A genomic hotspot of diversifying selection and structural change in the hoary bat (*Lasiurus cinereus*)

Robert S. Cornman

U.S. Geological Survey, Fort Collins Science Center, Fort Collins, Colorado, United States

ABSTRACT

Background: Previous work found that numerous genes positively selected within the hoary bat (*Lasiurus cinereus*) lineage are physically clustered in regions of conserved synteny. Here I further validate and expand on those finding utilizing an updated *L. cinereus* genome assembly and additional bat species as well as other tetrapod outgroups.

Methods: A chromosome-level assembly was generated by chromatin-contact mapping and made available by DNAZoo (www.dnazoo.org). The genomic organization of orthologous genes was extracted from annotation data for multiple additional bat species as well as other tetrapod clades for which chromosome-level assemblies were available from the National Center for Biotechnology Information (NCBI). Tests of branch-specific positive selection were performed for *L. cinereus* using PAML as well as with the HyPhy package for comparison.

Results: Twelve genes exhibiting significant diversifying selection in the *L. cinereus* lineage were clustered within a 12-Mb genomic window; one of these (*Trpc4*) also exhibited diversifying selection in bats generally. Ten of the 12 genes are landmarks of two distinct blocks of ancient synteny that are not linked in other tetrapod clades. Bats are further distinguished by frequent structural rearrangements within these synteny blocks, which are rarely observed in other Tetrapoda. Patterns of gene order and orientation among bat taxa are incompatible with phylogeny as presently understood, implying parallel evolution or subsequent reversals. Inferences of positive selection were found to be robust to alternative phylogenetic topologies as well as a strong shift in background nucleotide composition in some taxa.

Discussion: This study confirms and further localizes a genomic hotspot of protein-coding divergence in the hoary bat, one that also exhibits an increased tempo of structural change in bats compared with other mammals. Most genes in the two synteny blocks have elevated expression in brain tissue in humans and model organisms, and genetic studies implicate the selected genes in cranial and neurological development, among other functions.

Subjects Evolutionary Studies, Genomics, Zoology
Keywords Hoary bat, Evolutionary rate analysis, Comparative genomics, Adaptation

INTRODUCTION

As high-quality genome references with well-supported annotations continue to be released, it has become increasingly common to compare the evolution of genomic features

within a clade, often in conjunction with ecological, life-history, or morphological data (*Lewin et al., 2018*). These comparative analyses can illuminate the genomic architecture of trait evolution, bringing to bear diverse data types such as cytogenetics, gene family evolution, evolutionary rates of orthologous genes, gene expression patterns, and methylation profiles (*Wei et al., 2002*). Protein evolutionary rates are among the most accessible and widely applied comparisons, and several tests have been proposed to identify positive diversifying selection within a gene tree (*Muse & Gaut, 1994*; *Yang & Nielsen, 2000*; *Seo, Kishino & Thorne, 2004*; *Smith et al., 2015*). As nonsynonymous substitution rates statistically higher than synonymous rates are only expected under adaptive evolution, ratio-based tests can confirm hypotheses of diversifying selection on candidate genes relevant to fitness (*Aguileta et al., 2009*) as well as discover episodes of selection agnostically by scanning whole genomes (*Heger & Ponting, 2007*; *Kosiol et al., 2008*). In addition to revealing mechanisms of adaptation, genomic surveys of substitution patterns can be important in other contexts such as evolutionary medicine, *e.g.* by quantifying levels of protein constraint (*Lindblad-Toh et al., 2011*) and estimating the functional significance of mutations (*Cingolani et al., 2012*).

A number of studies have investigated rates of protein evolution within the various orders of mammals, some of which (*Shen et al., 2010*; *Hawkins et al., 2019*; *Cornman & Cryan, 2022*, but see *Jebb et al., 2020*) have indicated high rates of positive selection on coding sequence in bats (order Chiroptera). A recent study (*Cornman & Cryan, 2022*) identified candidates of positive selection in the hoary bat lineage, including genes clustered in regions of conserved synteny (discrete regions in which ancestral gene sets are colinear in comparative genomic alignments despite extensive background divergence in genome content and organization). For example, six genes showed evidence of positive selection in the vicinity of the cat-eye critical region, which overlaps a conserved synteny block in tetrapods, so-named for its association with the 'cat-eye' spectrum of congenital developmental disorders in human (*Footz et al., 2001*). In addition to containing clustered signatures of positive selection, this region was structurally divergent in the hoary bat lineage as well, including rearrangements not apparent in other mammalian clades (*Cornman & Cryan, 2022*). Furthermore, several other genes associated with cranial dysmorphy in humans were selection candidates in *L. cinereus*, collectively indicating that cranial development may have been a phenotypic target of selection in the divergence of the hoary bat lineage (*Cornman & Cryan, 2022*). This hypothesis is consistent with previous morphometric studies that have documented associations between cranial morphology and ecological divergence among related bat species and at higher taxonomic levels in bats (*Evin et al., 2008*; *Hedrick & Dumont, 2018*; *Arbour, Curtis & Santana, 2021*).

