

Positive Selection and Gene Expression Analyses from Salivary Glands Reveal Discrete Adaptations within the Ecologically Diverse Bat Family Phyllostomidae

Michael W. Vandewege ^{1,2,*}, Cibele G. Sotero-Caio^{1,3}, and Caleb D. Phillips¹

¹Department of Biological Sciences, Texas Tech University

²Present address: Department of Biology, Eastern New Mexico University, Portales, NM

³Present address: Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, United Kingdom

Data deposition: Raw RNASeq files have been deposited at SRA under the BioProject accession PRJNA451361.

Accepted: 17 July 2020

*Corresponding author: E-mail: mike.vandewege@gmail.com.

Abstract

The leaf-nosed bats (Phyllostomidae) are outliers among chiropterans with respect to the unusually high diversity of dietary strategies within the family. Salivary glands, owing to their functions and high ultrastructural variability among lineages, are proposed to have played an important role during the phyllostomid radiation. To identify genes underlying salivary gland functional diversification, we sequenced submandibular gland transcriptomes from phyllostomid species representative of divergent dietary strategies. From the assembled transcriptomes, we performed an array of selection tests and gene expression analyses to identify signatures of adaptation. Overall, we identified an enrichment of immunity-related gene ontology terms among 53 genes evolving under positive selection. Lineage-specific selection tests revealed several endomembrane system genes under selection in the vampire bat. Many genes that respond to insulin were under selection and differentially expressed genes pointed to modifications of amino acid synthesis pathways in plant-visitors. Results indicate salivary glands have diversified in various ways across a functional diverse clade of mammals in response to niche specializations.

Key words: submandibular glands, selection, molecular evolution, immunity, Chiroptera.

Significance

The bat family Phyllostomidae is one of the most ecologically diverse families of mammals. Salivary glands are hypothesized to facilitate adaptation to novel diets because their secreted products are the first to come in contact with food and pathogens. We sequenced expressed transcripts from phyllostomid salivary glands and found strong signals of selection among immune-related genes. Selection and gene expression signals among specific lineages were less clear but pointed to modifications of the endomembrane transport system and metabolic pathways. Although we could not strongly link gene evolution to dietary adaptations, results indicated diversification in response to niche specializations.

Introduction

Bats (Order Chiroptera) make up 20% of mammalian diversity and several groups have been alluded to as examples of extreme phenotypic and genomic changes leading to species diversity (Ray et al. 2006; Pritham and Feschotte 2007;

Dumont et al. 2012; Hayden et al. 2014; Platt et al. 2014; Phillips and Baker 2015; Sotero-Caio et al. 2015). In particular, the leaf-nosed bats (Phyllostomidae) include >200 extant species, representing the most ecologically diverse mammal family with their wide range of dietary strategies practiced (Baker et al. 2016; Cirranello et al. 2016). Morphometric and

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

evolutionary models suggest the ecological and phenotypic diversity among lineages and species are associated with strong selection on dietary specialization features (Monteiro and Nogueira 2011; Rojas et al. 2011; Dumont et al. 2012; Rossoni et al. 2017; Hedrick et al. 2019).

Bats exhibit several characteristics rendering them interesting to examine, most obviously their ability to fly and echolocate. Furthermore, bats act as vectors for zoonotic diseases including SARS coronavirus, Ebola, Nipah virus, and rabies (Calisher et al. 2006; Smith and Wang 2013; Olival et al. 2017). A lesser examined characteristic thought to be directly linked to their dietary multiplicity and adaptation is the anatomical diversity of their salivary glands (Phillips et al. 1998; Tandler et al. 2001). In mammals, there are typically three major pairs of salivary glands: the submandibular, parotid, and sublingual glands, and all use intracellular processes that involve the synthesis, modification, packaging, and secretion of proteins in membrane-bound granules. These glands are considered part of the digestive system as they secrete digestive enzymes, however, their products perform additional functions including antimicrobial resistance and biochemical communication (Dobrosielski-Vergona 1993; Talley et al. 2001; Bloss et al. 2002; Safi and Kerth 2003; Vandewege et al. 2013).

Tandler and Phillips (1998) have shown that among 55 species of bats, secretory products were correlated with diet, especially in insectivorous species. Further, Phillips et al. (2014) found evidence that salivary glands play a role in lipid metabolism in *Myotis lucifugus*. Because of this wide variation and their direct link to immunity, diet, and reproduction, submandibular glands (SMGs) and their products are hypothesized to play an important role in the adaptive radiation of mammals (Phillips and Tandler 1996).

Linking genetic variation and gene products to selection and adaptation is still a challenge in evolutionary biology. More recently, the implementation of next-generation sequencing has facilitated the search for signatures of adaptation through DNA and/or RNA comparative analyses. Indeed, sequencing transcriptomes offers an in-depth, data-rich method to identify selective pressures that have influenced the evolution of tissue-specific genes. Of interest are genes that have undergone positive selection to adapt physiological, immunological, and ecological processes to new environments (Daugherty and Malik 2012; Hawkins et al. 2019).

Selection analyses among bat genes have generally been limited to few species for specific purposes (Shen et al. 2010; ZhaNg et al. 2013). Hawkins et al. (2019) analyzed orthologs from 18 different bat genomes and transcriptomes and found most genes under selection were related to immunity and collagen production. Here, Phyllostomidae was examined specifically because of their extensive and relatively rapid radiation to new feeding strategies. We sequenced the SMG transcriptomes of nine phyllostomid bats representing different subfamilies and different diets, and through analysis of

orthologs characterized how selection on coding sequence and expression differences have shaped SMGs.

Materials and Methods

Nine species from seven out of the 11 recognized subfamilies were chosen to maximize representation of the phylogenetic and dietary diversity of Phyllostomidae (fig. 1). We also included two insectivorous bats, *M. lucifugus* from Vespertilionidae and *Pteronotus parnellii* from Mormoopidae, as outgroups. Tissues were extracted and frozen in liquid nitrogen within 5 min after euthanasia. Additional details from tissue loans provided by the Natural Science Research Laboratory (NSRL) of the Museum of Texas Tech University can be found in [supplementary table S1, Supplementary Material](#) online.

