

Published in final edited form as:

Nat Ecol Evol. 2018 April ; 2(4): 659–668. doi:10.1038/s41559-018-0476-8.

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

M. Lisandra Zepeda Mendoza^{#a,*}, Zijun Xiong^{#b,c}, Marina Escalera-Zamudio^d, Anne Kathrine Runge^a, Julien Thézé^e, Daniel Streicker^f, Hannah K. Frank^g, Elizabeth Loza-Rubio^h, Shengmao Liu^c, Oliver A. Ryderⁱ, Jose Alfredo Samaniego Castruita^a, Aris Katzourakis^e, George Pacheco^a, Blanca Taboada^j, Ulrike Löber^d, Oliver G. Pybus^e, Yang Li^c, Edith Rojas-Anaya^h, Kristine Bohmann^a, Aldo Carmona Baez^{a,k}, Carlos F. Arias^j, Shiping Liu^c, 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,*}, and M. Thomas P. Gilbert^{a,o,r,*}

^aCentre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark

^bState Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China

^cChina National GeneBank, BGI-Shenzhen, Shenzhen, 518083, China

^dDepartment of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany

^eDepartment of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom

^fInstitute of Biodiversity, Animal Health and Comparative Medicine & MRC-University of Glasgow Centre for Virus Research, University of Glasgow, G12 8QQ, Glasgow, United Kingdom

^gDepartment of Biology, Stanford University, 371 Serra Mall, 94305-5020 Stanford CA, USA

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence should be addressed to M.L.Z.M. (lisandracy@gmail.com), T.S.-P. (thomas@cbs.dtu.dk) or M.T.P.G. (tgilbert@snm.ku.dk).

Author Contributions

M.L.Z.M. and Z.X. contributed equally to the work. M.L.Z.M. and M.T.P.G. devised the study and supervised all parts of the project. M.L.Z.M., Z.X., M.E.Z., S.L. and G.Z. designed the genomic and comparative genomic methods. M.L.Z.M. and T.S.-P. designed the metagenomics methods. M.L.Z.M. led the data integration, bioinformatic analyses and did the primary interpretation of analytical outcomes in close collaboration with M.E.Z., A.D.G., D.S. and M.T.P.G. The design of the viral analyses was done by J.T., A.K., O.G.P., M.E.Z., U.L., and B.T., as well as the performance of the viral bioinformatic analyses. Z.X., Y.L., S.L., and M.L.Z.M. performed the genomic methods. A.C.B., and J.A.S.C. provided support on a variety of bioinformatic analyses. S.L. performed the laboratory protocols for the genomic analysis. G.P. performed validation of the chosen putatively lost genes. O.A.R., M.B., T.S.-P., M.T.P.G., and G.Z. aided in the generation of the common vampire bat genome. A.K.R. and M.E.Z. performed the laboratory metagenomic protocols. H.K.F., E.L.R., E.R.A., K.B., M.F.B., and N.W. carried out the collection and export of samples for the metagenomic datasets. C.F.A., E.R.A., and E.L.R. carried out the collection of the samples used for the viral analyses. M.L.Z.M. drafted the manuscript. M.E.Z., J.T., D.S., H.K.F., K.B., A.D.G., M.F.B., M.B., G.Z., T.S.-P. and M.T.P.G. provided critical and substantial contribution to the various draft versions of the paper. All authors approved the final version.

Competing financial interests

The authors declare no competing financial interests.

^hCentro Nacional de Investigación Disciplinaria en Microbiología Animal-INIFAP, Km 15.5, Blvd. Reforma, Zedec Sta. Fé, 01219 Ciudad de México, D.F., Mexico

ⁱSan Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road. Escondido, 92027 CA, USA

^jDepartamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Chamilpa 62210, Cuernavaca Morelos, Mexico

^kUndergraduate Program for Genomic Sciences, Center for Genomic Sciences, National Autonomous University of Mexico, Av. Universidad s/n Col. Chamilpa 62210, Cuernavaca, Morelos, Mexico

^lDepartment of Veterinary Medicine, Freie Universität Berlin, 14163 Berlin, Germany

^mCenter for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, 2000 Frederiksberg, Denmark

ⁿAustralian Wildlife Forensic Services, Department of Environment and Agriculture, Curtin University, 6102 Perth, Australia

^oTrace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, 6102 Perth, Australia

^pCentre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, 2100 Copenhagen, Denmark

^qCenter for Biological Sequence Analysis, Department of Bio and Health Informatics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

^rNorwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway

These authors contributed equally to this work.

Abstract

Adaptation to specialized diets often requires modifications at both genomic and microbiome levels. We applied a hologenomic approach to the common vampire bat (*Desmodus rotundus*), one of the only three obligate blood-feeding (sanguivorous) mammals, to study the evolution of its complex dietary adaptation. Specifically, we assembled its high-quality reference genome (scaffold N50=26.9 Mb, contig N50=36.6 Kb) and gut metagenome, and compared them against those of insectivorous, frugivorous, and carnivorous bats. Our analyses showed *i*) a particular common vampire bat genomic landscape regarding integrated viral elements, *ii*) a dietary and phylogenetic influence on gut microbiome taxonomic and functional profiles, and *iii*) that both genetic elements harbor key traits related to the nutritional (e.g. vitamins and lipids shortage) and non-nutritional challenges (e.g. nitrogen waste and osmotic homeostasis) of sanguivory. These findings highlight the value of a holistic study of both host and microbiota when attempting to decipher adaptations underlying radical dietary lifestyles.

The order Chiroptera (bats) exhibits a wide variety of dietary specializations, and includes the only three obligate blood-feeding mammalian species, the vampire bats (family

Phyllostomidae, subfamily *Desmodontinae*). Blood is a challenging dietary source since it consists of a ~78% liquid phase and a dry matter phase consisting of ~93% proteins and only ~1% carbohydrates¹, providing very low levels of vitamins², and potentially containing blood-borne pathogens. Vampire bats have evolved numerous key physiological adaptations to this lifestyle, for which the associated genomic changes have not yet been fully characterized due to the lack of an available reference genome. These adaptations include morphological specializations (such as claw-thumbed wings and craniofacial changes including sharp incisors and canines), infrared sensing capacity³ for the identification of easily accessible blood vessels in prey⁴, and renal adaptations to the high protein content in its diet⁵ (such as a high glomerular filtration rate and effective urea excretion). Furthermore, given the high risk of exposure to blood-borne pathogens, another important trait in the vampire bat is its immune system⁶.

Besides genomic adaptations, the role of host-associated microbiota may play an additional, possibly equally important, role in the evolution of vertebrate dietary specialization⁷. Although the functional role of the vampire bat gut microbiome has not been studied, analyses of obligate invertebrate sanguivorous organisms⁸ have shown that the gut microbiota contributes to blood meal digestion⁹, provision of nutrients absent from blood¹⁰, and to immunological protection¹¹. Studies on mammals have shown that the gut microbiome is a key aspect of an organism's digestive capacities (energy harvest, nutrient acquisition, and intestinal homeostasis)¹², and that it also affects phenotypes related to the immune and neuroendocrine systems^{13,14}. Furthermore, changes in the gut microbiome are associated with diseases such as diabetes, obesity, irritable bowel syndrome and inflammatory bowel disease^{15–17}. In response to the growing awareness of the key roles that host-microbiome relationships can play across the spectra of life, various studies have advocated for the “hologenome” concept^{18–20}. In brief, this argues that natural selection acts on both the host and its microbiome (together forming the holobiont), thus evolutionary studies should incorporate both. The extreme dietary adaptation of vampire bats provides a suitable model to investigate the effect of selection across genome and microbiome, and thus explores the question of the role of associated microbiome in the evolution of specialized diets.

In this study, we explore the contributions of both the common vampire bat's nuclear genome and gut microbiome to its adaptation to obligate sanguivory. To this end, we generated both the common vampire bat genome and fecal metagenomic datasets as a proxy to study its gut microbiome, as well as fecal metagenomic datasets from other non-sanguivorous bat species. We used these datasets for comparative genomic and metagenomic analyses. Specifically, we analyzed the common vampire bat genomic landscape, the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS), putative gene loss and gene family expansion/contraction, and computational predictions of the functional impact of amino acid substitutions. We also performed microbial taxonomic and functional profiling, identified the microbial taxonomic and functional core of the common vampire bat, and enriched microbial taxa and functions. Following a hologenomic approach, we identified elements in both the host genome and microbiome that could have played relevant roles in adaptation to sanguivory.