While the hoary bat assembly analyzed by *Cornman & Cryan (2022)* had a high scaffold N50 (35.1 Mb; *Cornman et al., 2021*), contiguity was not at chromosomal level, such that clustering of selection candidates and conservation of synteny could not be fully evaluated. Few other chromosome-scale assemblies of bat genomes were available at that time, preventing a reconstruction of the sequence of structural changes. However, a recent scaffolding effort with "Hi-C" chromatin mapping (*Belton et al., 2012*) has revealed that two of the three clusters discussed by *Cornman & Cryan (2022)* form a contiguous block

on a large metacentric chromosome. Additional high-quality genomes of bat species have also become available since that study, which enable improved sampling of gene trees and thus increased statistical power of selection tests within bats, as well as improved comparisons of synteny. Here I report that expanded ortholog alignments strengthen inferences of diversifying selection within the hoary bat lineage (*i.e.*, since the divergence of genus *Lasiurus* from other analyzed genera). Co-localization of the two synteny blocks was not observed in a survey of other tetrapod orders and appears not to be the ancestral state of Chiroptera, yet has an apparently homoplasious distribution within this group, implying parallel evolution of complex features or a similarly complex reversal. Numerous smaller changes in gene order are found within these synteny blocks as well, particularly within suborder Yangochiroptera.

MATERIALS AND METHODS

The analysis of *Cornman & Cryan (2022)* was based on a 10X linked-read assembly (accession GCA_011751065.1 of the National Center for Biotechnology Information (NCBI), *Cornman et al., 2021*). This was subsequently scaffolded by the DNAZoo Consortium (*Dudchenko et al., 2017*) using a HiC chromatin contact-mapping data set (NCBI accession SRX8933264) and made publicly available at https://www.dnazoo.org/assemblies/Aeorestes_cinereus. (Note that *Aeroestes cinereus* and *Lasiurus cinereus* are synonymous; we follow the latter usage here as it predominates in the literature.) To evaluate assembly coverage patterns, confirm the X chromosome, and detect assembly artifacts such as collapsed repeats, an existing genomic short-read data set for *L. cinereus* (PRJNA559902; *Pinzari et al., 2020*) was aligned to the genome assembly with bowtie2 v. 2.4.5 (*Langmead & Salzberg, 2012*) using the "fast" and "end-to-end" parameter switches, then filtered with SAMtools v. 1.12 (*Li et al., 2009*) at a mapping quality of 30 (Phred-scaled). Mapped reads were summed per major scaffold using the bedcov function of SAMtools, then divided by chromosome length as well as the total number of reads mapped per sample to those scaffolds. The resulting values were then normalized to the average of all sample-scaffold pairs.

Coding-sequence alignments were generated from precomputed orthologs sets for each gene, available from NCBI *via* the corresponding gene pages, with the exception that *L. cinereus* orthologs were taken from *Cornman & Cryan (2022)*. Some transcripts in the pre-computed ortholog groups differed structurally from the others due to alternative splicing, in which case substituting a different isoform of the same gene often sufficed to correct the alignment. In other cases, un-annotated exons could be extracted by searching genomic sequence with coding sequence of a related taxon. Rarely, the coding sequence of a congener was used when available and substantially more complete. If none of these alternatives yielded at least a partial coding sequence for a genus, that genus was deleted from the guide tree and evolutionary rates were computed from the remaining taxa. If multiple taxa were unavailable, the gene was not analyzed. Ortholog sets were aligned at the nucleotide level with MAFFT v. 7.480 (*Katoh et al., 2002*), trimmed of untranslated regions, and realigned at the protein level. Low-complexity or gapped regions for which

codon-level orthology was questionable were deleted from alignments between unambiguously conserved codons. Analyzed ortholog alignments are available in File S1.

The guide tree for all evolutionary rate analyses reported in the Results follows the phylogenetic analysis of *Amador et al. (2018)*, however the consistency of results was qualitatively assessed by examining three other tree topologies as well (Fig. S1). For example, *Agnarsson et al. (2011)* supported a closer relationship of *Lasiurus* and *Pipistrellus*, with *Eptesicus* an outgroup to the pair, which is also consistent with overall karyotype (*Bickham, 1987*). This alternative topology is labeled "tree 2" in Fig. S1. A third topology was tested in which the location of *Miniopterus* follows *Agnarsson et al. (2011)*, as sister to phyllostomids rather than vespertilionids ("tree 3"). A fourth topology ("tree 4") is a pruned version of tree 1, in which taxa with greatly increased GC content of some tested genes (see Results) were removed to ensure that this shift in nucleotide compositions did not influence the conclusions drawn.

