

# Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding

Moritz Blumer<sup>1,2,3,4,#</sup>, Tom Brown<sup>1</sup>, Mariella Bontempo Freitas<sup>5</sup>, Ana Luiza Destro<sup>5</sup>, Juraci A. Oliveira<sup>6</sup>, Ariadna Morales<sup>4,7,8</sup>, Tilman Schell<sup>7,8</sup>, Carola Greve<sup>7,8</sup>, Martin Pippel<sup>1</sup>, David Jebb<sup>1,2,3</sup>, Nikolai Hecker<sup>1,2,3</sup>, Alexis-Walid Ahmed<sup>4,7,8</sup>, Bogdan Kirilenko<sup>4,7,8</sup>, Maddy Foote<sup>9</sup>, Axel Janke<sup>7,10</sup>, Burton K. Lim<sup>11</sup> and Michael Hiller<sup>1,2,3,4,7,8,\*</sup>

<sup>1</sup> Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany.

<sup>2</sup> Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany.

<sup>3</sup> Center for Systems Biology Dresden, 01307 Dresden, Germany.

<sup>4</sup> Goethe-University, Faculty of Biosciences, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany

<sup>5</sup> Federal University of Viçosa, Department of Animal Biology, Viçosa, Brazil

<sup>6</sup> Federal University of Viçosa, Department of General Biology, Viçosa, Brazil

<sup>7</sup> LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany

<sup>8</sup> Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany

<sup>9</sup> Native Bat Conservation Program, Toronto Zoo, 361A Old Finch Avenue, Toronto, Canada

<sup>10</sup> Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

<sup>11</sup> Royal Ontario Museum, Department of Natural History, 100 Queen's Park, M5S 2C6 Toronto, Ontario, Canada

# Present address: Department of Genetics, University of Cambridge, Downing Street, CB2 3EH Cambridge, UK

\*To whom correspondence should be addressed:

Michael Hiller

LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany

Tel: +49 69 7542-1398

E-Mail: [Michael.Hiller@senckenberg.de](mailto:Michael.Hiller@senckenberg.de)

# Abstract

Feeding exclusively on blood, vampire bats represent the only obligate sanguivorous lineage among mammals. To uncover genomic changes associated with adaptations to this unique dietary specialization, we generated a new haplotype-resolved reference-quality genome of the common vampire bat (*Desmodus rotundus*) and screened 26 bat species for genes that were specifically lost in the vampire bat lineage. We discovered previously-unknown gene losses that relate to metabolic and physiological changes, such as reduced insulin secretion (*FFAR1*, *SLC30A8*), limited glycogen stores (*PPP1R3E*), and a distinct gastric physiology (*CTSE*). Other gene losses likely reflect the biased nutrient composition (*ERN2*, *CTRL*) and distinct pathogen diversity of blood (*RNASE7*). Interestingly, the loss of *REP15* likely helped vampire bats to adapt to high dietary iron levels by enhancing iron excretion and the loss of the 24S-hydroxycholesterol metabolizing enzyme *CYP39A1* could contribute to their exceptional cognitive abilities. Finally, losses of key cone phototransduction genes (*PDE6H*, *PDE6C*) suggest that these strictly-nocturnal bats completely lack cone-based vision. These findings enhance our understanding of vampire bat biology and the genomic underpinnings of adaptations to sanguivory.

# Introduction

Vampire bats are the only obligate sanguivorous lineage among tetrapods <sup>1</sup>. This exceptional dietary specialization is reflected in all aspects of their biology, including morphology, physiology and behavior <sup>2,3</sup>. To detect prey, the common vampire bat (*Desmodus rotundus*) exhibits a well-developed olfactory system <sup>4</sup>, advanced low-frequency hearing abilities <sup>5</sup> and, unique among mammals, the ability to sense infrared radiation <sup>6</sup>. Compared to other bats, vampire bats have exceptional terrestrial locomotion skills to sneak up on their prey <sup>7</sup>. Razor-sharp enamel-less upper incisors help cutting through the prey's skin and anticoagulants in their saliva prevent the prey's blood from coagulating during feeding <sup>8,9</sup>.

As the sole nutritional source, blood represents a challenging diet for several reasons. First, blood has a high fluid content of 78% and a comparatively low caloric value, making it necessary for a vampire bat to ingest as much as 1.4 times their body weight in blood during a single meal <sup>10,11</sup>. To enable the ingestion of large amounts of blood, their stomach experienced a functional shift towards a distensible organ primarily engaged in storage and fluid absorption <sup>12</sup>. Second, blood has a high iron content compared to other diets, which mainly stems from hemoglobin-derived heme and ferric iron transported by transferrin <sup>13-15</sup>. Third, the dry mass of blood has a highly skewed nutritional composition, providing mostly proteins (93%) with very little lipids and carbohydrates (1% each) <sup>11</sup>.

Due to a low carbohydrate intake, vampire bats exhibit lower basal insulin levels than other mammals <sup>10,11,16</sup>. Similar to human type 2 diabetes patients, vampire bats feature a reduced glucose-stimulated insulin secretion response, resulting in hyperglycemia upon an experimental glucose overload <sup>16</sup>. Glycogen and lipid stores are also reduced in vampire bats, which contributes to their fasting vulnerability and early deaths after 48-72 h of fasting <sup>17,18</sup>. To compensate for their fasting vulnerability, vampire bats share regurgitated blood with roost mates that failed to obtain a nightly meal <sup>19,20</sup>.

To gain insights into the molecular basis of vampire bat adaptations to sanguivory, comparative studies using the genome of the common vampire bat detected signatures of selection in genes involved in responses to nutrient starvation, metabolism, nitrogen waste disposal, the coagulation cascade, and immunity <sup>21,22</sup>. Common vampire bats have fewer *TAS2R* taste receptor genes than other bats, indicating a reduced sense of bitter taste reception <sup>23,24</sup>. In addition to genetic changes, vampire bats also possess a gut microbiome very different from that of other bats, and genes encoded by their gut microbiome further contribute to meeting the challenges of sanguivory <sup>21</sup>. Despite these advances, our understanding of which genomic changes are important for adaptations to sanguivory remains incomplete.

A limitation of previous comparative studies was the restricted taxonomic representation. For example, Zepeda Mendoza *et al.* <sup>21</sup> compared the *D. rotundus* genome to nine other bat genomes, with *D. rotundus* as the only representative of Phyllostomidae bats (Figure 1A). With this limited taxonomic resolution one cannot differentiate between those genomic changes that are shared between all phyllostomid bats (comprising more than 200 species) and those changes that evolved specifically in the vampire bat lineage and thus could be relevant for adaptations to sanguivory. To uncover genomic changes that evolved specifically in the vampire bat lineage, we generated a state-of-the-art haplotype-resolved chromosome-level assembly of *D. rotundus* and made use of many additional phyllostomid genomes that include the

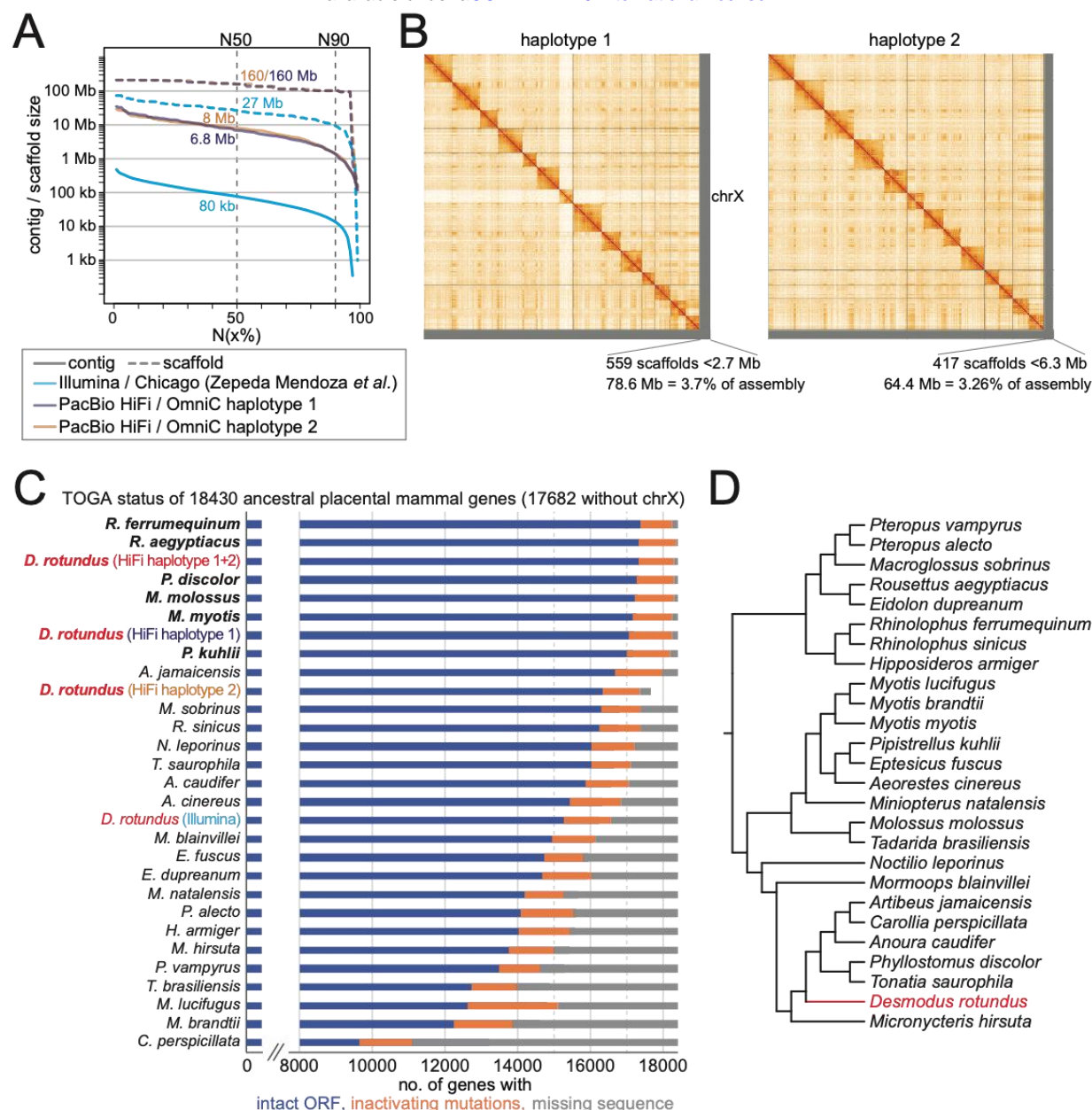
closest sister species of vampire bats (Supplementary Table 1). Since the loss of ancestral genes can be an important evolutionary force and previous studies revealed many associations between gene losses and phenotypic differences, including dietary adaptations<sup>25-29</sup>, we conducted a genome-wide screen for genes that are specifically lost in *D. rotundus* and that lack inactivating mutations in 25 other bat species. This screen revealed 3 known and 10 previously-unknown gene losses. Many of these novel gene losses have clear associations with vampire bat traits and some gene losses may contribute to coping with the challenges imposed by sanguivory.

## Results

### A new haplotype-resolved reference-quality *D. rotundus* assembly

To perform an accurate and comprehensive screen for gene losses in the common vampire bat, a genome assembly with high completeness, contiguity and base accuracy is desirable. The existing Illumina assembly of the common vampire bat<sup>21</sup> has more than 50,000 assembly gaps, indicating that a subset of genes will have missing sequences. This assembly also exhibits base errors that mimic gene-inactivating mutations, including cases where multiple putative mutations in the same gene are all erroneous (Supplementary Figure 1). Therefore, we generated a new reference-quality *D. rotundus* assembly. We used PacBio circular consensus (HiFi) sequencing<sup>30</sup> to produce 32X coverage in reads with an average length of 9.1 kb and used the Dovetail Omni-C protocol to produce 67X coverage of chromosome conformation capture Illumina read pairs. With these data, we obtained two haplotype-resolved chromosome-level assemblies that have only 596 and 557 assembly gaps, contig N50 values of 6.85 and 7.98 Mb and scaffold N50 values of 160.1 and 160.1 Mb (Figure 1A, Supplementary Table 1). The contig N50 metric is ~90-times higher than that of the Illumina-based *D. rotundus* assembly<sup>21</sup>. The karyotype of *D. rotundus* is 2n=28<sup>31</sup>. All 13 autosome pairs are represented by chromosome-level scaffolds (Figure 1B). Haplotype 1 contains the X chromosome as an additional chromosome-level scaffold. Haplotype 2 contains the Y chromosome, which was assembled as several scaffolds. Overall, more than 96% of both haplotypes are contained in chromosome-level scaffolds (Figure 1B). Using Merqury<sup>32</sup>, we estimated a very high base accuracy of QV 64.2 and 64.5 for both haplotypes, indicating one error per ~2.6 million base pairs.

To systematically assess gene completeness among the *D. rotundus* and other available bat genomes, we applied TOGA (Tool to infer Orthologs from Genome Alignments), a method that infers orthologs from whole genome alignments<sup>33</sup>, to 18,430 ancestral placental mammal coding genes (Methods). In comparison to the Illumina *D. rotundus* assembly, the number of ancestral genes with missing sequences dropped from 1,841 to 128 in our combined haplotype assemblies, whereas the number of genes with an intact reading frame increased from 15,295 to 17,301 (Supplementary Table 1). This indicates a substantially higher gene completeness in our assemblies, similar to the most contiguous bat assemblies generated to date<sup>34</sup> (Figure 1C).