RNA Isolation, Sequencing, and Assembly

RNA was extracted from SMGs of each bat using Trizol (Invitrogen, Carlsbad, CA, USA) following manufacturer protocols. Oligo-dT magnetic beads were used to enrich for mRNA strands with poly-A tails and a strand-specific paired-end cDNA library was prepared using a ScriptSeq kit (Epicentre, Madison WI USA). Libraries were sequenced on Illumina platform (see [supplementary table S1, Supplementary Material](#) online). For each species, paired-end reads were filtered for quality using Trimmomatic 0.36 (Bolger et al. 2014) with the parameters LEADING:20 TRAILING:20 SLIDINGWINDOW:1:30 MINLEN:75 (MINLEN:50 was used for *Desmodus rotundus*), and de novo transcript assemblies were constructed from clean read pairs with Trinity v2.7 (Grabherr et al. 2011) using default settings.

Gene Discovery and Annotation

Putative open reading frames and translated peptide sequences were identified using TransDecoder (Haas et al. 2013) and the resulting peptide sequences were processed through the Trinotate pipeline to identify functional properties and Gene Ontology (GO) annotations associated with biological processes, molecular functions, and cellular components. To summarize, peptide sequences were queried against the SwissProt database (Dimmer et al. 2012) using BlastP (Altschul et al. 1997) and the Pfam (Finn et al. 2010) database using hmmer 3.2.1 (Eddy 2011). Peptides were also scanned for a signal peptide using signalP (Petersen et al. 2011), and transmembrane domains using tmhmm 2.0 (Krogh et al. 2001).

Orthology Clustering

Orthology assignment is still a major challenge in bioinformatics and evolutionary biology. Here, we developed a process to

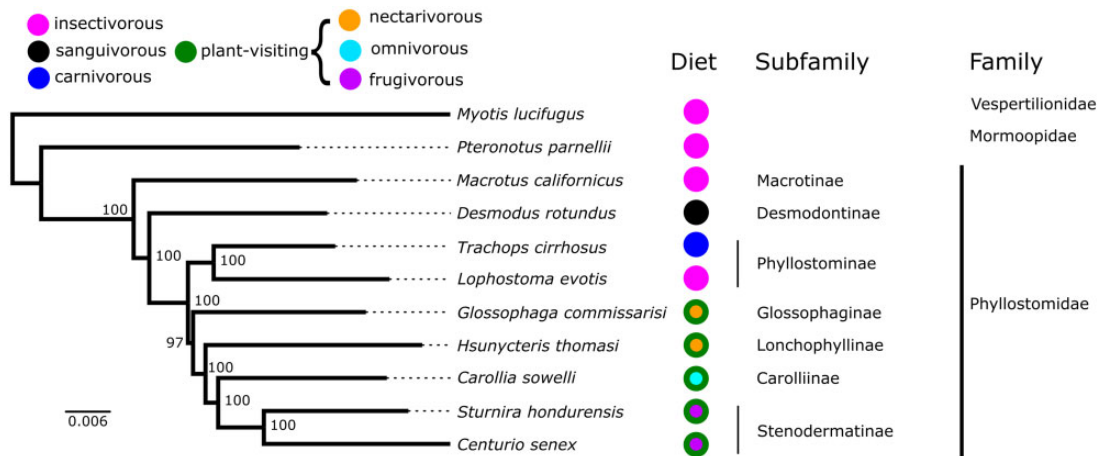


FIG. 1.—Maximum likelihood phylogenetic tree constructed from 500 randomly sampled genes demonstrating sampled species relationships and general dietary strategies. Node support was derived from 100 bootstrap replicates.

filter out similar transcripts produced by Trinity. The first step was to assume similar sequences would have similar Trinotate annotations. We parsed the Trinotate output to identify the best sequence representative for each unique SwissProt annotation. To choose the best gene representative among multiple coding sequences (CDSs) with the same SwissProt annotation, we multiplied the length of the CDS to the percent identity of the SwissProt hit. This metric correlated with E value, but effectively acted as a bit score when E values were identical. The CDS with the highest metric was chosen to represent the annotation. If a CDS did not have a SwissProt annotation, it was removed. We then ran combined best sequences from all species, and ran this data set through the OrthoMCL (Li et al. 2003) pipeline to identify orthologous groups. Only single gene ortholog groups were used in downstream selection tests.

Poor ortholog assignment can influence sequence relationships in a phylogeny, and because the relationships among phyllostomids are robust, a reasonable test of ortholog assignment would be to reconstruct a phylogenetic tree and determine if the resulting tree reflects previously described relationships. Therefore, we reconstructed a phylogenetic tree from 500 randomly sampled single gene orthologs shared among all 11 individuals. Each orthologous group was translated, aligned using LINSi parameters in MAFFT (Kato and Standley 2013) and reverse translated to construct a codon alignment. Resulting alignments were concatenated and we used RAxML (Stamatakis 2014) to find the best tree from the unpartitioned data set using the ML and rapid bootstrapping algorithm, a GTRGAMMA model of nucleotide substitution, and 100 bootstrap replicates.

Identifying Genes under Selection

Single gene orthologous groups that were found in seven or more phyllostomids were tested for evidence of selection

using the maximum likelihood approach described by Goldman and Yang (1994). Codon alignments were constructed as above and the best tree for each gene was estimated using RAxML. We used CODEML in PAML (Yang 2007) to estimate the role of selection on gene evolution by comparing the rate of nonsynonymous substitution per nonsynonymous site (d_N) to the rate of synonymous substitution per synonymous site (d_S). The d_N/d_S ratio can be used as a sensitive measure of selective pressure; however, in most cases, the overall d_N/d_S is <1 and only a few amino acid sites are evolving quickly. Therefore, to determine whether a gene was evolving adaptively, we calculated the likelihood of models that allow d_N/d_S to vary among codon sites (M1a, M2a, M7, and M8). For all genes, we used likelihood ratio tests (LRTs) to compare nested models that allow and disallow codon site d_N/d_S to be >1 (M2a v M1a, M8 v M7), and to test for significant differences between nested models (Yang 1998). If both models that allow d_N/d_S rates to be >1 (M2a and M8) were significantly better fits to the data (LRT, $P < 0.05$), we inferred these genes were evolving under positive selection. We performed a false discovery rate (FDR) correction on the P values resulting from the M2a v M1a and M7 v M8 LRTs. FDRs were estimated using the q value function in the “ q value” R module (Storey and Tibshirani 2003; Storey 2020).