Results and Discussion

Genomic landscape

We sequenced and *de novo* assembled the ~2 Gb common vampire bat genome using Illumina sequencing technology (Supplementary Information 1). The genome is smaller than that of other mammals, but similar to previously reported bat genomes²¹. The initial assembly (~100X mean coverage, scaffold N50=5.5 Mb and N90=933 Kb, Supplementary Figs. 1, 2) was subsequently improved using the *in vitro* proximity ligation-based technology for assembly contiguity refinement developed by Dovetail Genomics²². We obtained a final high-quality assembly with scaffold N50=26.9 Mb and N90=9.46 Mb, contig N50=36.6 Kb and N90= 8.8 Kb (Supplementary Table 1, Supplementary Fig. 3, Supplementary Information 1). We used our annotated common vampire bat genome (see Methods – Protein-coding gene and functional annotation) for comparative genomic analyses with publicly available bat genomes and other mammalian genomes (Supplementary Table 2, Fig. 1).

Repetitive elements can significantly contribute to genome evolution. Thus, for a genomic landscape characterization, we first compared transposable elements (TEs) in the common vampire bat genome to those within the genomes of non-phylostomid bats with other diets: the carnivorous *Megaderma lyra* (greater false vampire bat, Megadermatidae), the insectivorous *Pteronotus parnellii* (Parnell's mustached bat, Mormoopidae), and the frugivorous *Pteropus vampyrus* (large flying fox, Pteropodidae) (Fig. 1, Supplementary Table 3, Supplementary Note 1). We identified a 1.6 to 2.26-fold higher copy number of the MULE-MuDR transposon in the common vampire bat genome relative to the other bat genomes. The high mutagenic capacity of MULE-MuDR has been demonstrated to have played a critical role in the evolution of some plants²³. Furthermore, TEs in general may cause structural or functional changes within the genome and alter epigenetic regulation of the genes into which they are inserted²⁴. We therefore explored whether these elements might have also played a role in the evolution of sanguivory by analyzing their genomic location in the nuclear genome of the common vampire bat. We found that the identified common vampire bat TEs, MULE-MuDR elements in particular, were located within genomic regions enriched for gene ontology (GO) functions related to the challenges of sanguivory, such as antigen processing and presentation, defense response to viruses, lipid metabolism, and vitamin metabolism (Supplementary Information 2).

A sanguivorous diet facilitates exposure to blood-borne viruses that could lead to an increase in genomic invasion by retroviral and non-retroviral endogenous viral elements (EVEs). Thus, we next characterized their presence in the common vampire bat genome. Compared to previously published EVE studies on non-Chiropteran mammals, the common vampire bat exhibits a greater diversity of non-retroviral EVEs in terms of the number of integrations, including endogenized viral genes from Avian Bornaviridae and Parvoviridae/Dependovirus. However, these findings are not restricted to vampire bats and are similar to those in other bat species²⁵ (Supplementary Note 2, Supplementary Fig. 4, Supplementary Information 3). Surprisingly, and in contrast to the prior expectations given its sanguivorous diet, the diversity of endogenous retroviral elements (ERVs) in the common vampire bat is very low

compared to other bat species²⁶. The only proviral elements detected were DrERV27 and DrgERV, both present in low-copy numbers (Supplementary Note 3, Supplementary Fig. 5, Supplementary Information 4). We hypothesize that genome colonization by ERVs could have been restricted by the genomic adaptations in the common vampire bat genome against ERVs insertion and proliferation. In support of our hypothesis, we identified expansion of the anti-retroviral gene *TRIM5* family (Viterbi $P=0.00088$, Supplementary Information 5, Supplementary Note 4).

Genomic adaptations to sanguivory

Feeding specialization often requires morphological and physiological adaptations in traits such as the sensory apparatus (e.g. infrared sensing), locomotion, digestion, kidney function, and immunity (Figure 2A, Supplementary Information 6). For example, it has been shown that vampire bats have a loss of sweet taste genes and reduction of bitter taste genes²⁸. In agreement, such genes were also identified in our putative gene loss analysis (Supplementary Information 7, Supplementary Note 5). It is likely that the function of those genes is related to sanguivory, since sweet and bitter taste receptor genes influence glucose homeostasis in human²⁹. Interestingly, we found that the common vampire bat bitter taste receptor *TAS2R3* has experienced episodic positive selection and shows two species-specific positively selected sites (PSSs) on topological domains, one of them having a potential impact on protein function (PROVEAN score=-4.4) (Supplementary Table 4, Supplementary Fig. 6). Among the enriched GOs of the differentially evolving genes we identified functions related to regulation of RNA splicing, which could be relevant to sanguivory given that *D. rotundus* produces submandibular tissue specific splicing isoforms to counteract the prey's response to injury³⁰ (Supplementary Information 8, Supplementary Note 6). Regarding the recruitment of alternatively spliced forms, the ganglion-specific splicing of *TRPV1* has been found to underlie the vampire bat's infrared sensing ability³¹. Interestingly, we found that *PRKDI*, which directly modulates the product of *TRPV1*³², is positively selected and exhibits species-specific PSSs (branch-site test $P=3.39E-10$, branch test $P=1.39E-6$, Supplementary Information 8). These examples suggest an important role of alternative splicing as a form of regulatory evolution fundamental to sanguivory (Supplementary Note 7). However, it is clear that despite the number of detected genomic features related to sanguivory adaptation, they alone cannot address all the challenges posed by this diet (Fig. 2B).

Gut microbiome diet and phylogenetic influence

We generated Illumina shotgun metagenome data in order to compare the gut microbiomes of 13 fecal common vampire bat samples with those of non-sanguivorous non-phylostomid bats. Specifically, eight frugivorous *Rousettus aegyptiacus* (Egyptian fruit bat, Pteropodidae), five insectivorous *Rhinolophus ferrumequinum* (greater horseshoe bat, Rhinolophidae), and five carnivorous *Macroderma gigas* (ghost bat, Megadermatidae) bats (Supplementary Information 9). We obtained a median of 15.8 Gb of sequencing data (~37.6 million 100 bp paired-end reads) for each dietary category. After filtering low quality bases, adaptor sequences, and bat genome derived reads, we obtained a median of 2.77 Gb of high-quality data for each species, totaling 86.73 Gb data. We identified taxa and functions present only in the common vampire bat microbiome (gut microbiome core), as well as taxa

and functions that exhibit statistically significant differences in abundance or contribution to variation between the different microbiomes (Supplementary Information 8, 6, Supplementary Tables 5, 6, Supplementary Note 8).

It has been observed previously that similarity in the taxonomic composition of vertebrate gut microbiomes (including bats) can be influenced by the diet and the phylogenetic relationships of the respective host species³³. Overall, the common vampire bat microbiome taxonomic composition is more similar to that of the insectivorous and carnivorous bats, than to that of the frugivorous bat. This may reflect a phylogenetic influence on the microbiome taxonomic profile (Fig. 3A, Supplementary Fig. 7, and Supplementary Notes 9, 10). In contrast, vampire bat microbiome is strikingly different to that of the compared bats at the functional level, which was characterized by the KEGG annotations of the microbial non-redundant gene catalogues assembled from the metagenomic datasets. While there is little differentiation between the functional gut microbiomes of carnivorous, insectivores and frugivorous bats, the common vampire bat functional gut microbiome is almost completely distinct, and exhibits the least intra species variation between the samples (Fig. 3B, Supplementary Fig. 8, Supplementary Table 7). This suggests that the functional profile is less influenced by phylogeny than the taxonomic profile, and that the common vampire bat gut microbiome harbors a set of functions specialized to its extreme diet (Supplementary Note 11). Subsequently, we analyzed the comparative genomic and metagenomic results in a hologenomic framework to demonstrate how both components contribute to adaptation to sanguivory (Fig. 4, Supplementary Figs. 9-12, Supplementary Tables 7-9, Supplementary Information 8, 10).

The hologenomic framework of sanguivory

Viscosity and subsequent coagulation represent a challenge for ingestion and digestion of blood. Besides developing potent anticoagulants in its saliva³⁴, the common vampire bat hologenome addresses this challenge in various ways. For instance, *REG4*, involved in metaplastic responses of the gastrointestinal epithelium, was found to be under ongoing positive selection (M8a/M8 test $P=0.047$) with possible functional impact in its carbohydrate binding capacity, including binding of the anticoagulant heparin (Supplementary Table 4, Supplementary Information 8). Furthermore, we identified genes in the common vampire bat microbial functional core from pathways for degradation of heparan sulfate and dermatan sulfate, both being polysaccharides involved in blood coagulation (Supplementary Information 10). We also identified an enrichment in the common vampire bat microbial gene L-asparaginase (Fisher's $P=0.00027$), which decreases protein synthesis of coagulation factors³⁵ (Supplementary Information 11).

Besides specialized digestion, sanguivory poses other challenges related to the poor nutritional value of blood itself, as well as to the side effects that may arise due to the blood components as sole dietary source (Fig. 2B). We identified elements in both the genome and gut microbiome that might be involved in solving each of these challenges discussed next.