**Figure 1:** Haplotype-resolved chromosome-level assembly of *Desmodus rotundus*.

(A) Comparison of assembly contiguity between our haplotype-resolved assemblies and the previous Illumina-based assembly<sup>21</sup>. The graph shows contig (solid lines) and scaffold (dashed lines) sizes at the Y-axis, for which x per cent of the assembly consists of contigs and scaffolds of at least that size.

(B) HiC contact maps show 14 and 13 chromosome-level scaffolds for haplotype 1 and 2 that each comprise more than 96% of the respective haplotype assembly. Haplotype 1 contains the X chromosome. The Y chromosome is part of haplotype 2 and assembled into several shorter scaffolds.

(C) Status of 18,430 ancestral placental mammal genes in each assembly, as inferred by TOGA. Genes are classified into those that have an intact reading frame (blue), have inactivating mutations (orange) or have missing coding parts due to assembly gaps or fragmentation (grey). Assemblies are sorted by the number of intact genes. Long read-based assemblies (bold font) consistently exhibit more intact genes and fewer missing genes compared to short read-based assemblies (not bold). For the *D. rotundus* haplotype 2 assembly that does not contain the X chromosome, we excluded genes located on the X chromosome and only considered the remaining 17,682 genes. To provide a fair comparison with previous assemblies that collapse both haplotypes, we also computed statistics for the union of both *D. rotundus* haplotype assemblies (haplotype 1+2), which exhibits the third highest number of intact genes of all included genomes.

(D) Phylogeny of the bats analyzed in this study<sup>1,42,43</sup>.

## A genome-wide screen revealed known and novel vampire bat-specific gene losses

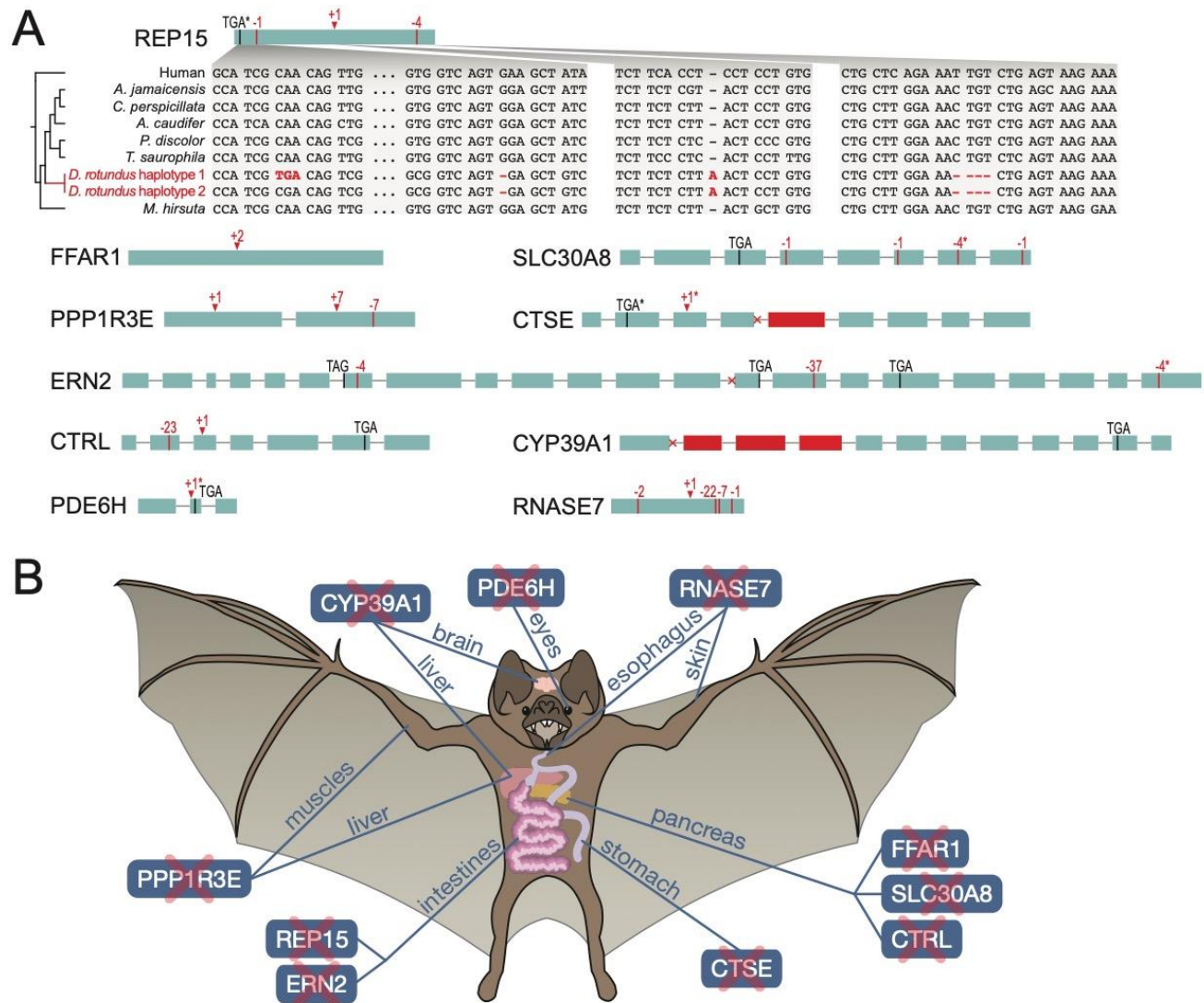
To detect genes that were specifically lost in the vampire bat lineage, we used our two haplotype assemblies and considered 25 other bat genomes<sup>21,23,34-41</sup> (<http://dnazoo.org/>), including *Micronycteris hirsuta* and *Phyllostomus discolor*, which represent the closest phylogenetic outgroup and sister lineages of vampire bats, respectively<sup>1,42,43</sup> (Figure 1D). Using the human gene annotation as the reference, we applied TOGA to detect genes that exhibit gene-inactivating mutations (premature stop codons, frameshifts, splice site disruptions, and deletions of exons or entire genes) specifically in *D. rotundus*.

This screen revealed 13 vampire bat-specific gene losses. Three of these losses have been reported before: the sweet taste receptor gene *TAS1R2*<sup>44</sup> and the bitter taste receptor genes *TAS2R5* and *TAS2R42*<sup>24</sup>. These gene losses indicate a reduced sense of taste reception in vampire bats. To our knowledge, the remaining 10 gene losses (*REP15*, *FFAR1*, *SLC30A8*, *PPP1R3E*, *CTSE*, *ERN2*, *CTRL*, *CYP39A1*, *PDE6H*, *RNASE7*) have not been reported before.

The inactivating mutations in these 10 genes are displayed in Figure 2A. In support of the validity of the underlying mutations, the 10 gene losses are also detected in the Illumina *D. rotundus* assembly and exhibit at least one shared inactivating mutation that is supported by both PacBio HiFi and Illumina reads. Further supporting gene loss, selection rate analysis using RELAX<sup>45</sup> showed that 9 of the 10 genes evolve under relaxed selection in *D. rotundus* to preserve the reading frame (significant for 7 of 9 genes, Supplementary Table 2). Finally, inspecting available *D. rotundus* RNA-seq data, we either found no relevant expression in tissues where expression would be expected or found that the aligned reads supported the gene-inactivating mutations, indicating that potential transcripts cannot be translated into a full length protein (Supplementary Figure 2).

Intriguingly, as detailed below, the loss of these genes in the common vampire bat relates to a number of adaptations to their highly specialized blood diet, such as reduced insulin secretion and glycogen synthesis and a distinct stomach physiology (Figure 2B). These gene losses further provide a possible mechanism contributing to vampire bats' exceptional social behavior, and indicate the complete lack of cone photoreceptor function in *D. rotundus*.





**Figure 2:** Inactivating mutations and affected organ systems of 10 previously-unknown vampire bat-specific gene losses. (A) Exon-intron structure visualization with inactivating mutations that are detected in the *Desmodus rotundus* genome. Premature stop codons are shown as black vertical lines, frameshifting deletions as red vertical lines and frameshifting insertions as red arrow heads. Donor or acceptor splice site mutations are indicated as a cross at the exon boundaries. Deleted exons are in red. Asterisks denote mutations that are heterozygous in our sequenced *D. rotundus* individual (present in only one of the two haplotype assemblies). The inset for *REP15* illustrates that inactivating mutations and thus gene loss was only detected in the common vampire bat. (B) Illustration of organs and anatomical sites where the 10 genes play important roles.

## Loss of *REP15* and enhanced iron excretion

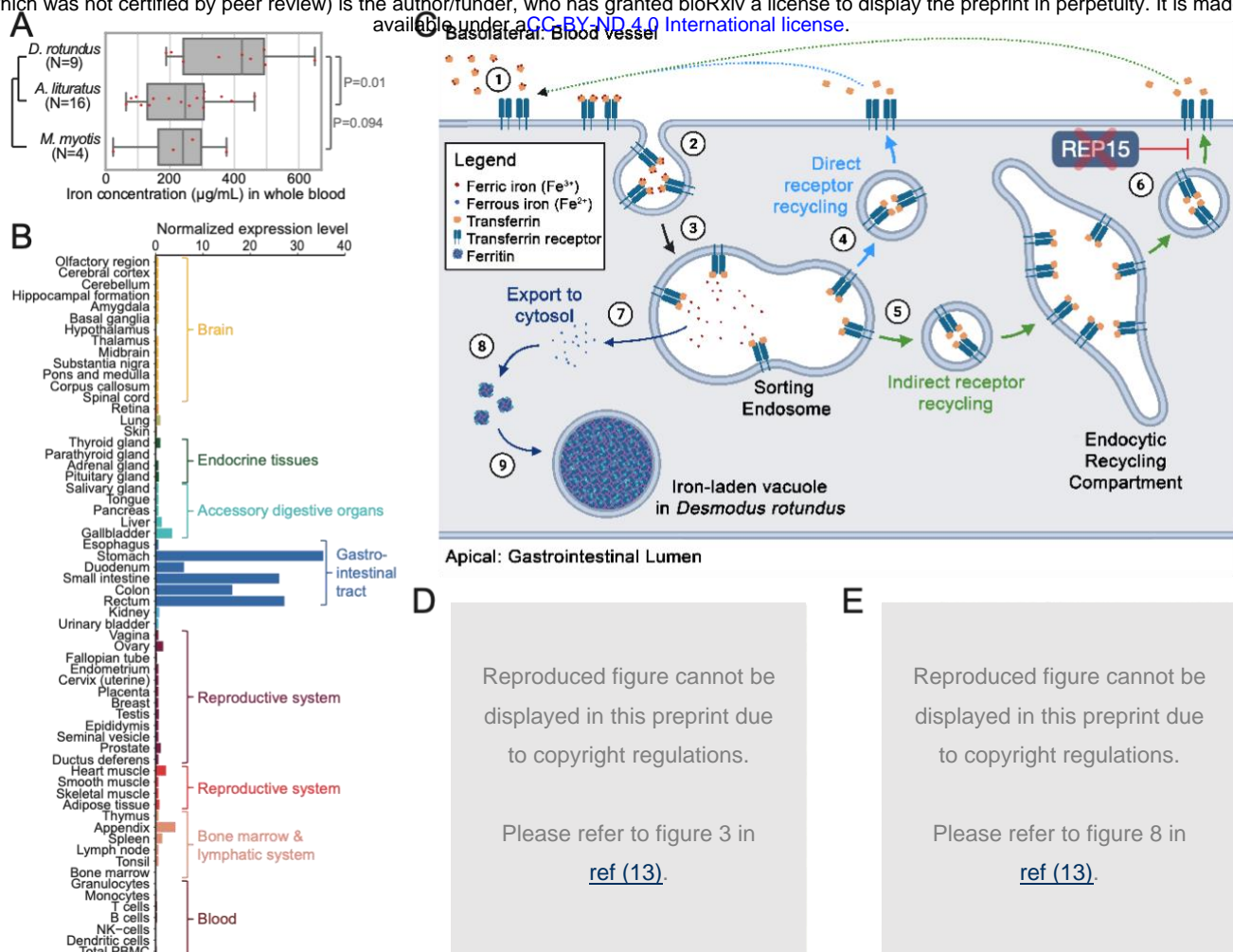
The loss of *REP15* (RAB15 effector protein), a gene involved in regulating cellular iron uptake <sup>46</sup>, is likely related to the obligatory iron-rich blood diet of vampire bats. Despite the importance of iron for various cellular processes, iron overload can have severe detrimental effects <sup>15</sup>. Remarkably, the common vampire bat tolerates extreme dietary iron levels without exhibiting adverse effects <sup>47</sup> – the relative amount of dietary iron was estimated to be 800-fold higher compared to humans <sup>13</sup>. However, blood iron concentration in vampire bats has never been determined to our knowledge. Therefore, we measured whole blood iron concentrations in wild *D. rotundus* and compared it to two other neotropical bats (Supplementary Table 3). We found that blood iron content was significantly higher in the common vampire bat compared to the fruit-eating bat *Artibeus lituratus* (family Phyllostomidae; t-test:  $P=0.01$ , Figure 3A). Vampire bats also had higher blood iron levels compared to the insectivorous outgroup bat *Myotis myotis* (family Vespertilionidae; not significant, likely due to the small sample size of four individuals; Figure 3A). This suggests that vampire bats would benefit from mechanisms to lower systemic iron levels.

*REP15* is specifically expressed in the gastrointestinal tract <sup>48</sup> (Figure 3B). Cellular overexpression of *REP15* decreases the amount of iron-transporting transferrin receptors on the cell surface <sup>46</sup> (Figure 3C). Since transferrin receptor availability is a limiting factor for cellular iron uptake <sup>49</sup>, *REP15* likely inhibits the uptake of iron from the bloodstream into gastrointestinal cells. Consistent with this inhibitory effect, the downregulation of *REP15* in colorectal cancer cells coincides with increased intracellular iron levels in these cells <sup>50,51</sup>.