We also conducted branch-site selection tests in CODEML (Yang and Nielsen 2002; Zhang et al. 2005) for ortholog groups that were present among all 11 species. Alignments were constructed as previously described, but we used the species tree generated above (fig. 1). The branch-site test for positive selection divides branches of a phylogenetic tree into foreground and background branches. The null model (Model A) restricts positive selection among codons in both foreground and background branches. The alternative (Model B) allows positive selection to occur among codon in foreground branches. Likelihoods between Model A and Model B were compared using a LRT. The FDR was also estimated

from the distribution of P values. We conducted independent branch-site tests where the *D. rotundus* branch was the foreground, a second test with *Trachops cirrhosus* as foreground, and a third test that included the entire plant-visitors clade as foreground (fig. 1).

Gene Expression

We mapped quality filtered paired-end RNAseq reads back to reference sequences using Bowtie 1.2.2 (Langmead et al. 2009) and default parameters of RSEM (Li and Dewey 2011). Only ortholog groups that were present in four or more species were present in reference sequences. Ortholog groups with <10 mapped reads across all 11 samples were removed prior to analyses. Normalization of raw read counts was performed using the estimateSizeFactors and estimateDispersions functions in DESeq2 v1.26 (Love et al. 2014). Patterns of species variation in expression levels were assessed by performing a phylogenetically corrected PCA using the *phytools* R package (Revell 2012) based on the blind variance stabilizing transformed data. We also conducted a differential expression analysis from the normalized count data between plant-visitors versus others (see fig. 1) and determined statistical significance when the adjusted P value was < 0.1. We repeated this analysis examining only nectarivores.

Functional Analyses

For all single gene orthologs tested for selection, we counted the number of reoccurring GO terms. We then used the GO term proportion to summarize the general function of genes tested in the SMGs of phyllostomids using REVIGO (Supek et al. 2011), which nests redundant and similar terms from long GO term lists by semantic clustering. We repeated the same analysis using GO terms described from orthologs under positive selection. To test for GO term enrichment for the genes under selection, PANTHER (Mi et al. 2016) was applied to identify statistically overrepresented GO terms from the list of genes under positive selection using the genes tested as the reference list. We used the Fisher's exact test and applied the FDR correction and statistical significance was determined below an adjusted P value <0.05. We estimated whether a protein was membrane bound, a receptor, immune related, or secreted by searching for specific GO terms. If "secretion," "extracellular space," or "extracellular region" was present among GO terms, we annotated the gene as secreted. If "immune," "defense," or "antimicrobial" was present, the gene was annotated as an immune. If "membrane" was present we called the gene membrane bound and if "receptor" was present, we called the gene a receptor (supplementary table S2, Supplementary Material online). We also used DAVID 6.8 (Huang et al. 2009) to map KEGG pathways to differentially expressed genes.

Results

To investigate putative proteins underlying the evolution of SMGs, we performed RNAseq on 11 Chiroptera species: nine phyllostomids with varying dietary strategies, and two outgroups representing the bat ancestral state of strict insectivory (fig. 1). Among the 11 samples, between 53,757 and 105,666 transcripts were assembled per species. Over 87% of the paired-reads mapped back to each assembly indicating most reads were used. Among the transcripts, we predicted between 17,820 and 54,578 ORFs, and among the ORFs, we found between 5,392 and 9,580 unique hits to the SwissProt database (supplementary table S1, Supplementary Material online). The most similar isoform to the reference sequence was chosen to represent the annotation. SwissProt annotations were not used a priori for ortholog clustering. However, if after clustering there were multiple SwissProt annotations in a single gene ortholog group, the most common annotation was used to describe the ortholog group.

In all, OrthoMCL grouped 79,817 annotations into 10,267 orthologous groups that included between 2 and 45 members (supplementary fig. S1A, Supplementary Material online). Among the ortholog groups, 6,694 were single gene families represented in four or more species and 2,247 were found in all 11 species (supplementary fig. S1B, Supplementary Material online). Among these 6,694 groups, 6,322 (94%) had the same SwissProt annotation among orthologs, suggesting overall consistency between OrthoMCL clustering and Trinotate annotations. In the remaining 372 groups, there was one dominant gene annotation and the outliers were likely a result of improper classification by Trinotate, given BLAST hits are often closely related paralogs and not true orthologs, in which case annotations were manually corrected.

To additionally test if the ortholog predictions were generally accurate and DNA sequences largely reflected species relationships, we generated a phylogenetic tree from the concatenated alignments of 500 randomly sampled orthologs that were present in all species. Although individual gene trees may not reflect a species trees, we expected that a consensus would emerge from a large volume of DNA sequences if orthologs were accurately predicted and aligned. We found that the resulting concatenated tree accurately reflected previous species trees generated by Dumont et al. (2012), Baker et al. (2016), and Rojas et al. (2016), with high bootstrap support at each node (fig. 1).