Hologenomic solutions to nutritional challenges

Low nutrient availability: Obligate sanguivory requires adaptation to very low levels of some nutrients, such as essential amino acids and vitamin B complex^{36,37}, and very high levels of others, such as salt³⁸. Our data clearly demonstrate how both the host and its associated gut microbiome have dealt with these challenges. In the genome, we found the gene *LAMTOR5* to be positively selected in the common vampire bat genome (FDR=2.02E-7, Supplementary Information 8). This gene is involved in the response to nutrient starvation³⁹, suggesting that the common vampire bat metabolism has adapted to the low nutrient content available in blood. Similarly, we identified KOs in the common vampire bat microbial core related to energy and carbohydrate metabolism (Fig. 4B, Supplementary Information 11). For example, when compared to the other bats, we identified an enrichment in the common vampire bat microbial genes involved in response to low nutrient availability (RelA/SpoT family protein Fisher's $P=0.0064$, and guanosine pentaphosphate Fisher's $P=0.0018$), and enzymes in the common vampire bat core involved in the reverse Krebs cycle (Supplementary Information 10), which is used by some bacteria to produce carbon compounds from CO₂ and water (abundant blood components). We speculate that the presence of such metabolic pathways indicates the growth of specific microbes in the gut's environmental conditions resulting from blood consumption.

The holobiont has also provided solutions to the lack of important nutrients in blood (Fig. 4C, Supplementary Information 8). For example, the *PDZD11* gene in the common vampire bat genome, involved in vitamin B₅ metabolism, evolved faster in the common vampire bat genome relative to the other examined bats (branch test $P=1.97E-10$). We further postulate that the microbiome also contributes in tackling the low nutritional challenge by providing necessary nutrients. For example, compared to the other bat microbiomes, the common vampire bat gut microbiome had the highest number of enriched enzymes related to the biosynthesis of cofactors and vitamins, such as carotenoid (Supplementary Fig. 8B, Supplementary Information 10, 11). Furthermore, we identified enzymes involved in the metabolism of butyrate, an important nutrient for cells lining the mammalian colon derived from bacterial fermentation⁴⁰, enriched in the common vampire bat gut microbiome as well as in the vampire bat gut microbiome core (Supplementary Information 10, 11).

Lipid and glucose assimilation: Besides vitamins and other nutrients, lipids and glucose may not be readily available in blood. Furthermore, vampire bats have a reduced capacity to store energy reserves. In agreement with this, we identified GO enrichment of lipid metabolism on genes with dN/dS values statistically higher or lower compared to other bats (Fig. 4AC, Supplementary Information 8). For example, we identified the gene *FFAR1*, which plays an important role in glucose homeostasis, as evolving faster in the common vampire bat compared to the other bats (branch test $P=3.68E-5$) and containing amino acid substitutions with a possible functional impact in its binding ability (Supplementary Table 4). This may enable the common vampire bat to better utilize the available glucose. The common vampire bat gut microbiome also exhibits unique solutions to the challenge (Fig. 4BD, Supplementary Information 10, 11). Differences in the carbohydrate and glycan metabolism functional profile were identified in the principal component analysis (PCA) comparing the different microbiomes (Supplementary Fig. 11), which place the vampire bat gut

microbiome profile within a cluster separate from those of the non-sanguivorous bats. Importantly, we identified enrichment in the microbial gene glycerol kinase in the common vampire bat (Fisher's $P=0.0027$), which plays a key role in formation of triacylglycerol and in fat storage, and its deficiency causes symptoms such as hypoglycemia and lethargy in a mouse model⁴¹.

Hologenomic solutions to non-nutritional challenges

Immunity: Due to its sanguivorous lifestyle, the common vampire bat risks direct contact with blood-borne pathogens from prey. Consequently, we observed >280 bacterial species known to be pathogenic to some mammalian species present exclusively in the common vampire bat gut microbiome (Supplementary Information 12). For example, we identified enrichment of genes from *Borrelia* (Supplementary Information 11) and *Bartonella* as one of the most abundant genera in the common vampire bat compared to the other bats (Supplementary Information 13). These bacteria are known to be transmitted by sanguivorous invertebrates (ticks, fleas, mosquitoes, and lice). This suggests that the abundance of this genus could be a shared pattern of sanguivorous species. While several studies have elucidated part of the expected genomic immunity-related adaptations⁶, analysis of the full genome enabled us to identify more elements related to immunity, such as defense response to virus and antigen processing and presentation (Fig. 4AC, Supplementary Information 8). For example, we identified the antimicrobial gene *RNASE7* to be positively selected (branch-site test $P=0.004$) and containing amino acid substitutions that may increase its bactericidal capacity (Supplementary Table 4). In addition, when compared to the gut microbiomes of non-sanguivorous bats, that of the common vampire bat contains a large abundance of potentially protective bacteria, such as *Amycolatopsis mediterranei* ($P<0.05$), which has been shown to produce antiviral compounds against bacteriophages and poxviruses⁴² (Supplementary Information 13).

Iron assimilation: Iron concentration represents a significant challenge to sanguivory. Although the concentration of free iron ion is not high in the blood, severe hemolysis (e.g. during digestion of blood) could result in high levels of iron that, if absorbed in excess, could accumulate and disrupt normal function in organs such as liver, heart and pancreas. Interestingly, we identified the light and heavy chains of the iron storing protein ferritin (*FTL*, *FTH1*) under gene family expansion in the common vampire bat genome (Viterbi $P=0$ and Viterbi $P=0.0012$, respectively, Supplementary Information 5). Additionally, we identified an enrichment of the iron storing protein ferritin (Fisher's $P=0.0014$), suggesting that the gut microbiome also contributes to solving this challenge (Fig. 4D, Supplementary Information 11).

Nitrogen waste and blood/osmotic pressure: The high level of protein abundance in the blood and its rapid ingestion could lead to accumulation of nitrogenous waste products, primarily urea, which could lead to renal diseases-like symptoms (e.g. high blood-pressure and fluid retention). This challenge is exacerbated by the abundance of salts in blood, which pose additional osmotic and blood pressure challenges. We see at the genome level that this is addressed by a higher rate of evolution in the common vampire bat genes compared to the other bats involved in disposal of excess nitrogen (Fig. 4AC, Supplementary Information 8),

such as *PSMA3* (branch test $P=2.08E-07$). This challenge seems to also be addressed by the gut microbiome. The PCA of the copy number of genes involved in amino acid metabolism distinguishes the common vampire bat in a single cluster separated from the other bat species analyzed (Supplementary Fig. 11), suggesting a specialized microbial amino acid metabolism capacity. We also identified enrichment in the common vampire bat microbial gene urease subunit alpha (*ureA*, Fisher's $P=0.016$) involved in urea degradation (Supplementary Information 11).

Conclusions

It is clear from our results that the common vampire bat has adapted to sanguivory through a close relationship between its genome and gut microbiome. We identified a phylogenetic and dietary impact on the common vampire bat gut microbiome and uncovered an unexpected genomic viral and repetitive element genomic makeup. We showed that extreme dietary specializations, such as that of the common vampire bat, provide a comparative framework with which to tease apart the relative role of genomes and microbiomes in adaptation. In conclusion, our study illustrates the benefits of studying the evolution of complex adaptations under a holobiome framework, and suggests that vertebrate adaptation studies that do not account for the action of the hologenome may fail to recover the full complexity of adaptation.

Methods

Genome sequencing and raw reads processing

We shotgun sequenced the *D. rotundus* genome using a wing biopsy from a sample collected by the NIH through the Catocin Wildlife and Zoo in Thurmont, Maryland, USA. Capturing method and dead preservation procedure of the specimen is unknown. The age and sex of the dead individual is unknown. Sampling permits were given to BGI for the sequencing of the specimen, originally as part of the BGI 10K genome project. Genomic DNA was extracted at the Laboratory of Genomic Diversity and was fragmented to 2-10 kb. Sequencing libraries were constructed with insert sizes 170 bp-10 kb, according to the Illumina protocol for sequencing on Illumina HiSeq2000 following the manufacturer's instructions. We sequenced reads of 49 bp for the long insert size libraries (2 kb, 5 kb and 10 kb), and 100 bp for the short insert size libraries (170 bp, 500 bp and 800 bp). Sequencing errors were corrected based on k -mer frequency and low-quality reads filtered out using SOAPfilter43 as follows: 1) Remove reads with $>10\%$ N's. 2) For short insert-size libraries (<2 kb), reads were removed if the quality score of $>60\%$ bases was <7 . For large insert-size libraries (>2 kb), reads were removed if the quality score of $>80\%$ bases was <7 . 3) Adapter sequences, and duplicate or identical reads were removed. 4) Read pairs were removed if Read1 and Read2 were completely identical. 5) For short insert size paired-end sequences, reads with overlapping length >10 bp between the Read1 and Read2 were removed.