Why is a gene that inhibits cellular iron uptake specifically lost in *D. rotundus*? We propose that *REP15* loss represents a strategy to boost iron excretion and thus helps vampire bats to reduce systemic iron levels. Loss of *REP15* is expected to enhance the accumulation of iron in cells of the gastrointestinal tract, where it is specifically expressed. Because the intestinal epithelium has a relatively fast turnover time <sup>52</sup>, iron is subsequently eliminated by shedding iron-containing gastrointestinal cells. This general mechanism has been recently suggested as an important factor controlling iron loss in mammals <sup>53</sup>.

Intriguingly, a 1980 study on the distribution of iron in the gastrointestinal tract of *D. rotundus* made observations that precisely match this hypothesis <sup>13</sup>. Using histochemical staining techniques along with electron microscopy and radiography, large accumulations of iron in ferritin-containing vacuoles were identified in gastrointestinal epithelial cells that frequently desquamated into the intestinal lumen <sup>13</sup> (Figure 3D,E). This observation is consistent with enhanced cellular iron uptake in gastrointestinal cells of vampire bats to which *REP15* loss could contribute. In addition to enhanced excretion, a previous study found that vampire bats limit gastrointestinal iron absorption by increased expression of hepcidin, a factor that inhibits intestinal iron absorption <sup>47</sup>. Furthermore, iron-storing ferritin genes are expanded in the common vampire bat genome <sup>21</sup> and these bats have high levels of iron-binding RFESD in the serum proteome <sup>54</sup>. Thus, limited iron absorption (mediated by increased hepcidin expression), a higher capacity for iron storage (mediated by ferritin and RFESD), and enhanced iron excretion (mediated by inactivating the inhibitory factor *REP15*) help vampire bats to cope with their iron-rich diet.





**Figure 3: Loss of *REP15* in *Desmodus rotundus* and enhanced iron excretion.**

(A) Measurements of iron concentration in whole blood of *D. rotundus*, *Artibeus lituratus* and *Myotis myotis* show that vampire bats have higher circulatory iron levels. P-values are computed with a two-sided t-test.

(B) *REP15* mRNA expression is highest in gastrointestinal tract tissues. Data was taken from the Human Protein Atlas <sup>48</sup> (<https://www.proteinatlas.org/ENSG00000174236-REP15/tissue>) and show the consensus RNA expression values that integrate three gene expression datasets.

(C) Illustration of transferrin receptor-mediated cellular iron uptake and function of *REP15* in intestinal epithelial cells. Transferrin, an abundant ferric iron-binding plasma protein, binds to transferrin receptors that are present only in the basolateral membrane (1). Transferrin-transferrin receptor complexes are internalized via endocytosis (2). In sorting endosomes, ferric iron is released (3) and the unladen complexes are either directly targeted back to the cell membrane (4) or sent to the endocytic recycling compartment (5). *REP15*, encoded by the gene that is lost in *D. rotundus*, specifically localizes to the endocytic recycling compartment (6) and inhibits recycling of the unladen complex to the cell membrane <sup>46</sup>, where transferrin and its receptor dissociate and the released transferrin can bind ferric iron again. Because the availability of transferrin receptors on the cell surface limits iron uptake <sup>49</sup>, the presence of *REP15* normally inhibits cellular iron uptake. In the sorting endosome, ferric iron (Fe<sup>3+</sup>) is reduced to ferrous iron (Fe<sup>2+</sup>) and exported to the cytosol (7), where ferritin acts as the major high-capacity iron storage protein (8). Interestingly, accumulations of ferritin and other "ferruginous" complexes enclosed in vacuoles were observed in intestinal epithelial cells of *D. rotundus* (9, panel D). Loss of *REP15* likely enhances iron accumulation in intestinal epithelial cells and shedding of these cells boosts iron excretion in *D. rotundus*.

(D) Light microscopy image, reproduced from Figure 3 in <sup>13</sup>, showing a longitudinal section of the upper villus half from the *D. rotundus* ileum. Prussian blue staining that indicates iron demonstrates the presence of iron-containing cytoplasmic granules in epithelial cells (arrow). In addition to delivering iron via the bloodstream, a macrophage-linked mechanism contributes to iron deposition in these epithelial cells <sup>13</sup>.

(E) Prussian blue-positive granules are present in the forming feces of *D. rotundus*, showing that these bats excrete iron by shedding iron-containing intestinal cells. The figure is reproduced from Figure 8 in <sup>13</sup>.

### Losses of *FFAR1* and *SLC30A8* and reduced insulin secretion

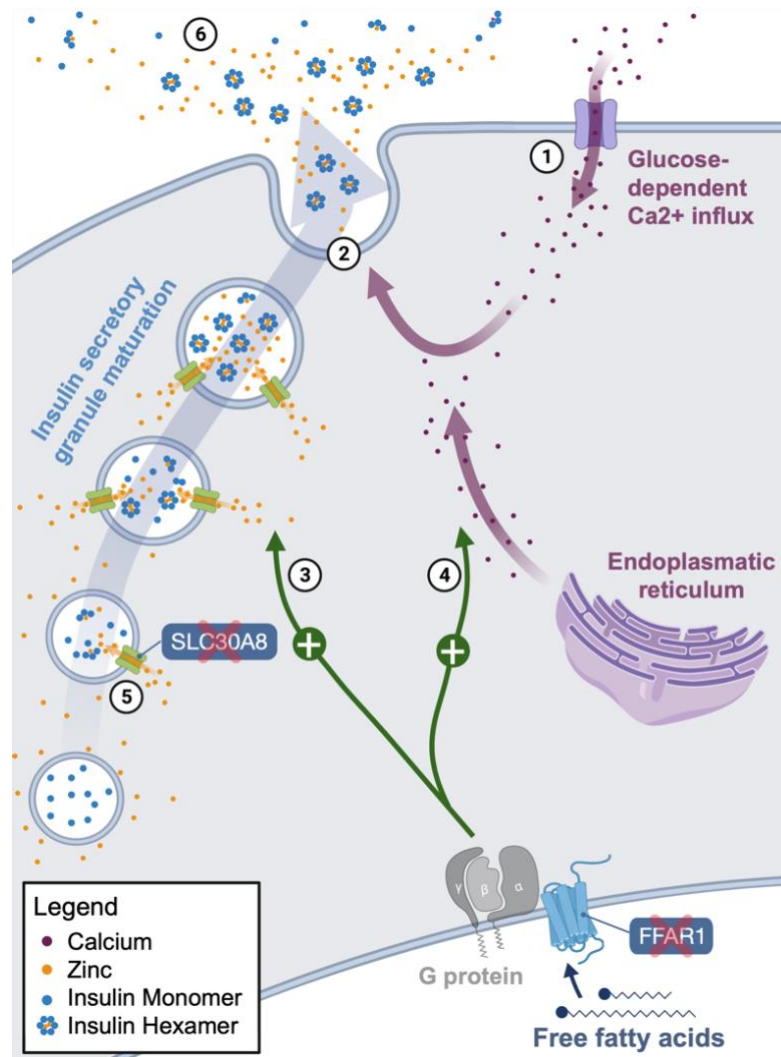
*FFAR1* (free fatty acid receptor 1) encodes a G protein-coupled receptor that is highly expressed in pancreatic beta cells and senses medium to long chain free fatty acids<sup>48,55</sup>. Free fatty acids augment glucose-stimulated insulin secretion and *FFAR1* is a major factor that mediates this effect<sup>55,56</sup> (Figure 4). Glucose-stimulated insulin secretion involves an initial, rapid phase, which appears to be largely dependent on the depletion of an existing pool of membrane-docked insulin secretory granules, and a second, prolonged phase, which requires the production of new granules to supply the ongoing secretion process<sup>55,57</sup>. Deletion of *FFAR1* in mice reduces the amplifying effect of free fatty acids in the prolonged phase by ~50%<sup>56,58</sup>. Consistent with this, *FFAR1* expression levels in human islets positively correlate with insulin secretion, which suggested that *FFAR1* deficiency could lower the insulin secretory capacities of beta cells and thus contribute to the development of type 2 diabetes<sup>59,60</sup>. In addition to directly stimulating insulin secretion in beta cells, *FFAR1* also amplifies insulin secretion indirectly<sup>55</sup>. First, *FFAR1* is expressed in enteroendocrine cells, where its activation triggers the release of incretin hormones that stimulate pancreatic insulin secretion<sup>61</sup>. Second, *FFAR1* is expressed in the brain and likely functions as a lipid sensor that influences insulin secretion through innervation of pancreatic islets<sup>62</sup>. Hence, *FFAR1* amplifies insulin secretion via multiple mechanisms.

The free fatty acid sensor *FFAR1* likely became dispensable for vampire bats for two reasons. First, the *FFAR1* ligand (fatty acids) is barely present in their blood diet. Second, the low dietary sugar content likely abolishes the need for a normal glucose-stimulated insulin secretion, including the prolonged insulin secretion phase to which *FFAR1* substantially contributes. Indeed, vampire bats exhibit lower basal insulin levels than other mammals and show a substantially reduced insulin secretion with prolonged hyperglycemia upon an experimental glucose overload<sup>10,11,16</sup>. While having similarities to the defective insulin secretion response in human type 2 diabetes patients, in vampire bats, this is likely an adaptation to prevent hypoglycemia, since glucose must remain in the bloodstream. Whereas a previous study detected signatures of positive selection in *FFAR1* and interpreted it as a means to improve glucose utilization in *D. rotundus*<sup>21</sup>, our analysis reveals that *FFAR1* is actually lost and suggests that this gene loss relates to reduced insulin secretion under normal conditions and the inability to cope with high glucose intake when experimentally challenged.

*SLC30A8* (solute carrier family 30 member 8) encodes the most highly expressed zinc transporter in pancreatic beta cells<sup>63,64</sup>. In secretory granules of beta cells, zinc is required for the formation of zinc-insulin hexamers (Figure 4), which protect insulin from degradation<sup>65,66</sup>. Therefore, loss of the zinc transporter *SLC30A8* may have been a consequence of or may contribute to the decreased insulin secretion phenotype of *D. rotundus*.

Another, not mutually exclusive hypothesis is that *SLC30A8* loss in vampire bats may be beneficial by reducing the requirements for body zinc levels. Even though sufficient dietary zinc is available from mammalian blood<sup>67,68</sup>, it might not be efficiently absorbed by vampire bats, since high dietary iron concentrations can inhibit zinc absorption<sup>69</sup> and zinc in turn stimulates iron absorption in intestinal cells<sup>70</sup>. Since blood has a high iron:zinc ratio of 66:1<sup>67</sup>, high dietary iron levels could induce zinc deficiency in vampire bats. Interestingly, *SLC30A8* knockout in mice causes extremely low zinc concentrations in beta cells<sup>71</sup>. Similarly, “natural *SLC30A8* knockouts” in mammalian herbivores such as guinea pig, sheep and

cow that also lost this gene <sup>72</sup> are associated with low pancreatic islet zinc concentrations <sup>73-75</sup>. Since beta cells are normally among the cell types with the highest zinc concentration <sup>76</sup>, it is possible that these herbivores may have lost *SLC30A8* because vegetarian diets are generally more limited in zinc content <sup>77,78</sup>. Repeated loss of *SLC30A8* in the vampire bat and in herbivorous mammals could therefore represent a strategy to restrict zinc utilization to other essential physiological functions. A precondition for this strategy would be that insulin stability and secretion does not depend on zinc anymore, as shown in guinea pigs, or that insulin secretion would not be essential anymore, as it seems to be the case for vampire bats.



**Figure 4:** Losses of *SLC30A8* and *FFAR1* in *Desmodus rotundus* relate to reduced insulin secretion.

Illustration of the roles of both genes in insulin synthesis and secretion in pancreatic beta cells. Glucose-stimulated insulin secretion is initiated by the opening of voltage-gated calcium channels (1) <sup>55</sup>. The resulting influx of calcium enhances insulin secretion by stimulating exocytosis of insulin secretory granules (2). *FFAR1* encodes a G protein-coupled receptor that upon activation enhances the prolonged insulin secretion phase via (F)-actin remodeling (3). *FFAR1* also triggers the release of calcium from the endoplasmic reticulum (4), which adds to the glucose stimulated calcium influx and amplifies exocytosis of secretory granules. *SLC30A8* transports zinc into maturing insulin secretory granules, where zinc is essential for the formation of insulin hexamers (5). Upon secretion, these hexamers dissociate, releasing bioactive insulin monomers and zinc into the circulation (6). Loss of *FFAR1* and *SLC30A8* in *D. rotundus* is likely related to vampire bats' reduced insulin secretion.

### Loss of *PPP1R3E* and impaired glycogen metabolism

*PPP1R3E* (protein phosphatase 1 regulatory subunit 3E) encodes a regulatory subunit of protein phosphatase 1 (PP1)<sup>79</sup>. PP1 plays a central role in regulating a switch between synthesis and breakdown of glycogen, the major short-term storage form of glucose<sup>80</sup>. By dephosphorylating glycogen synthase, PP1 activates this enzyme and promotes glycogen synthesis<sup>81</sup>. By dephosphorylating glycogen phosphorylase, PP1 inhibits this enzyme and consequently glycogen breakdown<sup>81</sup> (Supplementary Figure 3A). The activity of PP1 depends on a regulatory subunit, which is encoded by seven different genes with different expression patterns (*PPP1R3A* - *PPP1R3G*)<sup>79</sup>. These regulatory subunits are critical for PP1 activity as overexpression of *PPP1R3A*, *PPP1R3B*, *PPP1R3C*, or *PPP1R3G* in cellular or animal models increases glycogen content, whereas their knockout reduces glycogen content<sup>82-86</sup>. Even though no animal studies of *PPP1R3E* exist, the gene is transcriptionally regulated by insulin and *PPP1R3E* binds to glycogen<sup>79</sup>. Thus, *PPP1R3E* most likely functions like the other glycogen-targeting subunits. With the exception of *PPP1R3E* loss in *D. rotundus*, none of the seven regulatory subunit-encoding genes exhibit inactivating mutations in any analyzed bat (Supplementary Figure 3B).

