* **87,** 8493–8501 (2013).
94. Cui, J. *et al.* Identification of diverse groups of endogenous gammaretroviruses in mega- and microbats. *J. Gen. Virol.* **93,** 2037–45 (2012).
95. Hayward, J. A. *et al.* Identification of diverse full-length endogenous betaretroviruses in megabats and microbats. *Retrovirology* **10,** 35 (2013).
96. Escalera-Zamudio, M. *et al.* A novel endogenous betaretrovirus in the common vampire bat (Desmodus rotundus) suggests multiple independent infection and cross-species transmission events. *J. Virol.* **89,** 5180–4 (2015).
97. Paprotka, T. *et al.* Recombinant origin of the retrovirus XMRV. *Science* **333,** 97–101 (2011).
98. Cui, J. *et al.* Bats and Rodents Shape Mammalian Retroviral Phylogeny. *Sci. Rep.* **5,** 16561 (2015).
99. Nisole, S., Stoye, J. P. & Saïb, A. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* **3,** 799–808 (2005).
100. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424,** 99–103 (2003).
101. Evans, D. T., Serra-Moreno, R., Singh, R. K. & Guatelli, J. C. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol.* **18,** 388–396 (2010).
102. Li, Y., Li, X., Stremlau, M., Lee, M. & Sodroski, J. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *J. Virol.* **80,** 6738–44 (2006).
103. Cocka, L. J. & Bates, P. Identification of Alternatively Translated Tetherin Isoforms with Differing Antiviral and Signaling Activities. *PLoS Pathog.* **8,** e1002931 (2012).
104. Pardieu, C. *et al.* The RING-CH Ligase K5 Antagonizes Restriction of KSHV and HIV-1 Particle Release by Mediating Ubiquitin-Dependent Endosomal Degradation of Tetherin. *PLoS Pathog.* **6,** e1000843 (2010).
105. Hong, W. & Zhao, H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. *Proc. R. Soc. B Biol. Sci.* **281,** 20141079–20141079 (2014).
106. Spirov, A. V., Sabirov, M. A. & Holloway, D. M. *In Silico* Evolution of Gene Cooption in Pattern-Forming Gene Networks. *Sci. World J.* **2012,** 1–19 (2012).
107. Phillips, C. D. & Baker, R. J. Secretory Gene Recruitments in Vampire Bat Salivary Adaptation and Potential Convergences With Sanguivorous Leeches. *Front. Ecol. Evol.* **3,** 122 (2015).
108. Spence, J. D. Increased Coagulation With Aging: Importance of Homocysteine and Vitamin B12. *Circ. J.* **81,** 268 (2017).
109. Asano, T. *et al.* Neglect-induced pseudo-thrombotic thrombocytopenic purpura due to vitamin B12 deficiency. *Pediatr. Int.* **57,** 988–990 (2015).
110. Routh, J. K. & Koenig, S. C. Severe vitamin B12 deficiency mimicking thrombotic thrombocytopenic purpura. *Blood* **124,** 1844 (2014).
111. Pennica, D. *et al.* Cloning and expression of human tissue-type plasminogen activator cDNA in E. coli. *Nature* **301,** 214–21 (1983).
112. Mackay, D. R. *et al.* Aeromonas species isolated from medicinal leeches. *Ann. Plast. Surg.* **42,** 275–9 (1999).
113. Pidiyar, V., Kaznowski, A., Narayan, N. B., Patole, M. & Shouche, Y. S. Aeromonas culicicola sp. nov., from the midgut of Culex quinquefasciatus. *Int. J. Syst. Evol. Microbiol.* **52,** 1723–8 (2002).

114. Tetlock, A., Yost, C. K., Stavrinides, J. & Manzon, R. G. Changes in the gut microbiome of the sea lamprey during metamorphosis. *Appl. Environ. Microbiol.* **78,** 7638–44 (2012).
115. Müller, H. E., Pinus, M. & Schmidt, U. [Aeromonas hydrophila as a normal intestinal bacterium of the vampire bat (Desmodus rotundus)]. *Zentralbl. Veterinarmed. B* **27,** 419–24 (1980).
116. Kikuchi, Y. & Graf, J. Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. *Appl. Environ. Microbiol.* **73,** 1984–91 (2007).
117. Worthen, P. L., Gode, C. J. & Graf, J. Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. *Appl. Environ. Microbiol.* **72,** 4775–81 (2006).