Genome assembly

We estimated the genome size of *D. rotundus* using Kmerfreq44 by dividing the total number of 17-mers by the peak of the 17-mer Poisson distribution. High quality reads were

assembled using SOAPdenovo43 as follows. 1) Short insert library reads were assembled as initial contigs ignoring the sequence pair information. 2) Reads were aligned to the previously generated contig sequences. Scaffolds were constructed from short insert-size libraries to large insert-size libraries by weighting the paired-end relationships between pairs of contigs, with at least 3 read pairs required forming a connection between any two contigs. 3) Gaps in the scaffolds were closed using GapCloser43. Genome quality assessment was performed by downloading a publicly available *D. rotundus* transcriptome45 and aligning the transcripts to the genome using BLAT46.

Genome contiguity improvement

We prepared two Chicago libraries22 using 5µg of high molecular weight DNA obtained from *D. rotundus* 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. Permits for this were obtained from the San Diego Zoo Global. Capturing method and dead preservation procedure of the specimen is unknown. The age and sex of the dead individual is unknown. DNA was extracted with Qiagen Blood and Cell Midi kits according to the manufacturer's instructions. The steps required for building the Chicago libraries were performed as described in 22. The libraries were sequenced using Illumina HiSeq 2500 2X100 bp rapid run. Our initial *D. rotundus* assembly, shotgun sequence data, and Chicago libraries sequences were used by Dovetail Genomics as input data for HiRise, as described in 22. Genome assembly contiguity statistics were obtained using a minimum N track length of 1 to delimit the contig blocks within the scaffolds.

Protein-coding gene and functional annotation

Homology based gene prediction was performed using as reference 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 the *D. rotundus* assembly using tblastn47 and linked the blast hits into candidate gene loci with genBlastA48. We filtered out those candidate loci with homologous block length <90% of the query length. We extracted genomic sequences of candidate gene loci, including the intronic regions and 3 kb upstream/downstream sequences. The sequences were passed to GeneWise49 to search for accurately spliced alignments. We filtered out pseudogenes containing more than one frame error for single-exon genes. Potentially pseudogenized single exons were removed if they were part of a multi-exon gene. We then aligned protein sequences of these genes against Uniprot using blastp and filtered out genes without matches. We also filtered out genes that had >80% repeat regions. *De novo* gene prediction was performed with AUGUSTUS50 using a published common vampire bat transcriptome45 as training dataset and with masked TE-related repeats. We filtered out partial and <150 bp predicted genes. Genes that aligned over 50% of their length to annotated TEs were filtered out. Finally, we built a non-redundant (nr) gene set with the homology-based evidence prioritized over the *de novo* evidence. If *de novo* genes were chosen in the reference gene set, we only retained those with >30% of their length aligning against Uniprot51 and that contained at least 3 exons. The integrated gene set was translated into amino acid (aa) sequences, which were used to search the InterPro database with iprscan_4.852. We used BLAST to search the metabolic pathway database in KEGG53 and homologs in the SwissProt and TrEMBL databases in UniProt. The quality and

annotation of the *D. rotundus* genome was assessed and compared to that of *E. europaeus*, *R. ferrumequinum*, *M. brandtii*, *M. davidii*, *M. lucifugus*, and *P. parnellii* with BUSCO54.

Repeat annotation

Repeat annotation was performed on the genomes of *D. rotundus*, *P. parnellii*, *M. lyra*, and *P. vampyrus*. Transposable elements (TEs) were identified using RepeatMasker55 and RepeatProteinMask against Repbase TE library56. We used RepeatScout, PILER-DF and RepeatModeler-1.0.555,57 to construct a *de novo* TE library, which was then used by RepeatMasker to predict repeats. We predicted tandem repeats using TRF58. LTR_Finder59 was used to detect long terminal repeats (LTRs). The Repbase-based annotations and the *de novo* annotations were merged.

Non-retroviral EVEs

We constructed a comprehensive library of all non-retroviral virus protein sequences available in GenBank and EMBL. We used DIAMOND60 to search these sequences against the *D. rotundus* genome. We extracted the matching aa sequences and performed reciprocal blastp-like searches61 using DIAMOND with the selected subset of *D. rotundus* aa sequences and the set of nr protein sequences. *D. rotundus* genome sequences were considered of viral origin if they unambiguously matched viral proteins in the reciprocal best hits. Putative viral ORFs were inferred from automated alignments, using exonerate62. For each putative *D. rotundus* viral peptide, we retrieved the function and predicted the taxonomic assignment by comparison to the best reciprocal blastp-like hit viral proteins. Phylogenetic analyses were performed by aligning the sequences using MAFFT63 and curated using AliView64. ML inferences were performed on each multiple aa alignment using RAxML65. Support for nodes was obtained from 100 non-parametric bootstrap iterations, and the root of ML trees was determined by midpoint rooting. We confirmed the presence of selected virus by constructing double indexed Illumina libraries66 on DNA derived from the spleen tissue of four different *D. rotundus*. Libraries were pooled and sequenced using 150 bp PE using the Illumina platform NextSeq 500. The raw paired-end reads were quality assessed, merged and filtered by mapping against the bacterial, human and chiropteran reference genome database with SMALT (<https://www.sanger.ac.uk/resources/software/smalt/>). Reads >150 bp were run through a viral assignment pipeline67 using blastx68 to search against the viral nr protein database. Viral-matching reads were assigned a taxonomy with MEGAN569. Reads matching viral sequences were manually verified by reciprocal blastx analysis. Positive viral hits were defined as those with at least two different reads matching two different proteins or two different regions within the same viral protein.

ERVs

We identified flanking LTR regions using RepeatMasker70 and RMBLAST (<http://www.repeatmasker.org/RMBlast.html>), Tandem Repeats Finder58 and RepBase71 to the *D. rotundus* genome. We used blastn and tblastx with the retro-viral reference genome sequences in the RefSeq database of NCBI against the *D. rotundus* genome. Reads matching each of the RefSeq sequences were extracted and sorted by GI, collapsed by id and manually verified by reciprocal blastn. Tblastx was re-run and the kept reads were mapped to the *D.*

rotundus genome with SMALT and manually verified by reciprocal blastp. ERV sequences validation was done with tblastx against the *Retroviridae*. The retrieved ERV sequences were mapped using SMALT v0.7.6 against the genomes of *D. rotundus*, *M. natalensis*, *E. fuscus*, *M. lucifugus*, *M. brandtii*, *M. davidii*, *P. alecto*, *R. aegyptiacus*, and *P. parnellii*. Tblastx of the full genomic sequences of DrERV and DrgERV was also performed against Chiroptera. Selected genomic contigs were further analyzed by reciprocal blastx against the non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF database restricting the search to the *Retroviridae*, or against the DrgERV sequences. Retrieved sequences were aligned using MUSCLE in SeaView72. The best-fit aa substitution model for each alignment was identified using jModelTest273 and trees were inferred under ML criteria using RAxML65.

Orthologous gene families

Using the *D. rotundus* genome against a range of other mammalian species, we performed clustering of orthologous genes using two strategies. 1) Identifying single-copy orthologs in the species by using the TreeFam method74. 2) Identifying 1:1 orthologs by building pairwise orthologs between *D. rotundus* and the other species and uses a RBH plus synteny approach as described in 75.

dN/dS analyses

We used PAML codeml76 on the cleaned CDS alignments from the two sets of orthologous families and the corresponding phylogenetic tree as reported in 77. In order to identify genes with accelerated evolution in the *D. rotundus* lineage we ran the two-ratio branch model. As null model, we used the one-ratio model. Using these results, we performed likelihood ratio tests (LRT) to identify genes with significant p value (*P*). To adjust for multiple testing, we used the False Discovery Rate (FDR) method. To identify genes with positively selected sites in *D. rotundus* we also used a branch-site with PAML codeml76. LRT and FDR were computed as done for the branch model tests.

Gene family expansion/contraction

We used CAFE78 with the results from the single-copy orthologous gene families. In order to obtain the dated tree required as input for CAFE, we obtained divergence times from the fossil dating records from TimeTree79. We concatenated the CDSs aligned with MAFFT and used PAML mcmctree to determine split times with the approximate likelihood calculation method. Genes with >200 copies in one of the species were filtered out.

Putative gene loss

We identified genes putatively lost in *D. rotundus* as previously described in 80. The human gene set was downloaded from Ensembl and obtained an nr gene catalogue to be used as reference by collapsing redundant genes and keeping the longest ORF. We used *Bos taurus*, *Equus caballus*, and *E. europaeus* as outgroups. Genes identified as missing in *D. rotundus* were searched with blastp against the protein catalogues of the other available bats. Further validation on selected genes was performed by a) evaluating the conservation of their syntenic regions compared to other species, and b) the GC content of the syntenic regions

compared to the *D. rotundus* genome average GC content; c) searching them against the published *D. rotundus* transcriptome; and d) through PCR assays.