The specific loss of *PPP1R3E* likely relates to the low glycogen concentrations in *D. rotundus*<sup>17</sup>. In fact, fed vampire bats have hepatic glycogen stores that are ~85% smaller than in fruit-eating bats and ~60% smaller than in other mammals fed on high-protein diets<sup>17</sup>. Since sufficient glycogen stores are important to withstand periods of fasting, loss of *PPP1R3E* and the associated smaller glycogen stores may also contribute to the observed starvation vulnerability of vampire bats.

### Loss of *CTSE* and altered stomach function

The loss of *CTSE* (cathepsin E) in *D. rotundus* may be a consequence of extensive morphological and physiological modifications of their stomach, which is unparalleled by any other mammalian species<sup>12,87</sup>. Most notably, the stomach of *D. rotundus* experienced fundamental remodeling from a compact muscular organ involved in mechanical and chemical digestion towards a distensible structure that functions primarily to store large amounts of ingested blood and serves as a major site of fluid absorption<sup>12,87,88</sup> (Supplementary Figure 4). *CTSE* is an intracellular protease that is highly expressed in gastrointestinal tissues, particularly in the stomach<sup>48,89</sup>. In the stomach, *CTSE* protein normally localizes to the canaliculi (invaginations) of parietal cells, which secrete hydrochloric acid, indicating that *CTSE* might be involved in secretory processes<sup>90</sup>. In *D. rotundus*, the canaliculi of parietal cells are specifically enriched in iron<sup>13</sup>, suggesting that their gastric parietal cells experienced a functional shift from hydrochloric acid secretion towards participating in iron excretion, which could have rendered *CTSE* dispensable. Consistent with this, *CTSE* has been reported lost in platypus<sup>91</sup>, which also lacks an acid-secreting stomach. However, *CTSE* is also lost in *Cetartiodactyla* that possess acid-secreting stomachs<sup>92</sup> and, in addition to a possible role in gastric acid secretion, *CTSE* also has immune-related functions<sup>89</sup>. Thus, other *CTSE* functions could have also contributed to or represent the primary reason of *CTSE* loss in *D. rotundus*.

### Loss of *ERN2* and low dietary fat content

The loss of *ERN2* (endoplasmic reticulum to nucleus signaling 2) in *D. rotundus* is likely a consequence of the low fat content of their blood diet. *ERN2* is specifically expressed in gastrointestinal epithelia<sup>93</sup> and encodes a transmembrane protein that inhibits the production of chylomicrons, which are lipoprotein particles that transport dietary lipids absorbed in enterocytes to other tissues<sup>94,95</sup>. Enterocytes of *ERN2*



knockout mice were shown to secrete more chylomicrons, resulting in increased lipid absorption and a more pronounced hyperlipidemia on high-fat diets<sup>94,95</sup>. Interestingly, on a standard diet, *ERN2* knockout mice did not exhibit hyperlipidemia, indicating that *ERN2* limits intestinal lipid absorption only under conditions where excessive fat is available. Since blood has a very low fat content<sup>11</sup>, it is conceivable that regulatory mechanisms, which normally limit lipid absorption, became dispensable during vampire bat evolution, resulting in loss of *ERN2*.

### **Loss of the pancreatic chymotrypsin *CTRL***

Pancreatic proteases such as chymotrypsinogens and trypsinogens are among the most essential digestive enzymes<sup>96</sup>. They are produced by acinar cells of the exocrine pancreas and are secreted as inactive zymogens which, upon activation in the small intestine, digest proteins<sup>96</sup>. *CTRL* (chymotrypsin like) encodes one of four chymotrypsinogen isozymes and is considered a predictive biomarker for human pancreatic cancer<sup>97</sup>. In mice, *CTRL* represents a minor chymotrypsin isoform, constituting ~10% of the chymotrypsinogen pool<sup>98</sup>. *In vitro* experiments showed that *CTRL* cleaves trypsinogens and thus inhibits the activation of protein-digesting trypsin<sup>98</sup>. Consistently, *CTRL* knockout in mice leads to a reduced activation of chymotrypsins and a slightly higher trypsin activity<sup>98</sup>, which could be relevant for the unique, protein-rich diet of vampire bats. Interestingly, recent evidence also indicates that acinar proteases including *CTRL* contribute to proliferation of beta cells<sup>99</sup>, providing another possible explanation for the loss of this gene in a bat species that exhibits a reduced beta cell mass<sup>16</sup>.

### **Loss of *CYP39A1* and advanced social behavior**

The loss of *CYP39A1* (cytochrome P450 family 39 subfamily A member 1) could have contributed to the evolution of vampire bat's exceptional social behaviors and cognitive abilities<sup>20,100,101</sup>. *CYP39A1* encodes an oxysterol 7- $\alpha$ -hydroxylase enzyme<sup>102</sup>. Apart from a minor contribution to hepatic bile acid synthesis (Supplementary Figure 5), *CYP39A1* is the only enzyme expressed in the brain that can degrade the cholesterol metabolite 24S-hydroxycholesterol<sup>103-106</sup>. 24S-hydroxycholesterol is an intermediate by-product during cholesterol elimination from the brain<sup>107</sup>. Unlike cholesterol, 24S-hydroxycholesterol can pass the blood-brain barrier. Therefore, the loss of the 24S-hydroxycholesterol-degrading *CYP39A1* is expected to result in elevated systemic levels of this metabolite, which was observed in humans with *CYP39A1* loss-of-function alleles<sup>108</sup>.

Since 24S-hydroxycholesterol also plays important neurophysiological roles, we hypothesize that *CYP39A1* loss may be connected to the exceptional cognitive and social capabilities of *D. rotundus*. 24S-hydroxycholesterol is a potent allosteric activator of *N*-methyl-D-aspartate receptors (NMDAR), which are glutamate-gated ion channels that mediate synaptic plasticity and memory formation<sup>109,110</sup>. Consistent with a positive effect of 24S-hydroxycholesterol on NMDAR activity and cognitive function, increased cerebral 24S-hydroxycholesterol levels improve spatial memory function in mice<sup>111,112</sup>, whereas reduced levels cause impaired learning and memory function<sup>113,114</sup>. NMDAR activity also exerts a strong influence on social behavior, as NMDAR agonists improve social behavior in rodents, whereas antagonists impair it<sup>115</sup>.

Intriguingly, vampire bats, the only lineage in our dataset that lost the 24S-hydroxycholesterol-metabolizing *CYP39A1* gene, are distinguished from other bats by their exceptional social behavior and cognitive abilities. They reciprocally share regurgitated blood with roost mates that failed to obtain a nightly meal and



would otherwise face starvation<sup>19</sup>. The decision with whom to share blood is primarily driven by recognizing contact calls of individuals that provided past blood donations, which demonstrates exceptional long term social memory<sup>20,100</sup>. Furthermore, non-kin adoption of orphaned offspring occurs in vampire bat colonies<sup>101</sup>. Vampire bats also form long-lasting social bonds, evident by observations that individuals who cooperated in captivity retain their social network when released into the wild<sup>116</sup>. Vampire bats, as perhaps the most socially-advanced bats, were also found to have the largest relative neocortical volume among 276 measured bat species<sup>117</sup>. Interestingly, the 24S-hydroxycholesterol-interacting subunit of NMDAR (*GLUN2B*) is highly expressed at birth and plays an important role in cortical development<sup>110,118</sup>, raising the possibility that *CYP39A1* loss may also be connected to the large neocortical volume of *D. rotundus*.

### Loss of cone phototransduction genes suggests rod monochromacy

Our strict genomic screen revealed that *PDE6H* is exclusively lost in vampire bats. *PDE6H* encodes the  $\gamma$ -subunit of the cone phosphodiesterase, a component of the phototransduction cascade in cone cells<sup>119</sup>. Loss of function mutations in *PDE6H* cause total color blindness in humans<sup>119</sup>, but not in mice, where the rod phosphodiesterase isoform *PDE6G* compensates for *PDE6H* loss<sup>120</sup>. Given this cross-species variability, we investigated additional genes required for phototransduction in rods and cones in the 26 focal bat species. For rods, we found no phototransduction component (*RHO*, *GNAT1*, *GNB1*, *GNGT1*, *PDE6A*, *PDE6B*, *PDE6G*) to be lost. Among the genes required for cone phototransduction (*OPN1SW*, *OPN1LW*, *GNAT2*, *GNB3*, *GNGT2*, *PDE6C*), we confirmed previously-reported losses of the short-wavelength sensitive opsin (*OPN1SW*) in the vampire bat and six other bats<sup>121-124</sup> (Supplementary Figure 6). In addition, our analysis detected that *PDE6C* is lost in the vampire and four other bats (Supplementary Figures 6, 7). Inactivating mutations in *PDE6C* in both humans and mice abolish cone function entirely, resulting in rod monochromacy, a condition where only rod photoreceptors remain functional<sup>125,126</sup>.

While ancestral Chiroptera were inferred to have two types of cones and dichromatic vision, many bats, including *D. rotundus*, lack cones that express a functional short-wavelength sensitive opsin (*OPN1SW*)<sup>121-124</sup>. Whereas the lack of a functional *OPN1SW* alone suggested color blindness but maintenance of functional cones expressing the conserved *OPN1LW* (long wave sensitive opsin) gene, our discovery of repeated *PDE6C* losses suggests that *D. rotundus* and four other nocturnal Noctilionoidea may be functional rod monochromats. This condition, where all cone-based photoreception is abolished even if vestigial cone cells can still be detected, has been suggested for only a few other mammalian lineages so far<sup>127</sup>. In light of their ecology, rod monochromacy is plausible in vampire bats, which lost both *PDE6C* and *PDE6H*. These strictly nocturnal bats are most active during the darkest periods of the night and even avoid moonshine<sup>128</sup>. Furthermore, to locate their prey, vampire bats have evolved non-visual sensory adaptations such as highly specialized auditory adaptations<sup>129</sup>, and an infrared sensing capability, which is unique among mammals<sup>6</sup>.

### Loss of the immune-related *RNASE7* and a different pathogen profile of blood

Feeding exclusively on blood, the immune system of vampire bats is regularly challenged with blood-borne pathogens<sup>130</sup>. However, because blood exhibits a low bacterial abundance but different species composition<sup>131,132</sup>, their intestinal tract is likely exposed to a different diversity of pathogens. We discovered that the immune-related *RNASE7* (ribonuclease A family member 7) gene is lost in *D. rotundus* and intact in all other analyzed bats.

*RNASE7* encodes a secreted ribonuclease that has potent antimicrobial activity against various microorganisms<sup>133</sup>. In humans, *RNASE7* is highly expressed in most epithelia including skin, urothelium and respiratory tract epithelium<sup>48,133</sup>. The gene is significantly downregulated in the skin and urinary tract of diabetic patients (potentially because *RNASE7* expression can be induced by insulin signaling), which reduces innate immune defense capacities and likely contributes to the much higher incidence of bacterial skin and urinary tract infections in diabetic patients<sup>134,135</sup>. The exclusive loss of *RNASE7* in vampire bats, which might also be related to their reduced insulin secretory capacities, predicts that vampire bats have a reduced capacity for secreting bactericidal peptides. While a previous study found that *RNASE7* evolved under positive selection in vampire bats, which was interpreted as an adaptation towards increased exposition to blood-borne pathogens<sup>21</sup>, our analysis shows that *RNASE7* is clearly inactivated, raises the possibility that its loss may be a consequence of exposure to a different pathogen diversity.

## Discussion

Here, we used long read (HiFi) sequencing to generate a haplotype-resolved genome assembly of high completeness, contiguity and base accuracy for the common vampire bat *Desmodus rotundus*. Adding to previous studies<sup>34,136</sup>, our side-by-side comparison of gene content and inactivating mutations demonstrates that the HiFi assembly not only improves gene completeness but also base accuracy. A haplotype-resolved assembly also facilitates discriminating between homozygous and heterozygous (inactivating) mutations, which is more difficult in collapsed diploid assemblies. Using this new *D. rotundus* assembly and existing genomes of 25 other bats, we performed a genome-wide screen for genes that are specifically lost in the vampire bat lineage. This screen revealed 10 previously-unknown gene losses that are likely associated with derived phenotypic features of vampire bats, such as reduced insulin secretion (*FFAR1*, *SLC30A8*) and glycogen synthesis (*PPP1R3E*), a distinct gastric physiology (*CTSE*), and exceptional social behavior and cognitive abilities (*CYP39A1*). The loss of other genes (*ERN2*, *CTRL*, *REP15*) is likely related to the biased nutrient composition of blood, which features low fat, high protein and high iron contents. Consistent with an association to vampire bat-specific changes in metabolism and digestion, many of these genes are highly expressed in relevant organ systems such as the pancreas (*FFAR1*, *SLC30A8*, *CTRL*) and gastrointestinal tissues (*REP15*, *CTSE*, *ERN2*). Finally, gene loss can also indicate previously-unknown phenotypes, exemplified here by the loss of *PDE6H* and *PDE6C*, which suggest the complete lack of cone photoreceptor function (functional rod monochromacy). Similarly, the loss of *RNASE7* may indicate differences in the immune system between *D. rotundus* and other bats, which deserves further studies. In particular, while bats are generally known for tolerating a larger diversity of viral and other pathogens, it is conceivable that their specialized diet exposes *D. rotundus* to a distinct pathogen profile, largely constrained to blood-borne pathogens. This hypothesis could be tested using metagenomics or related technologies to systematically characterize and compare pathogen profiles of vampire and other bats, ideally in matched environments.