Parameters and likelihoods of two models were estimated with PAML (*Yang, 2007*). Setting the control variables "Model = 0" and "NSsites = 0" estimates one rate class ω across the phylogeny, whereas setting the control variables "Model = 2" and "NSsites = 0" and labeling the *L. cinereus* branch as foreground estimates two ω values, one for background taxa and one for the foreground taxon. The test statistic for branch-specific positive selection was then calculated as twice the difference in log-likelihood between the latter and the former, assuming a χ^2 distribution with one degree of freedom (*Yang, 2007*). False discovery rate (FDR) correction was performed with the Benjamini-Hochberg procedure of the R function p.adjust (*R Core Team R, 2018*). The aBSREL and BUSTED programs of the HyPhy package (*Kosakovsky Pond et al., 2020*) were also used to test for positive selection in order to evaluate overall consistency of inferences. False discovery was performed for the set of 34 tested genes (see Results), but separately for each tree topology and for each statistical test. This is because only the PAML results for tree 1 are reported in the Results; the other methods and topologies are reported for qualitative comparison only. FDR correction of aBSREL *P*-values were performed on the uncorrected values for the foreground branch only (*L. cinereus*). Model outputs are summarized in File S2.

Ortholog locations were extracted for each gene from the gene feature files accompanying each annotated genome downloaded from NCBI. Locations of *L. cinereus* genes were updated by splice-aware alignment of transcripts to the revised genome assembly with GMAP v. 2023-12-01 using default settings (*Wu & Watanabe, 2005*). Clustering of the original selection candidates from *Cornman & Cryan (2022)* was re-evaluated for the updated assembly in two ways. First, the cumulative proportions of selected and total genes were plotted in consecutive 1-Mb windows based on midpoint coordinate. Secondly, a null distribution for the expected number of selected genes in comparable windows was generated by selecting 1,000 random 12-Mb windows that were permitted to cross chromosome boundaries, with chromosomes concatenated in a random order for each iteration.

Bat "RefSeq" genome accessions generated by NCBI that were used in evolutionary rate and synteny analyses include *Eptesicus fuscus* (GCA_027574615.1; *Paulat et al., 2023*), *Pipistrellus kuhlii* (GCF_014108245.1; *Jebb et al., 2020*), *Myotis myotis* (GCF_014108235.1;

Jebb et al., 2020), *Molossus molossus* (GCF_014108415.1; *Jebb et al., 2020*), *Artibeus jamaicensis* (GCF_021234435.1; *Wang et al., 2020*), *Phyllostomus discolor* (GCF_004126475.2; *Jebb et al., 2020*), *Rousettus aegyptiacus* (GCF_014176215.1; *Jebb et al., 2020*), and *Rhinolophus ferrumequinum* (GCF_004115265.2; *Jebb et al., 2020*). Bat genomes used for evolutionary rate analysis only include *Sturnira hondurensis* (GCF_014824575.3; *Wang et al., 2020*), *Pteropus alecto* (GCF_000325575.1; *Zhang et al., 2013*), *Miniopterus natalensis* (GCF_001595765.1; *Eckalbar et al., 2016*), and *Hipposideros armiger* (GCF_001890085.2; *Dong et al., 2016*). Genome accessions of outgroup tetrapods used in synteny analyses included *Homo sapiens* (GCF_000001405.40; *International Human Genome Sequencing Consortium, 2001*), *Mus musculus* (GCF_000001635.27; *Church et al., 2009*), *Corvus cornix* (GCF_000738735.6; *Poelstra et al., 2014*), *Monodelphis domestica* (GCF_027887165.1; *Mikkelsen et al., 2007*), *Ornithorhynchus anatinus* (GCF_004115215.2; *Zhou et al., 2021*), *Ochotona princeps* (GCF_030435755.1; *Sjodin et al., 2021*), *Sus scrofa* (GCF_000003025.6; *Warr et al., 2020*), *Bos taurus* (GCF_002263795.3; *Rosen et al., 2020*), *Equus caballus* (GCF_002863925.1; *Kalbfleisch et al., 2018*), and *Felis catus* (GCF_018350175.1; *Pontius et al., 2007*).
Syntenic regions were evaluated using pre-computed data tracks of the University of California, Santa Cruz (UCSC) Genome Browser (*Lee et al., 2020*), including Gencode (*Frankish et al., 2021*) and RefSeq (*O'Leary et al., 2016*) annotation data for human.