Functional characterization

We performed GO analysis using GOrilla81 as well as manual characterization using the GO annotations of the human genes downloaded from Ensembl and literature mining. We used PROVEAN82 to characterize the functional impact of the non-synonymous substitutions present only on the proteins in *D. rotundus*.

Species specific PSS identification and protein modeling

We tested for positive selection and positive selected sites (PPS) using the complete CDS alignments for the proteins FFAR1, PLXNA4, REG4, RNAS7 and TA2R3 from various species downloaded from the OrthoMaM database83. Sequences were re-aligned using MUSCLE84 in SeaView72 and phylogenetic analysis was performed using PhyML85. We used PAML codeml76 to test: a) M1/M2 branch model constraining the *D. rotundus* node76, b) M8a/M8 site model, and c) branch-site model A (BSA) constraining the *D. rotundus* node. The BSA model was evaluated under a LRT against the null hypothesis, while PSS were scored under NEB and BEB. We performed modeling of the mentioned proteins from *D. rotundus* using PyMOL Molecular Graphics System (Schrödinger, LLC). The 3D protein models of the *D. rotundus* sequences were constructed using Phyre286 based on profile hidden Markov model analysis using the structures reported in 87–90.

Metagenomic data

We used fecal and anal swab samples from *D. rotundus*, *R. aegyptiacus*, *R. ferrumequinum*, and *M. gigas*. The *D. rotundus* samples were obtained from wild *D. rotundus* captured using mist nets and cloth bags and preserved in RNAlater according to the manufacturer's instructions. The *R. ferrumequinum* samples were collected in Woodchester Mansion, Gloucestershire, UK. The *M. gigas* samples from Australia were collected by the consulting company Biologic in Pilbarra, Australia. The *R. aegyptiacus* fecal samples were collected from the Copenhagen Zoo. We have complied with all relevant ethical regulations for the collection of these samples. DNA from fecal samples was extracted using the PowerFecal DNA Isolation Kit (MoBio) with modifications to the manufacturer's instructions. DNA from the *D. rotundus* milk sample was extracted using a standard commercial kit (Qiagen), following manufacturer's instructions. Anal swab samples were extracted using the BiOstic® Bacteremia DNA Isolation Kit (MoBio) following the manufacturer's protocol. All samples were kept at -20°C. Libraries were built using the NEBnext DNA Library Prep Mast Mix Set for 454 (New England BioLabs) following the manufacturer's instructions. Samples were 100 bp paired-end sequenced on the Illumina 2500 HiSeq platform.

Taxonomic and functional identification

The reads were cleaned with Trimmomatic91 and prinseq-lite92. The datasets were mapped against the closest available bat genome and only the non-mapping reads were kept. MGMapper93 was used for the taxonomic identification of invertebrates, protozoa, fungi, virus, and bacteria. We kept the species identified with more than the 1st quartile coverage

from the coverage distribution of the corresponding database and filtered out those found on the corresponding extraction blanks. Rarefaction curves from each dataset were obtained using an in-house script with the MGmapper results. We then performed *de novo* assembly using IDBA_UD94, predicted genes using Prodigal95 and generated a nr gene catalogue with usearch96. The nr catalogue was searched with ublast96 against the Uniprot database51 for functional and taxonomic annotation. We used DIAMOND60 to search the unmapped reads against the Uniprot database keeping only the best hit for functional and taxonomic annotation. We annotated the Uniprot protein identifications using KEGG orthology (KO) and eggNOG ids.

Taxonomic and functional metagenomics comparison

We filtered based on the breadth of coverage and the identifications of the extraction blanks. We removed any non-microbial hit and any taxa in which the paired reads matched different genera or only one of the reads had a hit. The counts were normalized by percentage. We identified a microbial taxonomic and functional sanguivorous core by comparing the filtered sets of the bats and keeping as core those taxa and genes identified only in the *D. rotundus* samples. We manually examined the taxa from the filtered taxonomic identifications, and the KO and COG annotations from the filtered nr gene set catalogue. We compared the normalized abundance of taxa and functions between *D. rotundus* and the non-sanguivorous bats as follows: 1) Using the distributions of the different functional categories from *D. rotundus* and each non-sanguivorous bat species with a Wilcoxon rank-sum test. 2) Using the entire taxonomic and functional datasets, as well as down-sampling the normalized count values. Sampling values were the minimum, median and 3rd quartile values of the counts distributions. With the resulting datasets we calculated the Euclidean, Bray-Curtis, and Jaccard distance metrics with the R package vegan97, and used the Ward hierarchical clustering method using UPGMA and Ward, and performed PCAs with prcomp and the GPA method. 3) We identified taxa and functions significantly contributing to the variation between the *D. rotundus* and the non-sanguivorous bat species. We examined the rotation matrix from the PCA of the normalized counts, excluding the four deepest sequenced samples, of the species and genera microbial taxonomic levels and the KEGG functional pathways. We identified the most significantly abundant *D. rotundus* microbial taxa as those with a significantly higher median normalized counts values ($P < 0.05$) in *D. rotundus* and a median and mean normalized count values of 0 in the other 3 bat species for the first three PCs. We also identified significantly more abundant genes in *D. rotundus* by generating and annotating with KEGG an nr gene set with all the predicted genes from all the bat samples. The reads of the samples were mapped against this bat nr gene set, and a normalized count matrix was generated and used for Fisher tests on each of the functional pathways.

Data availability

The NCBI BioProject accession code for the genome assembly is PRJNA414273, the sequence reads are available at the NCBI sequence read archive (SRA) under the accession SRA619672. The BioProject code for the metagenomic sequencing data is PRJNA415003, and the reads can be accessed at SRA with the accession SRA620977.

Code availability

In-house scripts used for the processing of the data are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank María Luisa Méndez Ojeda, Gabor Czirjak, Luis Caballero, Omar Ríos, Alfredo Patraca, Carlos Tello, Daniel Becker, and Melody E. Roelke-Parker for the collection and exportation of vampire bat samples, as well as Polina Perelman and Marisa Korody for the preparation of genomic DNA. We thank Dovetail Genomics for the improvement of the genome contiguity of *D. rotundus*, in particular to R. Ed Green, Margot Hartley, and Brandon Rice. We furthermore thank Gareth Jones for his support in the capture of the *R. ferrumequinum* bats samples and for useful comments on the manuscript, and Rob Dunn and Lauren Nichols for their support at the early stages of the project. Likewise, we thank Antton Alberdi and Ostaizka Aizpurua for assistance regarding samples from *Rhinolophus euryale*, *Miniopterus schreibersii*, *Nyctalus leisleri*, and *Nyctalus lasiopterus*. We thank the Danish National High-Throughput DNA Sequencing Centre for assistance in generating the metagenome and Dovetail Illumina sequence data. We also thank the Instituto de Biotecnología-UNAM for giving us access to its computer cluster and Jerome Verleyen for his computer support. M.T.P.G. and M.L.Z.M. recognize Lundbeck Foundation R52-A5062, ERC Consolidator Grant (681396 – Extinction Genomics), and Danish National Research Foundation DNRF94 grants as the principal funding behind this work. M.E.Z. and A.D.G. acknowledge the Deutsche Forschungsgemeinschaft (DFG) (GR 3924/9-1 to A.D.G) and an international doctoral scholarship provided by the Consejo Nacional de Ciencia y Tecnología (CONACyT) and the German Academic Exchange Service (DAAD) (311664). O.G.P. and J.T. acknowledge funding from the European Research Council under FP7/2007-2013/ERC grant agreement no. 614725-PATHPHYLODYN. A. K. recognizes funding from the Royal Society. H.K.F. acknowledges funding from a Bing-Mooney Fellowship in Environmental Science and Conservation. Field work in Peru was funded by the US National Science Foundation (Grant DEB-1020966) and a Sir Henry Dale Fellowship to D. S., jointly funded by the Wellcome Trust and the Royal Society (Grant 102507/Z/13/Z). Funding for Costa Rican fieldwork came from the Stanford Woods Institute for the Environment Environmental Venture Project.