Our finding that vampire bats, which naturally exhibit a reduced capacity for insulin secretion<sup>16</sup>, have lost *FFAR1*, a beta cell receptor that stimulates insulin secretion, provides an interesting contrast to fruit eating bats of the phyllostomid and pteropodid families, which exhibit an increased capacity for insulin secretion.

These fruit bats have convergently adapted to sugar-rich diets <sup>137,138</sup> and have convergently lost *FFAR3* <sup>23,26</sup>, a different beta cell receptor that inhibits insulin secretion <sup>139</sup>. This suggests that losses of two different genes with opposite effects on insulin secretion are involved in opposite phenotypes in bats.

Which genes are lost during the course of evolution is influenced by several factors. One important factor underlying gene loss, including those identified here, is the absence of selection to preserve the function of certain genes as a consequence of adaptation to a new environment or different lifestyle. However, despite dispensability of gene function, several other factors constrain which genes are permitted to be lost in evolution <sup>25</sup>. For example, genes lost in mammals are generally depleted in essential functions and tend to have a lower degree of pleiotropy <sup>92</sup>. Indeed, the few known examples of pleiotropic gene losses in mammals refer to cases where most or all functions of the gene became dispensable (e.g. *KLK8*, *INSL5*, *RXFP4*, *SLC4A9* <sup>28,29,140</sup>) or where additional gene functions are compensated by functionally related proteins (e.g. *ACOX2*, *SLC27A5* <sup>141</sup>). Consistent with pleiotropy being a key restricting factor, the gene loss cases detected here largely refer to genes with specific functions and a low or no degree of pleiotropy. In addition to pleiotropy and essentiality, restricted expression patterns appear to be another factor permitting gene loss. For example, while the inactivated gene *ERN2* exhibits a specific expression in the gastrointestinal tract, the ubiquitously expressed and functionally similar *ERN1* gene <sup>93</sup> is intact in the vampire bat genome. Similarly, *REP15*, *SLC30A8*, *CTSE*, and *CTRL* have tissue-restricted expression patterns.

While dispensability of gene function is certainly the main explanation for gene losses in the vampire bat lineage, loss of ancestral genes can sometimes be beneficial in evolution and contribute to adaptations <sup>25,26,28,29,142</sup>. For example, the loss of *REP15* could have contributed to enhanced iron elimination through shedding of iron-containing gastrointestinal cells, which likely represents an adaptation of vampire bats to their iron-rich blood diet. Another example is *CYP39A1*, whose loss is expected to result in elevated systemic levels of 24S-hydroxycholesterol. Since elevated levels of this cholesterol metabolite can positively affect cognitive function in animal models <sup>111,112</sup>, it is conceivable that loss of *CYP39A1* in the vampire bat lineage may have contributed to their exceptional social behavior and cognitive abilities, which is a unique feature among bats.

In summary, our study provides novel insights into the genomic changes related to adaptations to sanguivory and the evolutionary importance of gene losses in general. Notably, a comprehensive understanding of how sanguivory, as a unique dietary specialization, has evolved, requires not only high-quality genomic resources but also data on the organismal biology of vampire bats. Our study reveals gaps in our knowledge of vampire bat physiology, metabolism and immunity, highlighting the need to better characterize the phenotypic side of adaptations to their unique diet.

# Methods

## Sample collection for sequencing

A *D. rotundus* male individual (ROMM126221) was collected in a mist net at Sabajo, Para, Suriname on 26 September 2017 as part of a larger environmental baseline study and euthanized with an overdose of isoflurane. The research permit was approved on 8 June 2017 by the Nature Conservation Division, Suriname Forest Service. Animal Use Protocol #2017-19 was approved by the Animal Care Committee of the Royal Ontario Museum. The use of wild mammals in research followed the guidelines of the American Society of Mammalogists <sup>143</sup>.

## DNA extraction, library preparation and whole genome sequencing

Sampled tissues were snap-frozen in liquid nitrogen immediately after dissection and stored at -80 °C until further processed. Mixed tissue of heart, liver, and spleen (23 mg) was used for high molecular weight DNA extraction using the Nanobind Tissue Big DNA Kit (Circulomics, MD) according to the manufacturer's instructions. Final DNA purity and concentrations were measured using NanoPhotometer® (Implen GmbH, Munich, Germany) and Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA). Two SMRTbell libraries were constructed following the instructions of the SMRTbell Express Prep kit v2.0 (Pacific Biosciences, Menlo Park, CA). The total input DNA for each library was approximately 5 µg. The libraries were loaded at an on-plate concentration of 80 pM using diffusion loading. Four SMRT cell sequencing runs were performed on the Sequel System IIe in CCS mode using 30 hour movie time with 2 hours pre-extension and sequencing chemistry V2.0. The Dovetail Omni-C library was prepared from mixed tissue (heart, liver, and spleen) using the Dovetail™ Omni-C™ kit (Dovetail Genomics, Scotts Valley, CA, USA) following the manufacturer's protocol (manual version 1.2 for mammalian samples). The Omni-C library was sequenced on a NovaSeq 6000 platform at Novogene (UK), generating 400 million 2 × 150 bp paired-end reads totaling 120 Gbp. The fragment size distribution and concentration of the final PacBio and Dovetail Omni-C library were assessed using a TapeStation (Agilent Technologies) and a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA), respectively.

## *D. rotundus* genome assembly

We called CCS reads (rq > 0.99) from the subreads.bam file using PacBio ccs (version 6.0.0, <https://github.com/PacificBiosciences/ccs>). We then created the two contig assemblies using hifiasm (version 0.15.1-r331) <sup>144</sup> with arguments -l2 --h1 --h2 using the CCS reads and the HiC reads as input. This resulted in one contig assembly for each of the haplotypes in this species, which both went independently into the following scaffolding, gap-closing and polishing steps.

For scaffolding, we used SALSA2 (git commit 1b76bf63efb973583647a1eb95863d33ee6e09ad) <sup>145</sup> and the Dovetail Omni-C mapping pipeline ([https://omni-c.readthedocs.io/en/latest/fastq\\_to\\_bam.html](https://omni-c.readthedocs.io/en/latest/fastq_to_bam.html)). Briefly, we mapped HiC reads using BWA-MEM (version 0.7.17-r1188) <sup>146</sup>. Alignments were filtered using pairtools (version 0.3.0) with arguments parse --min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30 to retain those alignments with high mapping quality, the most 5' alignments for conflicts and alignments without large gaps. We removed potential PCR duplicates using pairtools dedup. The resulting read-sorted bed file was used as input for SALSA2. We then performed a number of manual curation rounds to correct scaffolding errors and to scaffold those contigs which were not automatically scaffolded. To this end, we

used cooler (version 0.8.11, <https://github.com/open2c/cooler>) and HiGlass <sup>147</sup> to visually inspect the HiC maps and re-scaffolded the assembly using SeqKit (version 0.13.2) to re-arrange contigs and scaffolds into chromosome-level scaffolds.

After scaffolding, we closed additional assembly gaps using the lower-quality PacBio CLR read data that did not yield a CCS read. To this end, we mapped the original subreads.bam files to the scaffolded assembly using pbmm2 (version 1.3.0) with arguments --preset SUBREAD -N 1. Based on the read-piles created by reads spanning across gap regions, we computed consensus sequences for the gap regions and their 2 kb up/downstream flanks using gcpp (version 2.0.2-2.0.2). We filtered for high-confidence consensus sequences by requiring that no N's and no lower-case (acgt) characters remain in the consensus sequence. We then replaced the assembly gap and flanking region with high-confidence consensus sequences.

Despite the high accuracy of PacBio CCS reads, an assembly can still contain base errors due to errors in the consensus sequence calculation. In addition, lower base accuracy likely exists in the closed gap regions. To improve base accuracy, we polished both haplotype assemblies using all CCS reads. We tested both freebayes (version 1.3.2, <https://github.com/freebayes/freebayes>) and DeepVariant (version 1.1.0) <sup>148</sup>. Based on numbers of genes with inactivating mutations (computed by TOGA, see below) and the QV values computed by Merqury (version 1.0) <sup>32</sup>, DeepVariant vastly outperformed freebayes in correcting base errors. Therefore, for the final polishing round, we mapped all CCS reads to the scaffolded, gap-closed assemblies using pbmm2 with arguments: --preset CCS -N 1 and called variants using DeepVariant. We then filtered for sites with 'genotype 1/1' to specify that all or nearly all reads support an alternative sequence at this position and a 'PASS' filter value to specify that the site passed DeepVariant's internal filters. We then corrected base errors using bcftools consensus (version 1.12). Importantly, this procedure does not 'correct' any heterozygous or polymorphic regions but only those that are incorrect and not supported by any CCS reads.

### Repeat masking

We used RepeatModeler (<http://www.repeatmasker.org/>, parameter -engine ncbi) to generate a *de novo* repeat library for the new *D. rotundus* genome assembly and used the resulting library with RepeatMasker (version 4.0.9, parameters -engine crossmatch -s) to soft-mask the genome.

### Genome alignments

We used the human hg38 genome assembly as the reference and considered *D. rotundus* and 25 other bats as query species (Supplementary Table 1). To generate pairwise genome alignment chains as input for TOGA <sup>33</sup>, we first used LASTZ (version 1.04.03) <sup>149</sup> with parameters (K = 2400, L = 3000, Y = 9400, H = 2000 and the lastz default scoring matrix). These parameters are sensitive enough to capture orthologous exons between placental mammals <sup>150</sup>. Then, we used axChain <sup>151</sup> (default parameters except linearGap=loose) to compute co-linear alignment chains, RepeatFiller <sup>152</sup> (default parameters) to capture previously-missed alignments between repetitive regions and chainCleaner <sup>153</sup> (default parameters except minBrokenChainScore=75000 and -doPairs) to improve alignment specificity.



## Using TOGA to infer orthologous genes and detect gene losses

To compare gene completeness and to screen for gene losses, we used TOGA<sup>33</sup>, a method that uses pairwise genome alignment chains between an annotated reference genome (here human hg38 assembly) and other query species. Briefly, TOGA uses machine learning to infer orthologous loci for each reference transcript, utilizing that orthologous genes display more alignments between intronic and flanking intergenic regions<sup>33</sup>. TOGA then projects each reference transcript to its orthologous query locus using CESAR 2.0<sup>154</sup>, a Hidden Markov model method that takes reading frame and splice site annotation of the reference exons into account. CESAR avoids spurious frameshifts, is able to detect evolutionary splice site shifts and precise intron deletions<sup>154,155</sup>. Using the CESAR alignment, TOGA determines whether the transcript has inactivating mutations (frameshifting mutations, premature stop codons, splice site disrupting mutations, deletions of entire coding exons). TOGA also uses orthologous alignment chains to detect genes that are entirely deleted in the query assembly and distinguishes real exon or gene deletions from missing sequence caused by assembly gaps.

TOGA classifies each transcript into one of five categories. Classification 'intact', 'partially intact' and 'missing' refers to transcripts that lack inactivating mutations in the middle 80% of the coding region. TOGA specifically considers this central part of the coding region, since truly conserved genes can have inactivating mutations in the first or last 10% of the coding region (near the N- or C-terminus)<sup>33,155</sup>. For intact transcripts, the central part of the coding region is completely present in the assembly, whereas for partially intact ( $\geq 50\%$  present) and missing ( $< 50\%$  present) transcripts some coding parts are missing due to assembly gaps or fragmentation. Classification 'lost' refers to transcripts that have at least two inactivating mutations in at least two exons. All other transcripts, having a single inactivating mutation or mutations in only a single exon, are classified as 'uncertain loss', reflecting the possibility that only an exon but not the entire gene could be lost.

Considering all input transcripts of a gene, TOGA uses the precedence order 'intact', 'partially intact', 'uncertain loss', 'lost', 'missing' to provide a gene classification. That means, TOGA only classifies a gene as lost if all its transcripts are classified as lost (at all orthologous loci if there is more than one).

## Gene completeness among bat genomes

To compare gene completeness between assemblies, we first obtained a set of genes that likely already existed in the placental mammal ancestor. To this end, we used the human GENCODE V38 (Ensembl 104) gene annotation<sup>156</sup> as input and applied TOGA to the genomes of 11 afrotherian (represented by aardvark, cape golden mole, small Madagascar hedgehog, Talazac's shrew tenrec, cape elephant shrew, dugong, manatee, Asiatic elephant, African savanna elephant, cape rock hyrax, yellow-spotted hyrax) and five xenarthran (represented by southern two-toed sloth, Hoffmann's two-fingered sloth, nine-banded armadillo, giant anteater, southern tamandua) mammals. Since each input gene is by definition present in human (superorder Boreoeutheria), we considered a gene as ancestral if it is also conserved in the other two basal placental superorders Afrotheria and Xenarthra. Specifically, we selected genes that are classified by TOGA as intact or partially intact in at least one afrotherian and at least one xenarthran genome. Note that this requirement assures placental mammal ancestry irrespective of the exact basal split of placental mammals, which is difficult to resolve<sup>157</sup>. This procedure resulted in a set of 18,430 genes.