Adaptation among SMG Proteins

Among the 6,694 orthologs, 4,007 were single gene orthologs present in seven or more phyllostomids, and 2,920 had alignments with enough shared sites to produce output from CODEML. After correcting for FDR, we found 53 genes where models of evolution allowing positive selection were significantly better fit to the data than neutrality in both M2a and

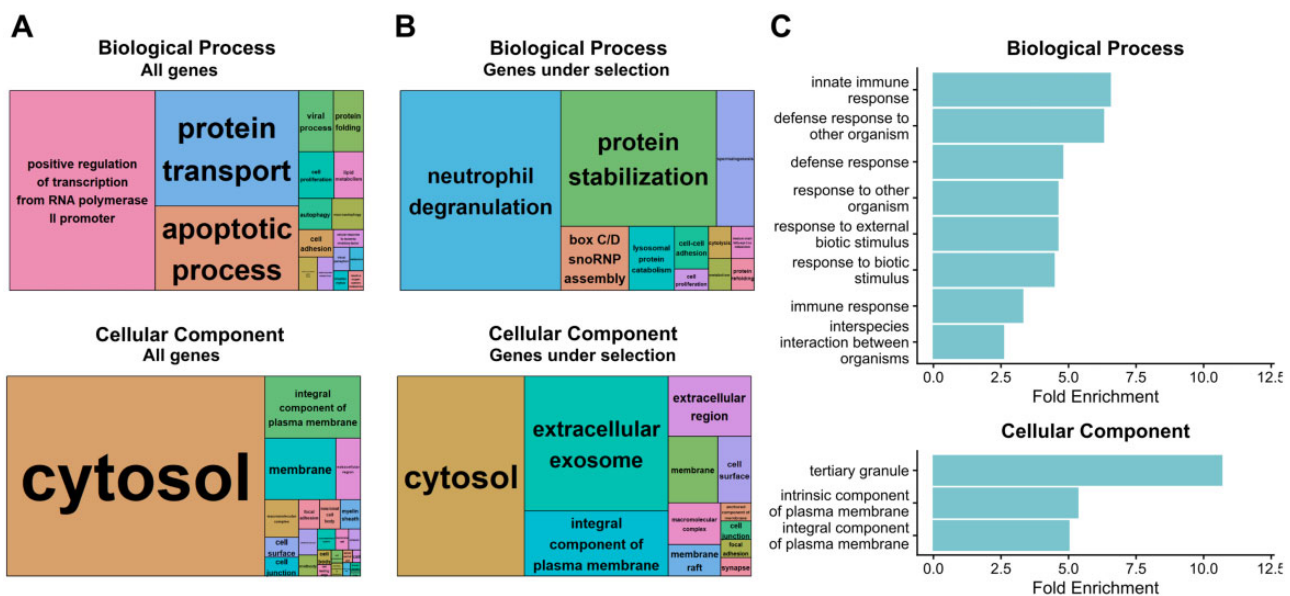


FIG. 2.—Tree map produced from the analysis of redundant GO terms in REVIGO for (A) all GO terms linked to genes tested and (B) GO terms linked to genes under selection. (C) Fold enrichment for Biological Processes and Cellular Components GO terms that were statistically overrepresented among genes under positive natural selection.

M8 tests (supplementary table S2, Supplementary Material online).

To contrast background salivary gene functions to those under selection, we summarized the GO terms of the 2,920 tested genes using the REVIGO server (fig. 2A). Overall, most of these gene products were localized to the cytosol, but many were also positioned in the plasma membrane and extracellular region. Common biological processes identified were related to transcription regulation, cell death, and protein synthesis and transport. By contrast, the biological process and cellular component terms of genes under selection were fairly different from those obtained in our global assessment of protein function (fig. 2B). These proteins had more terms related to the extracellular exosome or plasma membrane and were involved in the immune response (fig. 2B). Interestingly, no terms related to diet or metabolism (i.e., carbohydrate or lipid metabolism) were associated with genes under positive selection. We used PANTHER's overrepresentation test to determine whether any GO terms were significantly overrepresented among genes under selection. There were no molecular function terms enriched among the genes under selection (supplementary fig. S2, Supplementary Material online). The most enriched biological process term was innate immune response, but all enriched biological process terms were related to the defense against other organisms (fig. 2C). Consistent with receptor-like proteins under selection, terms associated with membrane surfaces were enriched in the cellular component category (fig. 2C). Upon survey of the multiple GO terms assigned to each gene, we found that eight of the 53 genes were involved in the immune response (supplementary table S2, Supplementary Material

online). Moreover, 50% of immunity-related loci had secretory annotations although a binomial test did not suggest secretion among immune proteins was enriched ($P=0.46$). Of the remaining 45 genes, 18 genes were membrane bound, two were secreted, and two had terms for both membranes and secretion.

Site tests can inform about positive selection on a protein among all species, but they cannot inform about episodic selection on specific lineages. Therefore, we conducted branch-site tests using the species tree as a reference on the most unique lineages, namely the plant-visitors, *D. rotundus*, and *T. cirrhosus*, and the plant-visiting species. We identified 2,247 single gene orthologs that were present among all 11 species and 1,590 had enough shared sites to be successfully tested in CODEML. After correcting for FDR, we found 24, 20, and 13 genes under positive selection in the plant-visitors, *D. rotundus*, and *T. cirrhosus*, respectively (supplementary table S3, Supplementary Material online) making up between 0.8% and 1.5% of genes tested. *BPIA2* (BPI fold-containing family A member 2), an immune response gene, was the only gene found to be under selection in both the plant-visitors and *D. rotundus* and approaching statistical significance in *T. cirrhosus* (supplementary table S3, Supplementary Material online). Interestingly, *BPIA2* was also under selection in site tests (supplementary table S2, Supplementary Material online), suggesting rapid evolution in codons among all phyllostomids. No GO terms were significantly overrepresented among genes under positive selection in any lineage and there were few overlapping GO terms among genes under selection in the three lineages, except for terms related to the Golgi body (supplementary fig. S3, Supplementary Material online).