118. Maltz, M. A. *et al.* Metagenomic analysis of the medicinal leech gut microbiota. *Front. Microbiol.* **5,** 151 (2014).
119. Harjes, J. *et al.* Draft Genome Sequence of the Antitrypanosomally Active Sponge-Associated Bacterium Actinokineospora sp. Strain EG49. *Genome Announc.* **2,** (2014).
120. Datzmann, T., von Helversen, O. & Mayer, F. Evolution of nectarivory in phyllostomid bats (Phyllostomidae Gray, 1825, Chiroptera: Mammalia). *BMC Evol. Biol.* **10,** 165 (2010).
121. Gilmour, D. *The biochemistry of insects*. (Academic Press, 1961).
122. Hilditch, T. P. *The chemical constitution of natural fats*. (Chapman and Hall, 1956).
123. Wang, J.-P., Hsu, M.-F. & Teng, C.-M. Antiplatelet effect of capsaicin. *Thromb. Res.* **36,** 497–507 (1984).
124. Liang, Y. T. *et al.* Capsaicinoids lower plasma cholesterol and improve endothelial function in hamsters. *Eur. J. Nutr.* **52,** 379–88 (2013).
125. Hashiba, Y. *et al.* Capsaicin-sensitive nerves exert an inhibitory effect on the development of fibrin-induced pulmonary edema in rats. *Am. Rev. Respir. Dis.* **140,** 652–8 (1989).
126. Kaplan, M. M. Novosphingobium Aromaticivorans: A Potential Initiator of Primary Biliary Cirrhosis. *Am. J. Gastroenterol.* **99,** 2147–2149 (2004).
127. Mosquera, J. D., Zabalza, M., Laniero, M. & Blanco, J. R. Endocarditis due to Gemella haemolysans in a patient with hemochromatosis. *Clin. Microbiol. Infect.* **6,** 566–568 (2000).
128. Collins, M. D., Hutson, R. A., Falsen, E., Sjoden, B. & Facklam, R. R. Description of Gemella sanguinis sp. nov., Isolated from Human Clinical Specimens. *J. Clin. Microbiol.* **36,** 3090–3093 (1998).
129. Shi, B.-H., Arunpairojana, V., Palakawong, S. & Yokota, A. Tistrella mobilis gen nov, sp nov, a novel polyhydroxyalkanoate-producing bacterium belonging to alpha-Proteobacteria. *J. Gen. Appl. Microbiol.* **48,** 335–43 (2002).
130. Gunnell, G. F. & Simmons, N. B. *Evolutionary history of bats: fossils, molecules and morphology*. (Cambridge University Press, 2012).
131. Straney, D. O., Smith, M. H., Greenbaum, I. F. & Baker, R. J. *Biochemical genetics. In Biology of Bats of the New World Family Phyllostomidae, Part III*. (Special Publication of the Museum of Texas Tech University, 1979).
132. Slaughter, B. H. *Evolutionary trends of chiropteran dentitions. In About Bats*. (Southern Methodsist University Press, 1970).
133. Schutt, B. *Dark Banquet: Blood and the Curious Lives of Blood-Feeding Creatures (Google eBook)*. (Random House LLC, 2008).
134. Ferrarezzi, H. & A., G. E. do. Systematic patterns and the evolution of feeding habits in Chiroptera (Archonta: Mammalia). *J. Comp. Biol.* **1,** 75–94 (1996).
135. Fenton, M. B. Wounds and the origin of blood-feeding in bats. *Biol. J. Linn. Soc.* **47,** 161–171 (1992).

136. Schutt, J. W. A. in *Bat biology and conservation* (eds. Kunz, T. H. & Racey, P. A.) 157–168 (Smithsonian Institution Press, 1998).
137. Baker, R. J., Bininda-Emonds, O. R. P., Mantilla-Meluk, H., Porter, C. A. & Van Den Bussche, R. A. in *Evolutionary History of Bats: Fossils, Molecules and Morphology* (eds. Gunnell, G. F. & Simmons, N. B.) 385–409 (Cambridge University Press, 2011). doi:10.1017/CBO9781139045599.012
138. Phillips, C. D. *et al.* Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol. Ecol.* **21,** 2617–27 (2012).
139. Bordenstein, S. R. & Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol.* **13,** e1002226 (2015).
140. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science.* **352,** 560–564 (2016).
141. Zhernakova, A. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science.* **352,** 565–569 (2016).