Next, we applied TOGA to *D. rotundus* and the 25 other bat genomes (Supplementary Table 1). Considering the 18,430 ancestral genes, we counted per species how many gene have an intact reading frame (TOGA classification 'intact'), have inactivating mutations (TOGA classifications 'loss' and 'uncertain loss') or have missing sequence (TOGA classifications 'partially intact' and 'missing'). This breakdown is shown in Figure 1C.

### Screen for gene losses

To identify gene losses that among bats occurred specifically in the vampire bat lineage, we used the above-generated TOGA data for the 26 bats. We first filtered for genes that have at least one inactivating mutation in the middle 80% of the coding sequence (TOGA classification 'lost' or 'uncertain loss') in both *D. rotundus* haplotype assemblies.

Base errors can also occur in other bat assemblies, which may result in missing real vampire bat-specific gene losses (false negatives). To reduce the false negative rate, we conducted the initial screen with less strict requirements and allowed per gene up to three non-vampire bat species to be classified as 'missing', 'partially missing' or 'uncertain loss', resulting in 52 candidate genes. Then, to assess which of these candidate genes are a true vampire specific gene losses, we mapped available Illumina reads of the same species to each included genome assembly using BWA-MEM (version 0.7.7-r441) <sup>146</sup> and removed duplicates with Picard (version 2.21.4) (<http://broadinstitute.github.io/picard>) (SRA identifiers are listed in Supplementary Table 4) and examined whether putative inactivating mutations are supported by raw reads and if they are heterozygous or homozygous. We excluded all genes as non-vampire specific losses that have homozygous mutations in at least one other bat species and furthermore required that not more than 60% of the coding sequence remains functional in *D. rotundus*. We kept *SLC30A8*, which has a heterozygous stop codon in exon 4 in *Micronycteris hirsuta*. We excluded genes belonging to the large and fast-evolving olfactory receptor family. This procedure revealed the 13 vampire specific gene losses that are discussed in the main text.

Finally, to confirm the correctness of all 10 previously-unknown vampire bat-specific gene losses, we manually inspected the pairwise alignment chains in the UCSC genome browser <sup>158</sup>, which showed that the remnants of all 10 genes are located in a conserved gene order (synteny) context. Inspecting chiropteran pairwise alignment chains also showed that none of the 10 genes exhibit duplications in *D. rotundus* or other bats, which excludes the possibility that a functional gene copy remained in *D. rotundus*. Finally, we further verified the correctness of the inactivating mutations by requiring that the gene is also classified as 'lost' or 'uncertain loss' in the Illumina *D. rotundus* assembly with at least one shared homozygous (based on aligned Illumina sequencing reads) inactivating mutation, thus providing base support from two sequencing technologies.

### Relaxed selection

To test the 10 genes for relaxation of selection, we obtained pairwise codon alignments between human and each query bat from TOGA. Codons affected by frameshifting insertions or deletions and premature stop codons were replaced by 'NNN' codons to maintain a reading frame. A multiple codon alignment was produced with MACSE v2 <sup>159</sup> and cleaned with HmmsCleaner <sup>160</sup> with default cost values. The resulting

alignments were used to investigate whether the gene evolves under relaxed selection with RELAX<sup>45</sup>, specifying the *D. rotundus* branch as the foreground and all other branches as background.

## Gene expression

We investigated RNA expression of the 10 target gene losses using available transcriptome sequencing data for *D. rotundus* from 11 tissues (heart, stomach, intestine, liver, gallbladder, pancreas, spleen, kidney, olfactory epithelium, eye and tongue). The paired-end sequencing read sets were downloaded from NCBI (accessions listed in Supplementary Table 5) and mapped to the new *D. rotundus* haplotype 1 assembly using STAR (version 2.7.3)<sup>161</sup>. Target loci were then manually examined in IGV<sup>162</sup>.

## Whole blood iron measurements

All animal captures were approved by the National Environmental Office (License number 77322-1, Sisbio, Brazil) and all experiments were approved by the Ethics Committee from Federal University of Viçosa (License number 10/2021, CEUA/UFV, Brazil). Nine *D. rotundus*, sixteen *Artibeus lituratus* and four *Myotis myotis* individuals were captured with mist nets in Viçosa, Minas Gerais, Brazil and euthanized through cervical dislocation followed by decapitation. All blood samples (100 - 400 µl) were dried using a stove and mineralized in a nitric-perchloric acid solution (1.5 mL total volume) until organic matter was removed. The final extract was used to determine iron concentrations through atomic absorption spectrophotometry (Shimadzu AA-6701F). Whole blood iron levels were compared between species with a two-sided t-test.

## Data availability

The haplotype-resolved *D. rotundus* assembly and the TOGA annotations are freely available at <https://bds.mpi-cbg.de/hillerlab/VampireBatGenome/>. The sequencing data and genome assembly is also uploaded to NCBI.

## Competing interests

The authors have no competing interests.

## Acknowledgments

We would like to thank the UCSC genome browser group for providing software and genome annotations, and the Computer Service Facilities of the MPI-CBG and MPI-PKS for their support. We also thank the Genome Technology Center (RGTC) at Radboudumc for the use of the Sequencing Core Facility (Nijmegen, The Netherlands). Furthermore, we want to thank Robert Cornman for support concerning the *Aeorestes cinereus* assembly, and David Lagman and Dan Larhammar for helpful feedback on the phototransduction cascade. Figures 3C, 4 and S3A created with Biorender (<https://biorender.com>). This work was supported by the Max Planck Society, the National Council for Scientific and Technological Development (CNPq, Brazil), funding from Environmental Services & Support (ESS) for fieldwork in Suriname, and the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK).

## References

- 1 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* Cambridge Studies in Morphology and Molecules: New Paradigms in Evolutionary Bio (eds Gregg F. Gunnell & Nancy B. Simmons) 385-409 (Cambridge University Press, 2012).
- 2 Wilkinson, G. S. Vampire bats. *Current biology : CB* **29**, R1216-R1217, doi:10.1016/j.cub.2019.10.052 (2019).
- 3 Schmidt, U. *Natural history of vampire bats*. (CRC Press, 2018).
- 4 Schmidt, U. Olfactory threshold and odour discrimination of the vampire bat. (*Desmodus rotundus*). *Period Biol* **75**, 89-92 (1973).
- 5 Heffner, R. S., Koay, G. & Heffner, H. E. Hearing in American leaf-nosed bats. IV: the Common vampire bat, *Desmodus rotundus*. *Hear Res* **296**, 42-50, doi:10.1016/j.heares.2012.09.011 (2013).
- 6 Gracheva, E. O. *et al.* Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* **476**, 88-91, doi:10.1038/nature10245 (2011).
- 7 Riskin, D. K., Parsons, S., Schutt, W. A., Jr., Carter, G. G. & Hermanson, J. W. Terrestrial locomotion of the New Zealand short-tailed bat *Mystacina tuberculata* and the common vampire bat *Desmodus rotundus*. *The Journal of experimental biology* **209**, 1725-1736, doi:10.1242/jeb.02186 (2006).
- 8 Davis, J. S., Nicolay, C. W. & Williams, S. H. A comparative study of incisor procumbency and mandibular morphology in vampire bats. *J Morphol* **271**, 853-862, doi:10.1002/jmor.10840 (2010).
- 9 Low, D. H. *et al.* Dracula's children: molecular evolution of vampire bat venom. *J Proteomics* **89**, 95-111, doi:10.1016/j.jprot.2013.05.034 (2013).
- 10 McNab, B. K. Energetics and the Distribution of Vampires. *Journal of Mammalogy* **54**, 131-144, doi:10.2307/1378876 (1973).
- 11 Breidenstein, C. P. Digestion and Assimilation of Bovine Blood by a Vampire Bat (*Desmodus rotundus*). *Journal of Mammalogy* **63**, 482-484, doi:10.2307/1380446 (1982).
- 12 Rouk, C. S. & Glass, B. P. Comparative gastric histology of five North and Central American bats. *J Mammal* **51**, 455-490 (1970).
- 13 Morton, D. & Wimsatt, W. A. Distribution of iron in the gastrointestinal tract of the common vampire bat: evidence for macrophage-linked iron clearance. *The Anatomical record* **198**, 183-192, doi:10.1002/ar.1091980206 (1980).
- 14 Graca-Souza, A. V. *et al.* Adaptations against heme toxicity in blood-feeding arthropods. *Insect Biochem Mol Biol* **36**, 322-335, doi:10.1016/j.ibmb.2006.01.009 (2006).
- 15 Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A Red Carpet for Iron Metabolism. *Cell* **168**, 344-361, doi:10.1016/j.cell.2016.12.034 (2017).
- 16 Freitas, M. B. *et al.* Reduced insulin secretion and glucose intolerance are involved in the fasting susceptibility of common vampire bats. *General and comparative endocrinology* **183**, 1-6, doi:10.1016/j.ygcen.2012.11.023 (2013).
- 17 Freitas, M. B., Welker, A. F., Millan, S. F. & Pinheiro, E. C. Metabolic responses induced by fasting in the common vampire bat *Desmodus rotundus*. *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology* **173**, 703-707, doi:10.1007/s00360-003-0383-3 (2003).
- 18 Freitas, M. B., Passos, C. B., Vasconcelos, R. B. & Pinheiro, E. C. Effects of short-term fasting on energy reserves of vampire bats (*Desmodus rotundus*). *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* **140**, 59-62, doi:10.1016/j.cbpc.2004.09.023 (2005).
- 19 Wilkinson, G. S. Reciprocal food sharing in the vampire bat. *Nature* **308**, 181-184, doi:10.1038/308181a0 (1984).
- 20 Carter, G. G. & Wilkinson, G. S. Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment. *Proceedings. Biological sciences / The Royal Society* **280**, 20122573, doi:10.1098/rspb.2012.2573 (2013).
- 21 Zepeda Mendoza, M. L. *et al.* Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat Ecol Evol* **2**, 659-668, doi:10.1038/s41559-018-0476-8 (2018).
- 22 Gutierrez-Guerrero, Y. T. *et al.* Genomic consequences of dietary diversification and parallel evolution due to nectarivory in leaf-nosed bats. *Gigascience* **9**, doi:10.1093/gigascience/giaa059 (2020).
- 23 Wang, K. *et al.* Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol Ecol*, doi:10.1111/mec.15542 (2020).
- 24 Hong, W. & Zhao, H. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. *Proceedings. Biological sciences / The Royal Society* **281**, 20141079, doi:10.1098/rspb.2014.1079 (2014).
- 25 Albalat, R. & Canestro, C. Evolution by gene loss. *Nature reviews. Genetics* **17**, 379-391, doi:10.1038/nrg.2016.39 (2016).
- 26 Sharma, V. *et al.* A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature communications* **9**, 1215, doi:10.1038/s41467-018-03667-1 (2018).
- 27 Sharma, V., Lehmann, T., Stuckas, H., Funke, L. & Hiller, M. Loss of RXFP2 and INSL3 genes in Afrotheria shows that testicular descent is the ancestral condition in placental mammals. *PLoS Biol* **16**, e2005293, doi:10.1371/journal.pbio.2005293 (2018).
- 28 Hecker, N., Sharma, V. & Hiller, M. Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 3036-3041, doi:10.1073/pnas.1818504116 (2019).