References

- Breidenstein CP. Digestion and Assimilation of Bovine Blood by a Vampire Bat (*Desmodus rotundus*). *J Mammal.* 1982; 63:482–484.
- Edwards MA, Kaufman ML, Storvick CA. Microbiologic assay for the thiamine content of blood of various species of animals and man. *Am J Clin Nutr.* 5:51–5. [PubMed: 13394540]
- Gracheva EO, et al. Molecular basis of infrared detection by snakes. *Nature.* 2010; 464:1006–11. [PubMed: 20228791]
- Kishida R, Goris RC, Terashima S, Dubbeddam JL. A suspected infrared-recipient nucleus in the brainstem of the vampire bat, *Desmodus rotundus*. *Brain Res.* 1984; 322:351–5. [PubMed: 6509324]
- Singer MA. Vampire bat, shrew, and bear: comparative physiology and chronic renal failure. *Am J Physiol Regul Integr Comp Physiol.* 2002; 282:R1583–92. [PubMed: 12010738]
- Escalera-Zamudio M, et al. The evolution of bat nucleic acid sensing Toll-like receptors. *Mol Ecol.* 2015; doi: 10.1111/mec.13431
- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol.* 2008; 6:776–788. [PubMed: 18794915]
- Graf J, Kikuchi Y, Rio RVM. Leeches and their microbiota: naturally simple symbiosis models. *Trends Microbiol.* 2006; 14:365–71. [PubMed: 16843660]
- Hornborstel H. Ueber die bakteriologischen Eigenschaften des Darmsymbionten beim medizinischen Blutegel (*Hirudo officinalis*) nebst Bemerkungen zur Symbiosefrage. *Zbl Bakteriol.* 1942; 148:36–47.
- Graf J. The effect of symbionts on the physiology of *Hirudo medicinalis*, the medicinal leech. *Invertebr Reprod Dev.* 2002; 41:269–275.

11. Indergand S, Graf J. Ingested Blood Contributes to the Specificity of the Symbiosis of *Aeromonas veronii* Biovar *Sobria* and *Hirudo medicinalis*, the Medicinal Leech. *Appl Environ Microbiol.* 2000; 66:4735–4741. [PubMed: 11055917]
12. Semova I, et al. Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish. *Cell Host Microbe.* 2012; 12:277–88. [PubMed: 22980325]
13. Hajela N, et al. Gut microbiome, gut function, and probiotics: Implications for health. *Indian J Gastroenterol.* 2015; 34:93–107. [PubMed: 25917520]
14. Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci.* 2012; 13:701–12. [PubMed: 22968153]
15. Qin J, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature.* 2012; 490:55–60. [PubMed: 23023125]
16. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature.* 2006; 444:1022–3. [PubMed: 17183309]
17. Sokol H, et al. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A.* 2008; 105:16731–6. [PubMed: 18936492]
18. Jefferson R. The hologenome. *Agriculture, Environment and the Developing World: A Future of PCR.* 1994
19. Bordenstein SR, Theis KR. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol.* 2015; 13:e1002226. [PubMed: 26284777]
20. Ballal SA, Gallini CA, Segata N, Huttenhower C, Garrett WS. Host and gut microbiota symbiotic factors: lessons from inflammatory bowel disease and successful symbionts. *Cell Microbiol.* 2011; 13:508–17. [PubMed: 21314883]
21. Smith JDL, Gregory TR. The genome sizes of megabats (Chiroptera: Pteropodidae) are remarkably constrained. *Biol Lett.* 2009; 5:347–51. [PubMed: 19324635]
22. Putnam NH, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. 2015
23. Kojima KK, Jurka J. Crypton transposons: identification of new diverse families and ancient domestication events. *Mob DNA.* 2011; 2:12. [PubMed: 22011512]
24. Robertson HM. Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol Gen Genet.* 1996; 252:761–6. [PubMed: 8917322]
25. Aiweesakun P, Katzourakis A. Endogenous viruses: Connecting recent and ancient viral evolution. *Virology.* 2015; 479–480:26–37.
26. Hayward JA, et al. Identification of diverse full-length endogenous betaretroviruses in megabats and microbats. *Retrovirology.* 2013; 10:35. [PubMed: 23537098]
27. 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.* 2015; 89:5180–4. [PubMed: 25717107]
28. Hong W, Zhao H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. *Proc R Soc B Biol Sci.* 2014; 281:20141079–20141079.
29. Dotson CD, et al. Bitter taste receptors influence glucose homeostasis. *PLoS One.* 2008; 3:e3974. [PubMed: 19092995]
30. Phillips CD, Baker RJ. Secretory Gene Recruitments in Vampire Bat Salivary Adaptation and Potential Convergences With Sanguivorous Leeches. *Front Ecol Evol.* 2015; 3:122.
31. Gracheva EO, et al. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature.* 2011; 476:88–91. [PubMed: 21814281]
32. Wang Y. The functional regulation of TRPV1 and its role in pain sensitization. *Neurochem Res.* 2008; 33:2008–12. [PubMed: 18528757]
33. Phillips CD, et al. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol Ecol.* 2012; 21:2617–27. [PubMed: 22519571]
34. Apitz-Castro R, et al. Purification and partial characterization of draculin, the anticoagulant factor present in the saliva of vampire bats (*Desmodus rotundus*). *Thromb Haemost.* 1995; 73:94–100. [PubMed: 7740503]

35. Truelove E, Fielding AK, Hunt BJ. The coagulopathy and thrombotic risk associated with L-Asparaginase treatment in adults with acute lymphoblastic leukaemia. *Leukemia*. 2013; 27:553–559. [PubMed: 23099335]
36. 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*. 2007; 73:1984–91. [PubMed: 17277211]
37. Worthen PL, Gode CJ, Graf J. Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. *Appl Environ Microbiol*. 2006; 72:4775–81. [PubMed: 16820471]
38. Guyton AC, Coleman TG, Young DB, Lohmeier TE, DeClue JW. Salt balance and long-term blood pressure control. *Annu Rev Med*. 1980; 31:15–27. [PubMed: 6994602]
39. Bar-Peled L, Schweitzer LD, Zoncu R, Sabatini DM. Ragulator Is a GEF for the Rag GTPases that Signal Amino Acid Levels to mTORC1. *Cell*. 2012; 150:1196–1208. [PubMed: 22980980]
40. Donohoe DR, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab*. 2011; 13:517–26. [PubMed: 21531334]
41. Huq AH, Lovell RS, Ou CN, Beaudet AL, Craigen WJ. X-linked glycerol kinase deficiency in the mouse leads to growth retardation, altered fat metabolism, autonomous glucocorticoid secretion and neonatal death. *Hum Mol Genet*. 1997; 6:1803–9. [PubMed: 9302256]
42. August PR, et al. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem Biol*. 1998; 5:69–79. [PubMed: 9512878]
43. Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012; 1:18. [PubMed: 23587118]
44. Liu B, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. 2013:47.
45. Francischetti IMB, 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*. 2013; 82:288–319. [PubMed: 23411029]
46. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res*. 2002; 12:656–64. [PubMed: 11932250]
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–10. [PubMed: 2231712]
48. She R, Chu JS-C, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res*. 2009; 19:143–9. [PubMed: 18838612]
49. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res*. 2004; 14:988–95. [PubMed: 15123596]
50. Stanke M, et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006; 34:W435–9. [PubMed: 16845043]
51. The universal protein resource (UniProt). *Nucleic Acids Res*. 2008; 36:D190–5. [PubMed: 18045787]
52. Hunter S, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009; 37:D211–D215. [PubMed: 18940856]
53. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28:27–30. [PubMed: 10592173]
54. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31:3210–2. [PubMed: 26059717]
55. Smit A, Hubley R, Green P. RepeatMasker Open-3.0.
56. Jurka J. Repeats in genomic DNA: mining and meaning. *Curr Opin Struct Biol*. 1998; 8:333–7. [PubMed: 9666329]
57. Smit, A., Hubley, R. RepeatModeler Open-1.0.
58. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; 27:573–80. [PubMed: 9862982]

59. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 2007; 35:W265–8. [PubMed: 17485477]
60. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2014; 12:59–60. [PubMed: 25402007]
61. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–402. [PubMed: 9254694]
62. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005; 6:31. [PubMed: 15713233]
63. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30:772–80. [PubMed: 23329690]
64. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* 2014; 30:3276–8. [PubMed: 25095880]
65. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006; 22:2688–90. [PubMed: 16928733]
66. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010; 2010.pdb.prot5448.
67. Taboada B, et al. Is There Still Room for Novel Viral Pathogens in Pediatric Respiratory Tract Infections? *PLoS One.* 2014; 9:e113570. [PubMed: 25412469]
68. Camacho C, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10:421. [PubMed: 20003500]
69. Huson DH, Weber N. Microbial community analysis using MEGAN. *Methods Enzymol.* 2013; 531:465–85. [PubMed: 24060133]
70. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* 2009; Chapter 4:Unit 4.10.
71. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* 2000; 16:418–20. [PubMed: 10973072]
72. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 2010; 27:221–4. [PubMed: 19854763]
73. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9:772.
74. 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.* 2014; 42:D922–5. [PubMed: 24194607]
75. Jarvis ED, et al. Whole Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science.* 2014
76. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007; 24:1586–91. [PubMed: 17483113]
77. Teeling EC, et al. A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record. *Science.* 2005; 307
78. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006; 22:1269–71. [PubMed: 16543274]
79. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol.* 2015; 32:835–45. [PubMed: 25739733]
80. Zhang G, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* 2014; 346:1311–1320. [PubMed: 25504712]
81. 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.* 2009; 10:48. [PubMed: 19192299]
82. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015; 31:2745–7. [PubMed: 25851949]
83. Ranwez V, et al. OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol Biol.* 2007; 7:241. [PubMed: 18053139]

84. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32:1792–7. [PubMed: 15034147]
85. Guindon S, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59:307–21. [PubMed: 20525638]
86. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015; 10:845–858. [PubMed: 25950237]
87. Kang Y, et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature.* 2015; 523:561–7. [PubMed: 26200343]
88. He H, Yang T, Terman JR, 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.* 2009; 106:15610–5. [PubMed: 19717441]
89. 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.* 2010; 402:682–695. [PubMed: 20692269]
90. 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.* 2007; 282:4626–33. [PubMed: 17150966]
91. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30:2114–20. [PubMed: 24695404]
92. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; 27:863–4. [PubMed: 21278185]
93. Hasman H, et al. Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *J Clin Microbiol.* 2014; 52:3136–3136.
94. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012; 28:1420–8. [PubMed: 22495754]
95. Hyatt D, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010; 11:119. [PubMed: 20211023]
96. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010; 26:2460–1. [PubMed: 20709691]
97. Oksanen J, et al. vegan: Community Ecology Package. R package version 2.3-5. 2016

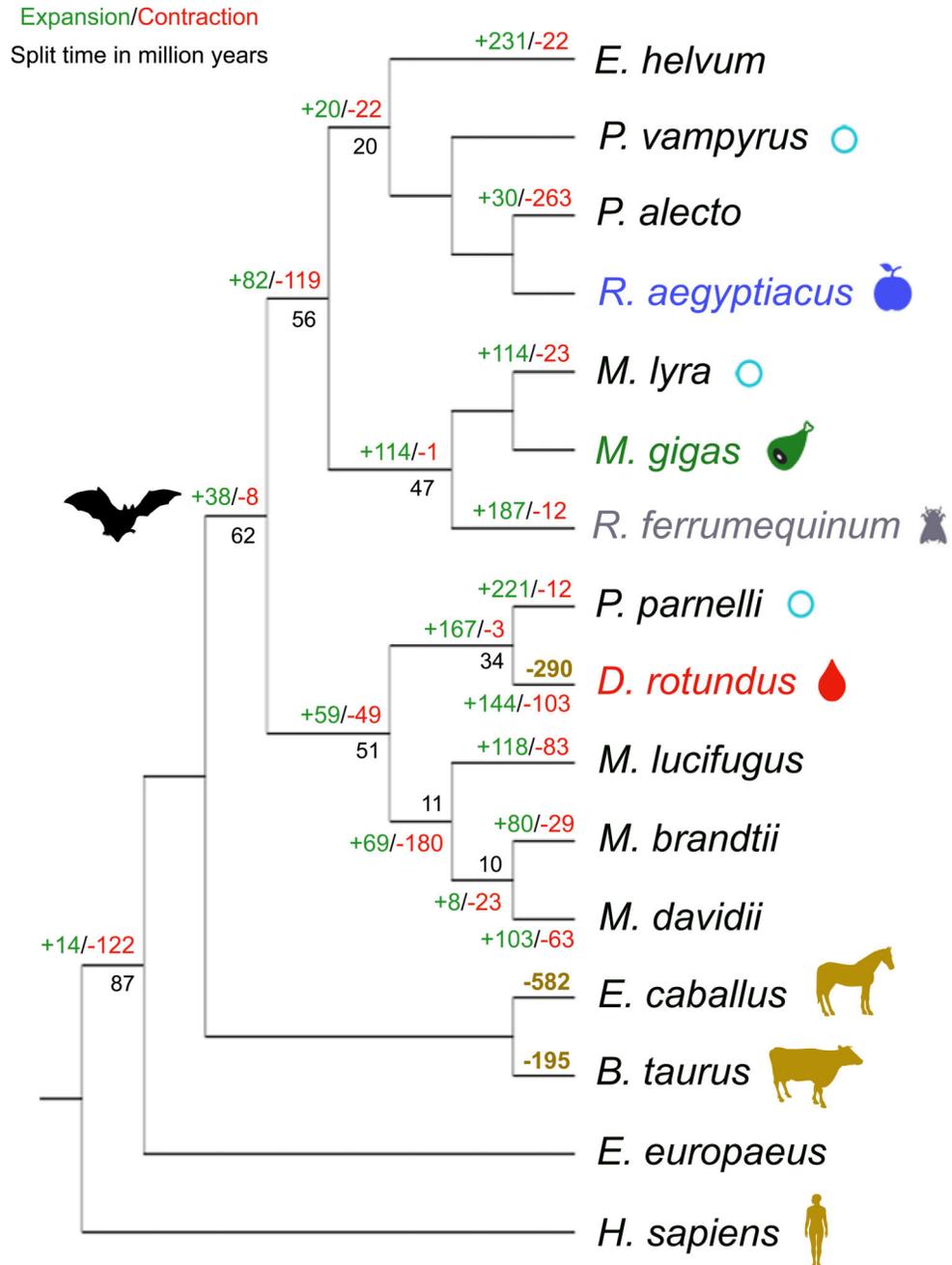


Figure 1. Comparative genomic analyses.

Species with a light blue circle were used for comparison of TEs. Species used for the identification of putative lost genes are represented with a golden silhouette and their corresponding number of putatively lost genes compared to the human genome. The number of gene families under expansion (green) and contraction (red) in Chiroptera using *E. europaeus* as outgroup are indicated. Estimated dates of nodes (in million years) are indicated in black. The gut microbiomes of bats with different diets derive from *Rousettus*

aegyptiacus (blue, frugivorous), *Macroderma gigas* (green, carnivorous), *Rhinolophus ferrumequinum* (grey, insectivorous), and the common vampire bat (red, sanguivorous).

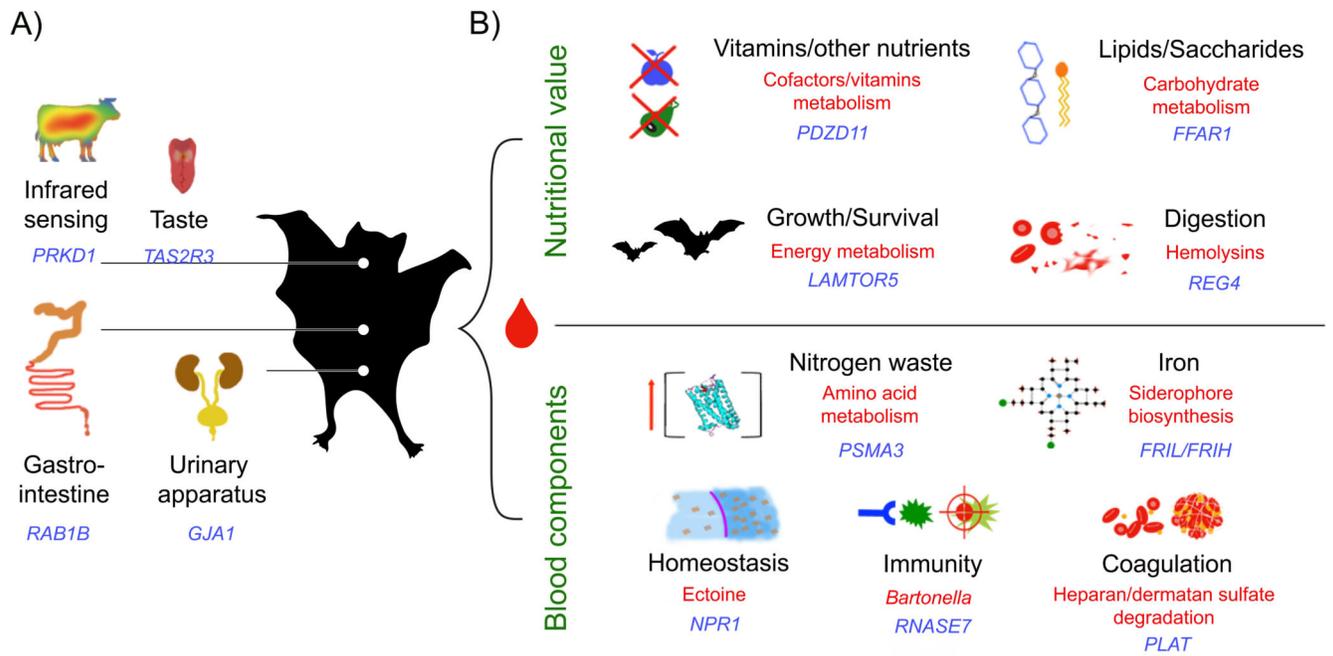


Figure 2. Dietary challenges overcome by the common vampire bat.