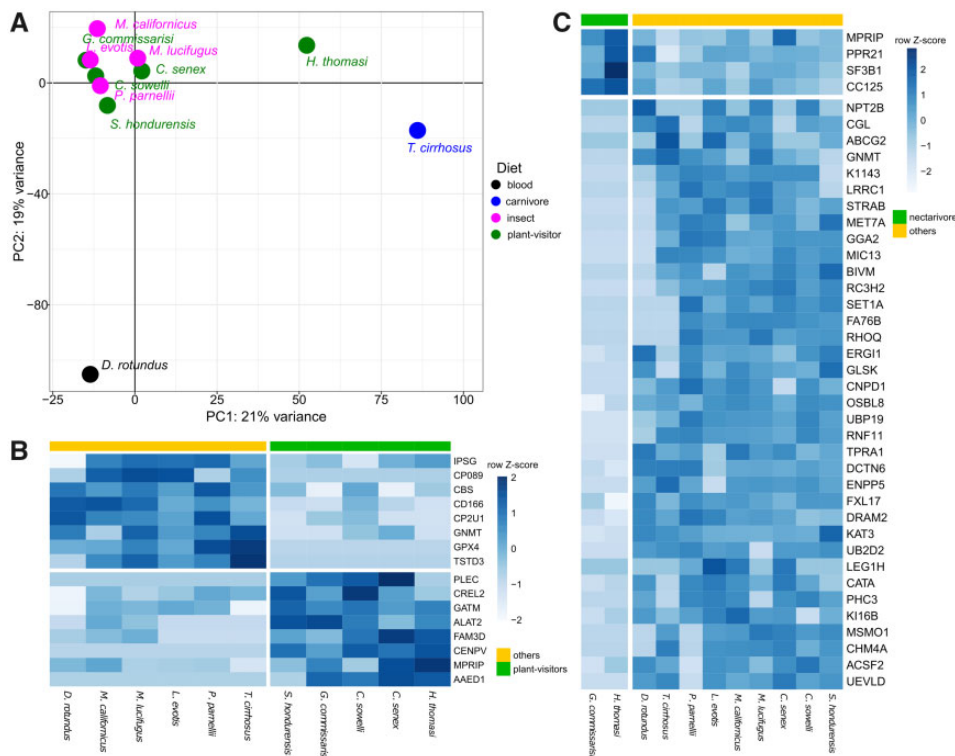


Fig. 3.—(A) Phylogenetically corrected PCA of the variance-stabilized gene expression data. Diets are color coded based on figure 1. (B) Expression heatmap of the 16 differentially expressed genes between plant-visitors and remaining species. (C) Expression heatmap of the 40 differentially expressed genes between nectarivores and remaining species.

Expression Profiles

We measured the expression of 6,692 orthologs among all 11 species. We performed a phylogenetically corrected PCA on the variance stabilizing transformed expression data and found that all species except *D. rotundus*, *T. cirrhosus*, and *Hsunnycteris thomasi* had relatively correlated expression profiles where most of the insectivores and plant-visitors clustered together (fig. 3A). As there were not any biological replicates and few dietary replicates among our samples, the most robust approach to identifying relevant differentially expressed genes was to compare the plant-visitors to the remaining species to identify expression differences that could explain how plant-visitors have adapted to a diet with a different macromolecular profile. Sixteen genes were differentially expressed in plant-visitors, eight were upregulated, and eight were downregulated (fig. 3B), and no GO terms or KEGG pathways were significantly enriched among these 16 genes. However, the most common pathways associated with these differentially expressed genes were metabolic pathways (supplementary table S4, Supplementary Material online). We performed another differential expression experiment by just examining the nectarivores. Interestingly, 36 out of 40 differentially expressed genes were downregulated in nectarivores (fig. 3C), many of these genes were also involved

in metabolic pathways (supplementary table S5, Supplementary Material online).

Discussion

Salivary glands secrete products that have important biological roles in diet/digestion, oral health, and communication, as well as display the most varied cellular ultrastructure in distinct taxa (Tandler and Phillips 1998). The great morphological variation observed in the ultrastructure of phyllostomid bat salivary gland granules led Phillips and Tandler (1996) to speculate that adaptation to novel niches drives salivary gland evolution. SMG acinar cells are responsive to a variety of extracellular signals which can affect gene expression, protein synthesis, and protein modifications, and this responsiveness can be driven by the density and distribution of cell surface receptors (Phillips et al. 1998). Salivary proteins are among the first to directly encounter food and introduced pathogens, which lends biological importance in the context of adaptive evolution. However, their role in the adaptation of mammals to novel niches has been largely under investigated. Here, we sequenced the SMG transcriptomes of nine phyllostomid bats with varying feeding strategies to illuminate the role of salivary glands in this adaptive radiation. Interestingly, the percentage

of genes under selection expressed in phyllostomid salivary glands was not necessarily greater than genes under selection among bats as a whole (Hawkins et al. 2019). Further, links to dietary changes were not strongly apparent, but signatures of selection did reveal modification of SMG functions.

Selection among Phyllostomid Genes

Among the sequenced transcriptomes, we were able to annotate between 5,392 and 9,580 expressed proteins, which were successfully clustered as 6,694 single gene orthologs. About 4,007 were present in seven or more phyllostomids and tested for selection among codon sites. Using codon selection models, we found only a small proportion (1.8%) of genes demonstrated signatures of positive selection. Consistently, Hawkins et al. (Hawkins et al. 2019) yielded the same frequency of genes under selection from the transcriptomes and genomes of 18 bat species using multiple tissue sources (e.g., spleen, trigeminal ganglia, inner ear, and embryos transcriptomes). This suggests that overall, the genes expressed in SMGs phyllostomids are not necessarily subjected to stronger selective pressures in coding regions than genes expressed in other tissues among bats.

Salivary glands are known to function in immunity through the localized function of immune cells (Fábián et al. 2012) and immunity was the only enriched term among genes under selection (fig. 2C). Eight of the 53 proteins inferred to be affected by positive selection were linked to immunity and defense, and five of these had a secretory component (supplementary table S2, Supplementary Material online). These eight genes had common functions associated with the innate and humoral immune response and stimulate the activation of immune cells to viral infected sites. For example, BPIA1 (BPI fold-containing family A member 1) and BPIA2 are secreted and known to localize to upper airways and prevent biofilm formation by Gram-negative bacteria (Liu et al. 2013; Prokopovic et al. 2014). A high exposure to pathogens unique to bats could explain this result, but positive selection among immune genes is not uncommon in animals (Roux et al. 2014; Xiao et al. 2015; van der Lee et al. 2017; Hawkins et al. 2019). Defense and immunity genes typically have a high evolutionary rate attributed to a continuous arms race between pathogens and host immune response (Jiggins and Kim 2007; Kosiol et al. 2008; Viljakainen et al. 2009), and our data provide insight about how positive selection has shaped bat oral tract defense against introduced pathogens.