- 29 Huelsmann, M. *et al.* Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Sci Adv* **5**, eaaw6671, doi:10.1126/sciadv.aaw6671 (2019).
- 30 Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* **37**, 1155-1162, doi:10.1038/s41587-019-0217-9 (2019).
- 31 Santos, N., Fagundes, V., Yonenaga-Yassuda, Y. & De Souza, M. J. Comparative karyology of Brazilian vampire bats *Desmodus rotundus* and *Diphylla ecaudata* (Phyllostomidae, Chiroptera): banding patterns, base-specific fluorochromes and FISH of ribosomal genes. *Hereditas* **134**, 189-194, doi:10.1111/j.1601-5223.2001.00189.x (2001).
- 32 Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245, doi:10.1186/s13059-020-02134-9 (2020).
- 33 Kirilenko, B. M. *et al.* TOGA: a machine-learning approach to integrate gene annotation with orthology inference at scale. (in prep.).
- 34 Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578-584, doi:10.1038/s41586-020-2486-3 (2020).
- 35 Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-482, doi:10.1038/nature10530 (2011).
- 36 Seim, I. *et al.* Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nature communications* **4**, 2212, doi:10.1038/ncomms3212 (2013).
- 37 Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456-460, doi:10.1126/science.1230835 (2013).
- 38 Eckalbar, W. L. *et al.* Transcriptomic and epigenomic characterization of the developing bat wing. *Nat Genet* **48**, 528-536, doi:10.1038/ng.3537 (2016).
- 39 Dong, D. *et al.* The Genomes of Two Bat Species with Long Constant Frequency Echolocation Calls. *Molecular biology and evolution* **34**, 20-34, doi:10.1093/molbev/msw231 (2017).
- 40 Pinzari, C. A. *et al.* Analysis of Genomic Sequence Data Reveals the Origin and Evolutionary Separation of Hawaiian Hoary Bat Populations. *Genome Biol Evol* **12**, 1504-1514, doi:10.1093/gbe/evaa137 (2020).
- 41 Zoonomia, C. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240-245, doi:10.1038/s41586-020-2876-6 (2020).
- 42 Datzmann, T., von Helversen, O. & Mayer, F. Evolution of nectarivory in phyllostomid bats (Phyllostomidae Gray, 1825, Chiroptera: Mammalia). *BMC Evol Biol* **10**, 165, doi:10.1186/1471-2148-10-165 (2010).
- 43 Rojas, D., Warsi, O. M. & Davalos, L. M. Bats (Chiroptera: Noctilionoidea) Challenge a Recent Origin of Extant Neotropical Diversity. *Systematic biology* **65**, 432-448, doi:10.1093/sysbio/syw011 (2016).
- 44 Zhao, H. *et al.* Evolution of the sweet taste receptor gene *Tas1r2* in bats. *Molecular biology and evolution* **27**, 2642-2650, doi:10.1093/molbev/msq152 (2010).
- 45 Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular biology and evolution* **32**, 820-832, doi:10.1093/molbev/msu400 (2015).
- 46 Strick, D. J. & Elferink, L. A. Rab15 effector protein: a novel protein for receptor recycling from the endocytic recycling compartment. *Mol Biol Cell* **16**, 5699-5709, doi:10.1091/mbc.e05-03-0204 (2005).
- 47 Stasiak, I. M. *et al.* Iron storage disease (hemochromatosis) and hepcidin response to iron load in two species of pteropodid fruit bats relative to the common vampire bat. *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology* **188**, 683-694, doi:10.1007/s00360-018-1155-4 (2018).
- 48 Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-1250, doi:10.1038/nbt1210-1248 (2010).
- 49 Elferink, L. A. & Strick, D. J. Functional properties of rab15 effector protein in endocytic recycling. *Methods Enzymol* **403**, 732-743, doi:10.1016/S0076-6879(05)03064-8 (2005).
- 50 Chen, E. *et al.* Alteration of tumor suppressor BMP5 in sporadic colorectal cancer: a genomic and transcriptomic profiling based study. *Mol Cancer* **17**, 176, doi:10.1186/s12943-018-0925-7 (2018).
- 51 Brookes, M. J. *et al.* Modulation of iron transport proteins in human colorectal carcinogenesis. *Gut* **55**, 1449-1460, doi:10.1136/gut.2006.094060 (2006).
- 52 Williams, J. M. *et al.* Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet Pathol* **52**, 445-455, doi:10.1177/0300985814559404 (2015).
- 53 Mercadante, C. J. *et al.* Gastrointestinal iron excretion and reversal of iron excess in a mouse model of inherited iron excess. *Haematologica* **104**, 678-689, doi:10.3324/haematol.2018.198382 (2019).
- 54 Neely, B. A. *et al.* Surveying the Vampire Bat (*Desmodus rotundus*) Serum Proteome: A Resource for Identifying Immunological Proteins and Detecting Pathogens. *J Proteome Res* **20**, 2547-2559, doi:10.1021/acs.jproteome.0c00995 (2021).
- 55 Mancini, A. D. & Poitout, V. The fatty acid receptor FFA1/GPR40 a decade later: how much do we know? *Trends Endocrinol Metab* **24**, 398-407, doi:10.1016/j.tem.2013.03.003 (2013).
- 56 Latour, M. G. *et al.* GPR40 is necessary but not sufficient for fatty acid stimulation of insulin secretion in vivo. *Diabetes* **56**, 1087-1094, doi:10.2337/db06-1532 (2007).
- 57 Pedersen, M. G., Tagliavini, A. & Henquin, J. C. Calcium signaling and secretory granule pool dynamics underlie biphasic insulin secretion and its amplification by glucose: experiments and modeling. *Am J Physiol Endocrinol Metab* **316**, E475-E486, doi:10.1152/ajpendo.00380.2018 (2019).
- 58 Lan, H. *et al.* Lack of FFAR1/GPR40 does not protect mice from high-fat diet-induced metabolic disease. *Diabetes* **57**, 2999-3006, doi:10.2337/db08-0596 (2008).



- 59 Tomita, T. *et al.* Expression of the gene for a membrane-bound fatty acid receptor in the pancreas and islet cell tumours in humans: evidence for GPR40 expression in pancreatic beta cells and implications for insulin secretion. *Diabetologia* **49**, 962-968, doi:10.1007/s00125-006-0193-8 (2006).
- 60 Del Guerra, S. *et al.* G-protein-coupled receptor 40 (GPR40) expression and its regulation in human pancreatic islets: the role of type 2 diabetes and fatty acids. *Nutr Metab Cardiovasc Dis* **20**, 22-25, doi:10.1016/j.numecd.2009.02.008 (2010).
- 61 Edfalk, S., Steneberg, P. & Edlund, H. Gpr40 is expressed in enteroendocrine cells and mediates free fatty acid stimulation of incretin secretion. *Diabetes* **57**, 2280-2287, doi:10.2337/db08-0307 (2008).
- 62 Yamashima, T. Dual effects of the non-esterified fatty acid receptor 'GPR40' for human health. *Prog Lipid Res* **58**, 40-50, doi:10.1016/j.plipres.2015.01.002 (2015).
- 63 Nicolson, T. J. *et al.* Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* **58**, 2070-2083, doi:10.2337/db09-0551 (2009).
- 64 Pound, L. D. *et al.* The physiological effects of deleting the mouse SLC30A8 gene encoding zinc transporter-8 are influenced by gender and genetic background. *PLoS one* **7**, e40972, doi:10.1371/journal.pone.0040972 (2012).
- 65 Chimienti, F., Devergnas, S., Favier, A. & Seve, M. Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* **53**, 2330-2337, doi:10.2337/diabetes.53.9.2330 (2004).
- 66 Dunn, M. F. Zinc-ligand interactions modulate assembly and stability of the insulin hexamer -- a review. *Biometals* **18**, 295-303, doi:10.1007/s10534-005-3685-y (2005).
- 67 Burguera, M., Burguera, J. L. & Alarcón, O. M. Flow injection and microwave-oven sample decomposition for determination of copper, zinc and iron in whole blood by atomic absorption spectrometry. *Analytica Chimica Acta* **179**, 351-357, doi:[https://doi.org/10.1016/S0003-2670\(00\)84479-7](https://doi.org/10.1016/S0003-2670(00)84479-7) (1986).
- 68 EFSA Panel on Dietetic Products, N. & Allergies. Scientific Opinion on Dietary Reference Values for zinc. *EFSA Journal* **12**, 3844, doi:<https://doi.org/10.2903/j.efsa.2014.3844> (2014).
- 69 Peres, J. M. *et al.* Improvement of zinc intestinal absorption and reduction of zinc/iron interaction using metal bound to the caseinophosphopeptide 1-25 of beta-casein. *Reprod Nutr Dev* **38**, 465-472, doi:10.1051/rnd:19980410 (1998).
- 70 Kondaiah, P., Sharp, P. A. & Pullakhandam, R. Zinc induces iron egress from intestinal Caco-2 cells via induction of Hephastin: A role for PI3K in intestinal iron absorption. *Biochemical and biophysical research communications* **523**, 987-992, doi:10.1016/j.bbrc.2020.01.023 (2020).
- 71 Pound, L. D. *et al.* Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. *The Biochemical journal* **421**, 371-376, doi:10.1042/BJ20090530 (2009).
- 72 Bosma, K. J. *et al.* Evidence that Evolution of the Diabetes Susceptibility Gene SLC30A8 that Encodes the Zinc Transporter ZnT8 Drives Variations in Pancreatic Islet Zinc Content in Multiple Species. *J Mol Evol* **87**, 147-151, doi:10.1007/s00239-019-09898-0 (2019).
- 73 Havu, N., Lundgren, G. & Falkmer, S. Zinc and manganese contents of micro-dissected pancreatic islets of some rodents. A microchemical study in adult and newborn guinea pigs, rats, Chinese hamsters and spiny mice. *Acta endocrinologica* **86**, 570-577 (1977).
- 74 Galabova, R., Gospodinov, C., Ogneva, V. & Petkov, P. [Histochemistry of the islands of Langerhans in the pancreas of the sheep (ovis aries)]. *Ann Histochem* **16**, 91-99 (1971).
- 75 Petkov, P. E., Gospodinov, C. & Galabova, R. Histochemistry of Langerhans' islets in the pancreas of cattle (Bos Taurus L.). *Histochemie* **23**, 127-137, doi:10.1007/BF00305846 (1970).
- 76 Zalewski, P. D. *et al.* Video image analysis of labile zinc in viable pancreatic islet cells using a specific fluorescent probe for zinc. *J Histochem Cytochem* **42**, 877-884, doi:10.1177/42.7.8014471 (1994).
- 77 Hunt, J. R. Bioavailability of iron, zinc, and other trace minerals from vegetarian diets. *Am J Clin Nutr* **78**, 633S-639S, doi:10.1093/ajcn/78.3.633S (2003).
- 78 Semrad, C. E. Zinc and intestinal function. *Curr Gastroenterol Rep* **1**, 398-403, doi:10.1007/s11894-999-0021-7 (1999).
- 79 Munro, S., Ceulemans, H., Bollen, M., Diplecxito, J. & Cohen, P. T. A novel glycogen-targeting subunit of protein phosphatase 1 that is regulated by insulin and shows differential tissue distribution in humans and rodents. *FEBS J* **272**, 1478-1489, doi:10.1111/j.1742-4658.2005.04585.x (2005).
- 80 Roach, P. J. Glycogen and its metabolism. *Curr Mol Med* **2**, 101-120, doi:10.2174/1566524024605761 (2002).
- 81 Petersen, M. C., Vatner, D. F. & Shulman, G. I. Regulation of hepatic glucose metabolism in health and disease. *Nat Rev Endocrinol* **13**, 572-587, doi:10.1038/nrendo.2017.80 (2017).
- 82 Aschenbach, W. G. *et al.* The muscle-specific protein phosphatase PP1G/R(GL)(G(M))is essential for activation of glycogen synthase by exercise. *J Biol Chem* **276**, 39959-39967, doi:10.1074/jbc.M105518200 (2001).
- 83 Mehta, M. B. *et al.* Hepatic protein phosphatase 1 regulatory subunit 3B (Ppp1r3b) promotes hepatic glycogen synthesis and thereby regulates fasting energy homeostasis. *J Biol Chem* **292**, 10444-10454, doi:10.1074/jbc.M116.766329 (2017).
- 84 Berman, H. K., O'Doherty, R. M., Anderson, P. & Newgard, C. B. Overexpression of protein targeting to glycogen (PTG) in rat hepatocytes causes profound activation of glycogen synthesis independent of normal hormone- and substrate-mediated regulatory mechanisms. *J Biol Chem* **273**, 26421-26425, doi:10.1074/jbc.273.41.26421 (1998).
- 85 Crosson, S. M., Khan, A., Printen, J., Pessin, J. E. & Saltiel, A. R. PTG gene deletion causes impaired glycogen synthesis and developmental insulin resistance. *The Journal of clinical investigation* **111**, 1423-1432, doi:10.1172/JCI17975 (2003).

86 Zhang, Y. *et al.* Ablation of PPP1R3G reduces glycogen deposition and mitigates high-fat diet induced obesity. *Mol Cell Endocrinol* **439**, 133-140, doi:10.1016/j.mce.2016.10.036 (2017).

87 Park, H. & Hall, E. R. The Gross Anatomy of the Tongues and Stomachs of Eight New World Bats. *Transactions of the Kansas Academy of Science (1903-)* **54**, 64-72, doi:10.2307/3626255 (1951).

88 Wimsatt, W. A. & Guerriere, A. Observations on the Feeding Capacities and Excretory Functions Of Captive Vampire Bats1. *Journal of Mammalogy* **43**, 17-27, doi:10.2307/1376876 (1962).

89 Zaidi, N., Hermann, C., Herrmann, T. & Kalbacher, H. Emerging functional roles of cathepsin E. *Biochemical and biophysical research communications* **377**, 327-330, doi:10.1016/j.bbrc.2008.10.034 (2008).

90 Saku, T., Sakai, H., Shibata, Y., Kato, Y. & Yamamoto, K. An immunocytochemical study on distinct intracellular localization of cathepsin E and cathepsin D in human gastric cells and various rat cells. *J Biochem* **110**, 956-964, doi:10.1093/oxfordjournals.jbchem.a123696 (1991).

91 Ordonez, G. R. *et al.* Loss of genes implicated in gastric function during platypus evolution. *Genome Biol* **9**, R81, doi:10.1186/gb-2008-9-5-r81 (2008).

92 Sharma, V. & Hiller, M. Losses of human disease-associated genes in placental mammals. *NAR Genom Bioinform* **2**, lqz012, doi:10.1093/nargab/lqz012 (2020).

93 Bertolotti, A. *et al.* Increased sensitivity to dextran sodium sulfate colitis in IRE1beta-deficient mice. *The Journal of clinical investigation* **107**, 585-593, doi:10.1172/JCI11476 (2001).

94 Iqbal, J. *et al.* IRE1beta inhibits chylomicron production by selectively degrading MTP mRNA. *Cell Metab* **7**, 445-455, doi:10.1016/j.cmet.2008.03.005 (2008).

95 Iqbal, J. *et al.* Increased intestinal lipid absorption caused by Ire1beta deficiency contributes to hyperlipidemia and atherosclerosis in apolipoprotein E-deficient mice. *Circulation research* **110**, 1575-1584, doi:10.1161/CIRCRESAHA.112.264283 (2012).

96 Whitcomb, D. C. & Lowe, M. E. Human pancreatic digestive enzymes. *Dig Dis Sci* **52**, 1-17, doi:10.1007/s10620-006-9589-z (2007).

97 Sun, D., Jin, H., Zhang, J. & Tan, X. Integrated whole genome microarray analysis and immunohistochemical assay identifies COL11A1, GJB2 and CTRL as predictive biomarkers for pancreatic cancer. *Cancer Cell Int* **18**, 174, doi:10.1186/s12935-018-0669-x (2018).