(**A**) Adaptational contributions to sanguivory accountable to genomic changes alone (blue labels) and (**B**) adaptational contributions to sanguivory within a hologenomic context (blue labels for host genes, red labels for gut microbial traits).

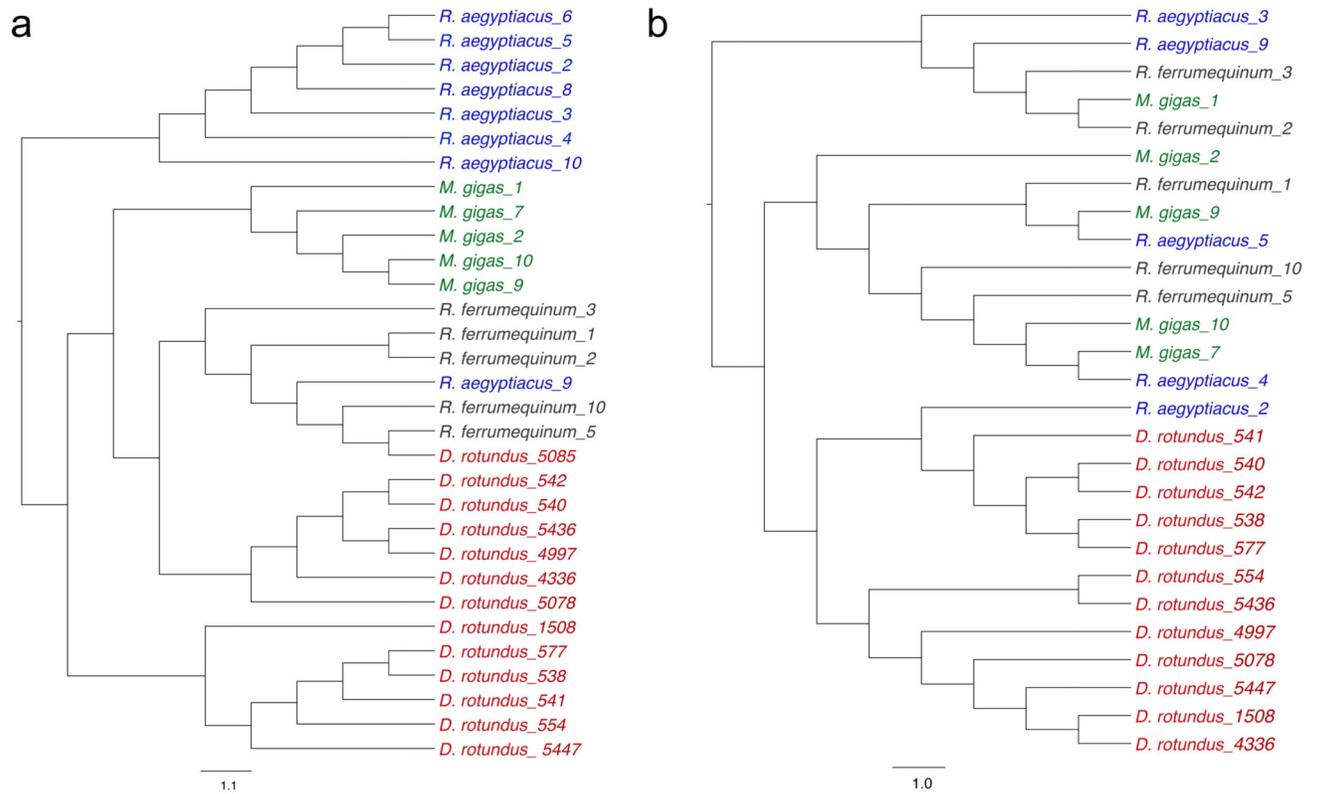


Figure 3. Comparison of the taxonomic and functional gut microbiome profiles. *D. rotundus* (sanguivorous, red), *R. ferrumequinum* (insectivorous, black), *M. gigas* (carnivorous, green), and *R. aegyptiacus* (frugivorous, blue). (A) Euclidean distance dendrogram of the microbial presence/absence identifications at the species taxonomical level. (B) Euclidean distance dendrogram from the Uniprot identified abundance functions from the normalized samples.

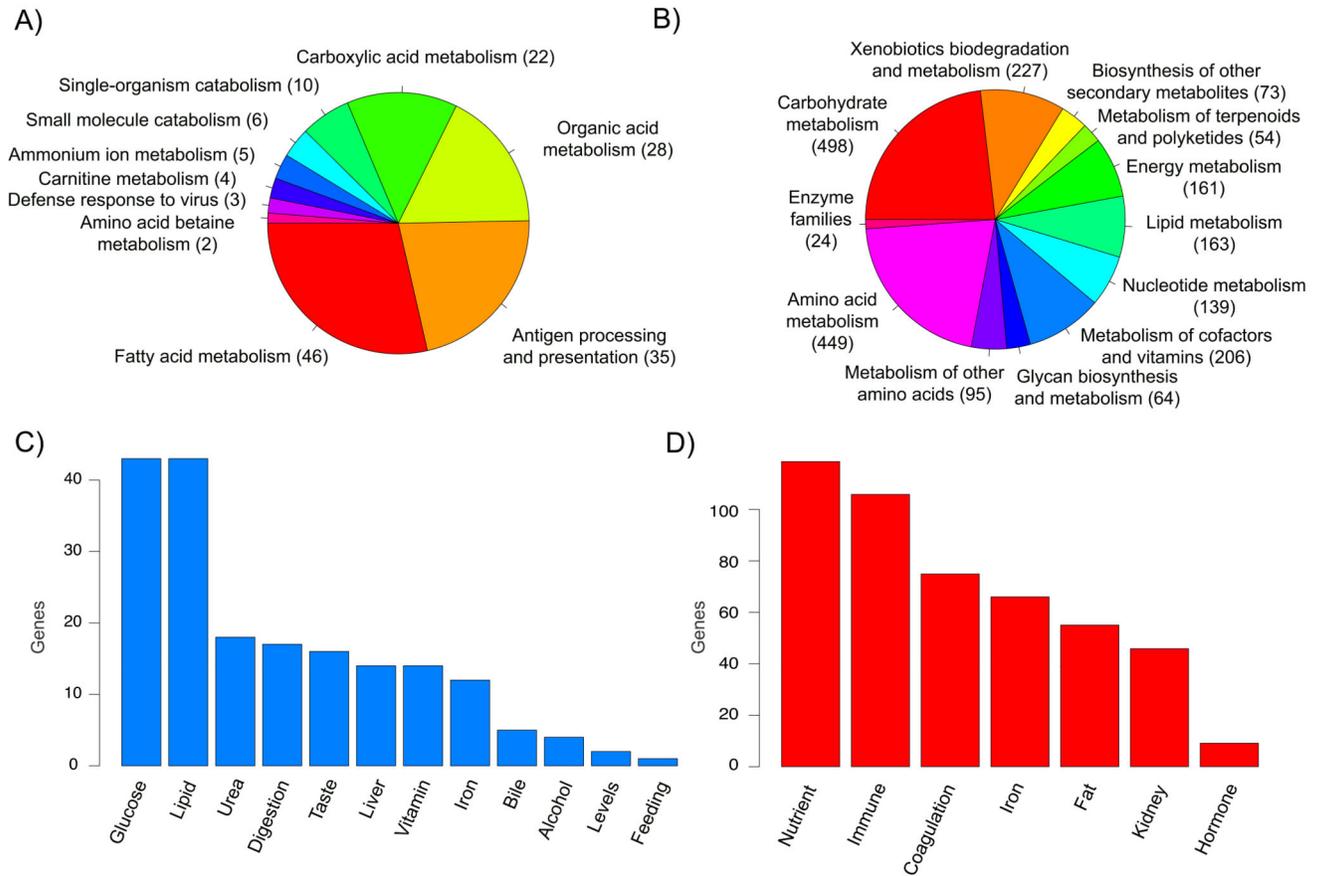


Figure 4. Traits in both the genome and gut microbiome with direct roles in the adaptation for sanguivory.

(A) GOs of the single copy orthologous genes with a dN/dS ratio that is statistically higher or lower in the common vampire bat in comparison to the other bats. (B) KEGG ontologies (KOs) from the common vampire bat gut microbiome functional core. (C) Genes with a dN/dS ratio that is statistically higher or lower in the common vampire bat in comparison to the other bats and that are directly associated to sanguivory challenges. (D) Taxa and genes from the common vampire bat gut microbiome core directly associated to the adaptation to sanguivory.