Lineage-Specific Adaptations

An unexpected facet of our results was a general paucity of genes under selection with roles related to metabolism among site tests. Therefore, to identify lineage-specific trends, we used branch-site models and tested for episodic selection in groups that exhibit the most divergence from the ancestral insectivorous trait: the common vampire bat *D. rotundus*, the

frog-eating bat *T. cirrhosus*, and all plant-visitors. There was a similar frequency of genes under positive selection as site tests, between 0.5% and 1.5%, and no GO terms were enriched among these genes. Most GO terms were not redundant among the three tests, except for associations with the Golgi body, which is expected given the importance of proper secretory function of salivary glands. The Golgi body is the central organizer of the membrane trafficking system. It is where proteins are sorted into vesicles and trafficked out of the cell. However, it is noteworthy that the vampire bat displayed an overrepresentation of Golgi-related ontology terms and genes under selection, when compared with the other two tested groups.

Apparent adaptation or specialization of the Golgi body and endomembrane system was present in the vampire bat *D. rotundus*. The common vampire bat belongs to the subfamily Desmodontinae, the only obligate sanguivores among amniotes, which may be the most ecologically divergent among the phyllostomids. Francischetti et al. (2013) previously performed a transcriptomic analysis of vampire bat salivary glands and found a diversity of anticoagulants, anti-inflammatory proteins, and neural disruptors hypothesized to enhance the efficiency of parasitizing animals for blood meals. Thus, unlike other phyllostomids, secreted proteins may not only be helpful to digestion but effective adaptations to blood-feeding with minimal death risk for the bat and its prey. Branch-site tests in *D. rotundus* identified 20 genes under selection (supplementary table S3, Supplementary Material online). Notable GO terms among these 20 genes pointed to the regulation of protein trafficking, that is, Golgi organization, protein N-linked glycosylation, trans-Golgi network, and positive regulation of secretion (supplementary fig. S3, Supplementary Material online). The genes linked to these terms were Golgi phosphoprotein 3-like (*GLP3L*), vesicle-associated membrane protein-associated A (*VAPA*), and NSF1 cofactor p47 (*NSF1C*). *VAPA* can mediate vesicle transport from the endoplasmic reticulum (ER) to the Golgi body (Lehto et al. 2005). *GLP3L* is generally expressed in secretory tissues, localizes to Golgi, is required for efficient anterograde trafficking, and knocking out *GLP3L* causes Golgi dispersal and impairs secretion (Ng et al. 2013). *NSF1C* seems to be relevant for Golgi reassembly after mitosis (Pécheur et al. 2002). Lastly, a gene under selection in the N-glycan biosynthesis pathway was *PMM2* (phosphomannomutase 2). *PMM2* specifically catalyzes the isomerization of mannose 6-phosphatase to mannose 1-phosphate. Deleterious mutations in *PMM2* tend to cause defects in protein glycosylation and subsequent congenital disorders (Matthijs et al. 1997). Vampire bats diverged from an insectivorous ancestor and subsequently underwent a rapid transition from insectivory to sanguivory (Datzmann et al. 2010; Dumont et al. 2012). Given our results, it is plausible the endomembrane system was modified during this transition. Moreover, given that some Golgi body-related genes appeared under selection in

all branch-site tests, this organelle played some role in the adaptive radiation of phyllostomids.

Another interesting result from branch-site tests was a response to insulin stimulus in the plant-visitors. Two genes under selection, ubiquitin-conjugating enzyme E2B (*UBE2B*) and CASP8 and FADD-like apoptosis regulator (*CFLAR*), were specifically associated with this term. *CFLAR* regulates downstream genes involved in lipid metabolism, glucose uptake, and oxidative stress. In mice, *CFLAR* was shown to reverse nonalcoholic steatohepatitis liver disease (Wang et al. 2017) that can be caused by insulin resistance and high blood sugar among other metabolic factors (Liu et al. 2019). The exact function of *UBE2B* is a little less clear but appears to be linked to muscle atrophy. *UBE2B* becomes more expressed in fasting rats and atrophying muscle cells, but becomes suppressed in response to insulin (Wing and Banville 1994; Polge et al. 2015). The plant-visitors make up a collection of frugivorous and nectarivorous species that ingest high amounts of sugars at once (Laska 1990). Interestingly, the phyllostomid great fruit-eating bat (*Artibeus lituratus*) exhibits no difference in serum insulin levels between fasted and fed states (Protzek et al. 2010). Moreover, serum insulin levels observed in *A. lituratus* were higher than those observed in mice (Fujiwara et al. 2007) and obese humans (Yassine et al. 2009). We can speculate physiological changes, that is, constantly high concentration of insulin, may have caused a selective response in insulin associated pathways.

Gene Expression Patterns

Selection pressures may not only occur at the sequence level but also the expression regulation level. Therefore, we paired selection tests with gene expression analyses. The distinct natural histories represented by *T. cirrhosus* and *D. rotundus* were reflected in the PCA. In fact, the only way *T. cirrhosus* stood out in this study was in the expression PCA. *Trachops cirrhosus* specializes on frogs (Tuttle and Ryan 1981) which is a unique feeding strategy in phyllostomids, but data from the vespertilionid *Nyctalus lasiopterus* suggests the transition from insects to carnivory may not require any major adaptations (Ibáñez et al. 2001). Consistently, selection pressures, given branch-site tests, were weakest in *T. cirrhosus* (supplementary table S3, Supplementary Material online).

The most remarkable observation from overall gene expression profiles was a general correlation among insectivores and plant-visitors. Even the distantly related *M. lucifugus* and *P. parnellii* were included in this cluster (fig. 3A). *Hsunnycteris thomasi* slightly deviated from the other plant-visitors, albeit this deviation was not as extreme as *T. cirrhosus* and *D. rotundus*. Morphological and genetic data suggest nectar-feeding derived independently in Glossophaginae and Lonchophyllinae (Datzmann et al. 2010). This independence may also have been captured by our PCA. Given overall similarity of expression profiles between plant-visitors and

insectivores, few genes were differentially expressed (fig. 3B), but more genes were differentially expressed when we examined just the nectarivores (fig. 3C) and many of these proteins were in amino acid metabolic pathways. The strength of differential expression among genes in amino acid pathways was not enough to be significantly enriched, but given carbohydrates are in higher abundance in plant-visitor diets, amino acid synthesis is linked to the products of glycolysis and the citric acid cycle, it is plausible this group has modified these pathways as adaptive responses.