98 Mosztbacher, D., Jancso, Z. & Sahin-Toth, M. Loss of chymotrypsin-like protease (CTRL) alters intrapancreatic protease activation but not pancreatitis severity in mice. *Scientific reports* **10**, 11731, doi:10.1038/s41598-020-68616-9 (2020).

99 Hostelley, T. L. *et al.* Exocrine pancreas proteases regulate beta-cell proliferation in zebrafish ciliopathy models and in murine systems. *Biol Open* **10**, doi:10.1242/bio.046839 (2021).

100 Carter, G. G. & Wilkinson, G. S. Common vampire bat contact calls attract past food-sharing partners. *Animal Behaviour* **116**, 45-51, doi:10.1016/j.anbehav.2016.03.005 (2016).

101 Razik, I., Brown, B. K. G., Page, R. A. & Carter, G. G. Non-kin adoption in the common vampire bat. *R Soc Open Sci* **8**, 201927, doi:10.1098/rsos.201927 (2021).

102 Li-Hawkins, J., Lund, E. G., Bronson, A. D. & Russell, D. W. Expression cloning of an oxysterol 7alpha-hydroxylase selective for 24-hydroxycholesterol. *J Biol Chem* **275**, 16543-16549, doi:10.1074/jbc.M001810200 (2000).

103 Grabovec, I. P. *et al.* Ligand-binding properties and catalytic activity of the purified human 24-hydroxycholesterol 7alpha-hydroxylase, CYP39A1. *J Steroid Biochem Mol Biol* **193**, 105416, doi:10.1016/j.jsbmb.2019.105416 (2019).

104 Norlin, M., Toll, A., Bjorkhem, I. & Wikvall, K. 24-hydroxycholesterol is a substrate for hepatic cholesterol 7alpha-hydroxylase (CYP7A). *J Lipid Res* **41**, 1629-1639 (2000).

105 Yantsevich, A. V. *et al.* Human steroid and oxysterol 7alpha-hydroxylase CYP7B1: substrate specificity, azole binding and misfolding of clinically relevant mutants. *FEBS J* **281**, 1700-1713, doi:10.1111/febs.12733 (2014).

106 Knabe, C., Sudhop, T., von Bergmann, K. & Lutjohann, D. Degradation of 24S-hydroxycholesterol in men is not regulated by CYP7A1. *Int J Clin Pharmacol Ther* **45**, 577-582, doi:10.5414/cpp45577 (2007).

107 Russell, D. W., Halford, R. W., Ramirez, D. M., Shah, R. & Kotti, T. Cholesterol 24-hydroxylase: an enzyme of cholesterol turnover in the brain. *Annual review of biochemistry* **78**, 1017-1040, doi:10.1146/annurev.biochem.78.072407.103859 (2009).

108 Stiles, A. R. *et al.* Genetic, anatomic, and clinical determinants of human serum sterol and vitamin D levels. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4006-4014, doi:10.1073/pnas.1413561111 (2014).

109 Paul, S. M. *et al.* The major brain cholesterol metabolite 24(S)-hydroxycholesterol is a potent allosteric modulator of N-methyl-D-aspartate receptors. *J Neurosci* **33**, 17290-17300, doi:10.1523/JNEUROSCI.2619-13.2013 (2013).

110 Wei, X., Nishi, T., Kondou, S., Kimura, H. & Mody, I. Preferential enhancement of GluN2B-containing native NMDA receptors by the endogenous modulator 24S-hydroxycholesterol in hippocampal neurons. *Neuropharmacology* **148**, 11-20, doi:10.1016/j.neuropharm.2018.12.028 (2019).

111 Maioli, S. *et al.* Is it possible to improve memory function by upregulation of the cholesterol 24S-hydroxylase (CYP46A1) in the brain? *PloS one* **8**, e68534, doi:10.1371/journal.pone.0068534 (2013).

112 Mast, N. *et al.* Cholesterol-metabolizing enzyme cytochrome P450 46A1 as a pharmacologic target for Alzheimer's disease. *Neuropharmacology* **123**, 465-476, doi:10.1016/j.neuropharm.2017.06.026 (2017).

113 Lund, E. G. *et al.* Knockout of the cholesterol 24-hydroxylase gene in mice reveals a brain-specific mechanism of cholesterol turnover. *J Biol Chem* **278**, 22980-22988, doi:10.1074/jbc.M303415200 (2003).

114 Kotti, T. J., Ramirez, D. M., Pfeiffer, B. E., Huber, K. M. & Russell, D. W. Brain cholesterol turnover required for  
geranylgeraniol production and learning in mice. *Proceedings of the National Academy of Sciences of the United  
States of America* **103**, 3869-3874, doi:10.1073/pnas.0600316103 (2006).

115 Zoicas, I. & Kornhuber, J. The Role of the N-Methyl-D-Aspartate Receptors in Social Behavior in Rodents.  
*International journal of molecular sciences* **20**, doi:10.3390/ijms20225599 (2019).

116 Ripperger, S. P. *et al.* Vampire Bats that Cooperate in the Lab Maintain Their Social Networks in the Wild. *Current  
biology : CB* **29**, 4139-4144 e4134, doi:10.1016/j.cub.2019.10.024 (2019).

117 Baron, G., Stephan, H. & Frahm, H. D. *Comparative Neurobiology in Chiroptera: Brain characteristics in functional  
systems, ecoethological adaptation, adaptive radiation, and evolution.* (Springer, 1996).

118 Wang, C. C. *et al.* A critical role for GluN2B-containing NMDA receptors in cortical development and function.  
*Neuron* **72**, 789-805, doi:10.1016/j.neuron.2011.09.023 (2011).

119 Kohl, S. *et al.* A nonsense mutation in PDE6H causes autosomal-recessive incomplete achromatopsia. *Am J Hum  
Genet* **91**, 527-532, doi:10.1016/j.ajhg.2012.07.006 (2012).

120 Brennenstuhl, C. *et al.* Targeted ablation of the Pde6h gene in mice reveals cross-species differences in cone and  
rod phototransduction protein isoform inventory. *J Biol Chem* **290**, 10242-10255, doi:10.1074/jbc.M114.611921  
(2015).

121 Li, L. *et al.* Retention and losses of ultraviolet-sensitive visual pigments in bats. *Scientific reports* **8**, 11933,  
doi:10.1038/s41598-018-29646-6 (2018).

122 Kries, K. *et al.* Colour vision variation in leaf-nosed bats (Phyllostomidae): Links to cave roosting and dietary  
specialization. *Mol Ecol* **27**, 3627-3640, doi:10.1111/mec.14818 (2018).

123 Sadier, A. *et al.* Multifactorial processes underlie parallel opsin loss in neotropical bats. *eLife* **7**,  
doi:10.7554/eLife.37412 (2018).

124 Simoes, B. F. *et al.* As Blind as a Bat? Opsin Phylogenetics Illuminates the Evolution of Color Vision in Bats.  
*Molecular biology and evolution* **36**, 54-68, doi:10.1093/molbev/msy192 (2019).

125 Chang, B. *et al.* A homologous genetic basis of the murine cpfl1 mutant and human achromatopsia linked to  
mutations in the PDE6C gene. *Proceedings of the National Academy of Sciences of the United States of America*  
**106**, 19581-19586, doi:10.1073/pnas.0907720106 (2009).

126 Grau, T. *et al.* Decreased catalytic activity and altered activation properties of PDE6C mutants associated with  
autosomal recessive achromatopsia. *Human molecular genetics* **20**, 719-730, doi:10.1093/hmg/ddq517 (2011).

127 Emerling, C. A., Widjaja, A. D., Nguyen, N. N. & Springer, M. S. Their loss is our gain: regressive evolution in  
vertebrates provides genomic models for uncovering human disease loci. *J Med Genet* **54**, 787-794,  
doi:10.1136/jmedgenet-2017-104837 (2017).

128 Crespo, R. F., Linhart, S. B., Burns, R. J. & Mitchell, G. C. Foraging Behavior of the Common Vampire Bat Related  
to Moonlight. *Journal of Mammalogy* **53**, 366-368, doi:10.2307/1379175 (1972).

129 Groger, U. & Wiegreb, L. Classification of human breathing sounds by the common vampire bat, *Desmodus  
rotundus*. *BMC Biol* **4**, 18, doi:10.1186/1741-7007-4-18 (2006).

130 Escalera-Zamudio, M. *et al.* The evolution of bat nucleic acid-sensing Toll-like receptors. *Mol Ecol* **24**, 5899-5909,  
doi:10.1111/mec.13431 (2015).

131 Castillo, D. J., Rifkin, R. F., Cowan, D. A. & Potgieter, M. The Healthy Human Blood Microbiome: Fact or Fiction?  
*Front Cell Infect Microbiol* **9**, 148, doi:10.3389/fcimb.2019.00148 (2019).

132 Scarsella, E., Sandri, M., Monego, S. D., Licastro, D. & Stefanon, B. Blood Microbiome: A New Marker of Gut  
Microbial Population in Dogs? *Vet Sci* **7**, doi:10.3390/vetsci7040198 (2020).

133 Rademacher, F. *et al.* The Antimicrobial and Immunomodulatory Function of RNase 7 in Skin. *Front Immunol* **10**,  
2553, doi:10.3389/fimmu.2019.02553 (2019).

134 Eichler, T. E. *et al.* Insulin and the phosphatidylinositol 3-kinase signaling pathway regulate Ribonuclease 7  
expression in the human urinary tract. *Kidney Int* **90**, 568-579, doi:10.1016/j.kint.2016.04.025 (2016).

135 Rodriguez-Carlos, A. *et al.* Host Defense Peptide RNase 7 Is Down-regulated in the Skin of Diabetic Patients with  
or without Chronic Ulcers, and its Expression is Altered with Metformin. *Arch Med Res* **51**, 327-335,  
doi:10.1016/j.arcmed.2020.03.006 (2020).

136 Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737-  
746, doi:10.1038/s41586-021-03451-0 (2021).

137 Protzek, A. O. *et al.* Insulin and glucose sensitivity, insulin secretion and beta-cell distribution in endocrine  
pancreas of the fruit bat *Artibeus lituratus*. *Comparative biochemistry and physiology. Part A, Molecular &  
integrative physiology* **157**, 142-148, doi:10.1016/j.cbpa.2010.05.016 (2010).

138 Michelmore, A. J., Keegan, D. J. & Kramer, B. Immunocytochemical identification of endocrine cells in the  
pancreas of the fruit bat, *Rousettus aegyptiacus*. *General and comparative endocrinology* **110**, 319-325,  
doi:10.1006/gcen.1998.7077 (1998).

139 Priyadarshini, M. & Layden, B. T. FFAR3 modulates insulin secretion and global gene expression in mouse islets.  
*Islets* **7**, e1045182, doi:10.1080/19382014.2015.1045182 (2015).

140 Hecker, N., Sharma, V. & Hiller, M. Transition to an Aquatic Habitat Permitted the Repeated Loss of the Pleiotropic  
KLK8 Gene in Mammals. *Genome Biol Evol* **9**, 3179-3188, doi:10.1093/gbe/evx239 (2017).

141 Sharma, V. & Hiller, M. Loss of enzymes in the bile acid synthesis pathway explains differences in bile composition  
among mammals. *Genome Biol Evol* **10**, 3211-3217, doi:10.1093/gbe/evy243 (2018).

142 Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* **64**, 18-23,  
doi:10.1086/302219 (1999).

- 143 Sikes, R. S., Animal, C. & Use Committee of the American Society of, M. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J Mammal* **97**, 663-688, doi:10.1093/jmammal/gyw078 (2016).
- 144 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170-175, doi:10.1038/s41592-020-01056-5 (2021).
- 145 Ghurye, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS computational biology* **15**, e1007273, doi:10.1371/journal.pcbi.1007273 (2019).
- 146 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 147 Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* **19**, 125, doi:10.1186/s13059-018-1486-1 (2018).
- 148 Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**, 983-987, doi:10.1038/nbt.4235 (2018).
- 149 Harris, R. S. *Improved pairwise alignment of genomic DNA* Ph.D. Thesis thesis, The Pennsylvania State University, (2007).
- 150 Sharma, V. & Hiller, M. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res* **45**, 8369-8377, doi:10.1093/nar/gkx554 (2017).
- 151 Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11484-11489, doi:10.1073/pnas.1932072100 (2003).
- 152 Osipova, E., Hecker, N. & Hiller, M. RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**, doi:10.1093/gigascience/giz132 (2019).
- 153 Suarez, H. G., Langer, B. E., Ladde, P. & Hiller, M. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596-1603, doi:10.1093/bioinformatics/btx024 (2017).
- 154 Sharma, V., Schwede, P. & Hiller, M. CESAR 2.0 substantially improves speed and accuracy of comparative gene annotation. *Bioinformatics* **33**, 3985-3987, doi:10.1093/bioinformatics/btx527 (2017).
- 155 Sharma, V., Elghafari, A. & Hiller, M. Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation. *Nucleic Acids Res* **44**, e103, doi:10.1093/nar/gkw210 (2016).
- 156 Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res*, doi:10.1093/nar/gkaa1087 (2020).
- 157 Foley, N. M., Springer, M. S. & Teeling, E. C. Mammal madness: is the mammal tree of life not yet resolved? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371**, doi:10.1098/rstb.2015.0140 (2016).
- 158 Lee, C. M. *et al.* UCSC Genome Browser enters 20th year. *Nucleic Acids Res* **48**, D756-D761, doi:10.1093/nar/gkz1012 (2020).
- 159 Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular biology and evolution* **35**, 2582-2584, doi:10.1093/molbev/msy159 (2018).
- 160 Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19**, 21, doi:10.1186/s12862-019-1350-2 (2019).
- 161 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 162 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).