Conclusions

We examined the evolutionary history of genes expressed in SMGs to test for links between genetic variation and signatures of selection. Given the ecological variation of these species, we expected to find strong selection signals and greater expression diversity. Indeed, we identified a strong selection signal among immune-related genes in phyllostomids, but lineage-specific adaptations were less clear, that is, fewer genes under selection among a wide array of pathways and few differentially expressed genes. From our data, we inferred modifications of the endomembrane system with a focus on the Golgi body, most apparent in the vampire bat. Further, lineage-specific adaptations have occurred in response to insulin changes and modifications of metabolic pathways in the plant-visitors, signaling unique, and lineage-specific adaptations have occurred in phyllostomids with diverse feeding strategies.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank Robert Baker and Carleton Phillips for previous work that inspired this research. This work was not possible without the previous efforts of TTU students, faculty, and staff toward sample collection and curation. Therefore, we are thankful for Nicté Ordóñez-Garza, Maria Sagot, Heath Garner, Robert Bradley, and the Texas Tech University Natural Science Research Laboratory (TTU NSRL) for tissue loans. We also acknowledge the High Performance Computing Center (HPCC) at Texas Tech University at Lubbock for providing HPC resources. URL: <http://cmsdev.ttu.edu/hpcc>. C.G.S.C. was supported with a fellowship (PDE) by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and by a postdoctoral fellowship (PNPD) from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil during the development of this study. These funding agencies had no role in any experimental aspect of this study.

Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.
- Baker RJ, Solari S, Cirranello A, Simmons NB. 2016. Higher level classification of phyllostomid bats with a summary of DNA synapomorphies. *Acta Chiropterol.* 18(1):1–38.
- Bloss J, Acree TE, Bloss JM, Hood WR, Kunz TH. 2002. Potential use of chemical cues for colony-mate recognition in the big brown bat, *Eptesicus fuscus*. *J Chem Ecol.* 28(4):819–834.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. 2006. Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev.* 19(3):531–545.
- Cirranello A, Simmons NB, Solari S, Baker RJ. 2016. Morphological diagnoses of higher-level phyllostomid taxa (Chiroptera: Phyllostomidae). *Acta Chiropterol.* 18(1):39–71.
- Datzmann T, von Helversen O, Mayer F. 2010. Evolution of nectarivory in phyllostomid bats (Phyllostomidae Gray, 1825, Chiroptera: Mammalia). *BMC Evol Biol.* 10(1):165.
- Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet.* 46(1):677–700.
- Dimmer EC, et al. 2012. The UniProt-GO Annotation database in 2011. *Nucleic Acids Res.* 40(D1):D565–D570.
- Dobrosielski-Vergona K. 1993. *Biology of the salivary glands*. Boca Raton: CRC Press.
- Dumont ER, et al. 2012. Morphological innovation, diversification and invasion of a new adaptive zone. *Proc R Soc B.* 279(1734):1797–1805.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol.* 7(10):e1002195.
- Fábián TK, Hermann P, Beck A, Fejérdy P, Fábián G. 2012. Salivary defense proteins: their network and role in innate and acquired oral immunity. *Int J Mol Sci.* 13(4):4295–4320.
- Finn RD, et al. 2010. The Pfam protein families database. *Nucleic Acids Res.* 38(Suppl 1):D211–D222.
- Francischetti IMB, et al. 2013. The “Vampirome”: transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat *Desmodus rotundus*, a vector of human rabies. *J Proteomics.* 82:288–319.
- Fujiwara Y, et al. 2007. Insulin hypersensitivity in mice lacking the V1b vasopressin receptor. *J Physiol.* 584(1):235–244.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11:725–736.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
- Haas BJ, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc.* 8(8):1494–1512.
- Hawkins JA, et al. 2019. A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species. *Proc Natl Acad Sci U S A.* 116(23):11351–11360.
- Hayden S, et al. 2014. A cluster of olfactory receptor genes linked to frugivory in bats. *Mol Biol Evol.* 31(4):917–927.
- Hedrick BP, et al. 2019. Morphological diversification under high integration in a hyper diverse mammal clade. *J Mamm Evol.* 1–13. Available from: <https://doi.org/10.1007/s10914-019-09472-x>
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.
- Ibáñez C, Juste J, García-Mudarra JL, Agirre-Mendi PT. 2001. Bat predation on nocturnally migrating birds. *Proc Natl Acad Sci U S A.* 98(17):9700–9702.
- Jiggins FM, Kim KW. 2007. A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol.* 20(3):965–970.
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kosiol C, et al. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4(8):e1000144.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305(3):567–580.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):R25.
- Laska M. 1990. Food transit times and carbohydrate use in three phyllostomid bat species. *Z Säugetierkd.* 55:49–54.
- Lehto M, et al. 2005. Targeting of OSBP-related protein 3 (ORP3) to endoplasmic reticulum and plasma membrane is controlled by multiple determinants. *Exp Cell Res.* 310(2):445–462.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Liu Y, et al. 2013. SPLUNC1/BPIFA1 contributes to pulmonary host defense against *Klebsiella pneumoniae* respiratory infection. *Am J Pathol.* 182(5):1519–1531.
- Liu Y, Xu W, Zhai T, You J, Chen Y. 2019. Silibinin ameliorates hepatic lipid accumulation and oxidative stress in mice with non-alcoholic steatohepatitis by regulating CFLAR-JNK pathway. *Acta Pharm Sin B.* 9(4):745–757.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):1–21.
- Matthijs G, et al. 1997. Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type 1 syndrome (Jaeken syndrome). *Nat Genet.* 16(1):88–92.
- Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 44(D1):D336–D342.
- Monteiro LR, Nogueira MR. 2011. Evolutionary patterns and processes in the radiation of phyllostomid bats. *BMC Evol Biol.* 11:137.
- Ng MM, Dippold HC, Buschman MD, Noakes CJ, Field SJ. 2013. GOLPH3L antagonizes GOLPH3 to determine Golgi morphology. *Mol Biol Cell.* 24(6):796–808.
- Olival KJ, et al. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546(7660):646–650.
- Pécheur EI, et al. 2002. Phospholipid species act as modulators in p97/p47-mediated fusion of Golgi membranes. *Biochemistry* 41(31):9813–9823.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8(10):785–786.
- Phillips CD, Baker RJ. 2015. Secretory gene recruitments in vampire bat salivary adaptation and potential convergences with sanguivorous leeches. *Front Ecol Evol.* 3:122.
- Phillips CJ, et al. 2014. Dietary and flight energetic adaptations in a salivary gland transcriptome of an insectivorous bat. *PLoS One* 9(1):e83512.
- Phillips CJ, Tandler B. 1996. Salivary glands, cellular evolution, and adaptive radiation in mammals. *Eur J Morphol.* 34(3):155–161.
- Phillips CJ, Weiss A, Tandler B. 1998. Plasticity and patterns of evolution in mammalian salivary glands: comparative immunohistochemistry of lysozyme in bats. *Eur J Morphol.* 36:19–26.
- Platt RN, et al. 2014. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol Biol Evol.* 31(6):1536–1545.

- Polge C, Attaix D, Taillandier D. 2015. Role of E2-Ub-conjugating enzymes during skeletal muscle atrophy. *Front Physiol.* 6:59.
- Pritham EJ, Feschotte C. 2007. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc Natl Acad Sci U S A.* 104(6):1895–1900.
- Prokopovic V, et al. 2014. Isolation, biochemical characterization and anti-bacterial activity of BPIFA2 protein. *Arch Oral Biol.* 59(3):302–309.
- Protzek AOP, et al. 2010. Insulin and glucose sensitivity, insulin secretion and β -cell distribution in endocrine pancreas of the fruit bat *Artibeus lituratus*. *Comp Biochem Physiol A Mol Integr Physiol.* 157(2):142–148.
- Ray DA, Pagan HJT, Thompson ML, Stevens RD. 2006. Bats with hATs: evidence for recent DNA transposon activity in genus *Myotis*. *Mol Biol Evol.* 24(3):632–639.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3(2):217–223.
- Rojas D, Vale Á, Ferrero V, Navarro L. 2011. When did plants become important to leaf-nosed bats? Diversification of feeding habits in the family Phyllostomidae. *Mol Ecol.* 20(10):2217–2228.
- Rojas D, Warsi OM, Dávalos LM. 2016. Bats (Chiroptera: Noctilionoidea) challenge a recent origin of extant neotropical diversity. *Syst Biol.* 65(3):432–448.
- Rossoni DM, Assis APA, Giannini NP, Marroig G. 2017. Intense natural selection preceded the invasion of new adaptive zones during the radiation of New World leaf-nosed bats. *Sci Rep.* 7(1):11076.
- Roux J, et al. 2014. Patterns of positive selection in seven ant genomes. *Mol Biol Evol.* 31(7):1661–1685.
- Safi K, Kerth G. 2003. Secretions of the interaural gland contain information about individuality and colony membership in the Bechstein's bat. *Anim Behav.* 65(2):363–369.
- Shen Y-Y, et al. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc Natl Acad Sci U S A.* 107(19):8666–8671.
- Smith I, Wang L-F. 2013. Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr Opin Virol.* 3(1):84–91.
- Sotero-Caio CG, et al. 2015. Integration of molecular cytogenetics, dated molecular phylogeny, and model-based predictions to understand the extreme chromosome reorganization in the neotropical genus *Tonatia* (Chiroptera: Phyllostomidae). *BMC Evol Biol.* 15:220.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Storey JD. 2020. qvalue: Q-value estimating for false discovery rate control. R package version 2.14.0. Available from: <http://github.com/jdstorey/qvalue>. Accessed July 27, 2020.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 100(16):9440–9445.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.
- Talley HM, Laukaitis CM, Karn RC. 2001. Female preference for male saliva: implications for sexual isolation of *Mus musculus* subspecies. *Evolution* 55(3):631–634.
- Tandler B, Gresik EW, Nagato T, Phillips CJ. 2001. Secretion by striated ducts of mammalian major salivary glands: review from an ultrastructural, functional, and evolutionary perspective. *Anat Rec.* 264(2):121–145.
- Tandler B, Phillips CJ. 1998. Microstructure of mammalian salivary glands and its relationship to diet. *Gland Mech Salivary Secret.* 10:21–35.
- Tuttle MD, Ryan MJ. 1981. Bat predation and the evolution of frog vocalizations in the neotropics. *Science* 214(4521):677–678.
- van der Lee R, Wiel L, van Dam TJP, Huynen MA. 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res.* 45(18):10634–10648.
- Vandewege MW, Phillips CJ, Wickliffe JK, Hoffmann FG. 2013. Evolution of the ABPA subunit of androgen-binding protein expressed in the submaxillary glands in New and Old World rodent taxa. *J Mol Evol.* 76(5):324–331.
- Viljakainen L, et al. 2009. Rapid evolution of immune proteins in social insects. *Mol Biol Evol.* 26(8):1791–1801.
- Wang PX, et al. 2017. Targeting CASP8 and FADD-like apoptosis regulator ameliorates nonalcoholic steatohepatitis in mice and nonhuman primates. *Nat Med.* 23(4):439–449.
- Wing SS, Banville D. 1994. 14-kDa ubiquitin-conjugating enzyme: structure of the rat gene and regulation upon fasting and by insulin. *Am J Physiol Endocrinol Metab.* 267:E39–E48.
- Xiao J, et al. 2015. Transcriptome analysis revealed positive selection of immune-related genes in tilapia. *Fish Shellfish Immunol.* 44(1):60–65.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15(5):568–573.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19(6):908–917.
- Yassine HN, et al. 2009. Effects of exercise and caloric restriction on insulin resistance and cardiometabolic risk factors in older obese adults – a randomized clinical trial. *J Gerontol A Biol Sci Med Sci* 64A(1):90–95.
- Zhang G, et al. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339(6118):456–460.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22(12):2472–2479.

Associate editor: Naruya Saitou