Sequence analysis was performed to investigate the evolution of a set of genes with partial homology to the transcription factor *Tbx1* (see Results). Manipulation and visualization of these sequence alignments was performed with BioEdit (*Hall, 1999*), whereas protein secondary structure was predicted with Jpred4 (*Drozdetskiy et al., 2015*) and core promoter sequences identified with the neural network tool of *Reese (2001)*.

Gene expression data were tabulated from two sources. Summaries of tissue-level expression in human and mouse orthologs were extracted from their NCBI Gene pages and are derived from *Fagerberg et al. (2014)* and *Yue et al. (2014)*, respectively. Tissues of elevated expression for each gene were obtained from the "UP_Table" expression output of the DAVID functional annotation tool. Gene ontology enrichment was performed with the AmiGO 2 web service (*Ashburner et al., 2000*; *The Gene Ontology Consortium et al., 2023*) using official gene symbols as input and the GO-Slim ontology terms for biological process.

Primary sequence data underlying these analyses are also available in a U.S. Geological Survey data release (*Cornman, 2024*).

RESULTS AND DISCUSSION

The updated *L. cinereus* genome assembly is comprised of 14 major scaffolds totaling 2.08 Gb, encompassing 98.9% of the total assembly length of 2.11 Gb. The assembly is similar in length to the chromosome-level assembly of the vespertilionid species *E. fuscus* (2.01 Gb) and matches the *L. cinereus* karyotype (*Bickham, 1987*) in terms of chromosome number and their relative lengths (Fig. S2). Among-individual coverage patterns in a population sample were consistent for all major scaffolds except Scaffold 10, which exhibited a bimodal coverage pattern unambiguously indicating the X chromosome. No scaffold was identified with coverage patterns indicative of the Y chromosome, as expected given that

the sequenced individual was female. Although not relevant to the present study, the smallest scaffolds were consistently undersampled in the Illumina short-read libraries used for coverage assessment (Fig. S2), an unanticipated effect which could potentially bias population-genomic analyses in this or related species.

To re-evaluate the clustering of positive selection candidates in this new assembly, the 9,447 tested single-copy orthologs from *Cornman & Cryan (2022)* were re-aligned, with all but two placed on one of the 14 major scaffolds. The cumulative distribution of positive selection candidates diverges from all tested genes (Fig. 1A), particularly in the region of the synteny blocks described below, which encompasses ten such genes within a 12-Mb span. No randomly resampled 12-Mb windows selected from randomly concatenated chromosomes contained more than ten positive selection candidates, and 99.9% of resampled windows were less than this value (Fig. 1B).

The ten clustered selection candidates identified above combine two of three clusters identified in *Cornman & Cryan (2022)*, which were on separate scaffolds of that assembly. They include four genes of the 'cat-eye' (CE) synteny block (*Cecr2*, *Cecr6*, *Mical3*, and *Slc25a18*) and three genes (*Postn*, *Frem2*, and *Proser1*) that lie within a second block of conserved synteny in tetrapods that was highlighted in *Cornman & Cryan (2022)*. For convenience, I designate this latter synteny block "NF" based on the upstream and downstream genes *Nbea* and *Foxo1* that bound the protein-coding members of the block in human. The CE and NF synteny blocks are operationally defined here based on pre-computed alignments with other taxa (Figs. 2 and 3) available from the UCSC Genome Browser (*Lee et al., 2020*). One positive-selection candidate, *Sacs*, lies within the NF synteny block in most bat species examined but not in mammals generally (further discussed below). Another positive-selection candidate, *Amer3*, lies between these two synteny blocks in *L. cinereus* whereas the final selection candidate is downstream of the CE synteny block (*Fgf9*). Two positive-selection candidates from *Cornman & Cryan (2022)* that also mapped to this region, *Rps13* and *Necap1*, were excluded from this analysis as they are likely retrogenes: the coding sequences of the annotated genes occur on single exons and both have unannotated, multiexon alignments elsewhere in the genome. Note that these synteny-block depictions are necessarily human-centric based on the available data tracks and do not imply that no other genes are present in these regions in other lineages or that humans have retained all genes that were present in these synteny blocks in the common ancestor.

Tests of positive selection in the *L. cinereus* lineage with the expanded set of 13 bat taxa were again significant for all ten of the previously identified selection candidates ($P < 0.01$ after false-discovery correction; Table 1). Given the larger number of taxa available for analysis and the clear relevance of this region to adaptation in *L. cinereus*, I also tested the remaining genes of the CE and NF synteny blocks (Figs. 2 and 3). For these additional tests, *Nbea* and *Trpc4* were also significant for the *L. cinereus* branch after FDR correction at $P < 0.01$ (Table 2). Thus, a total of 12 genes within a 12-Mb genomic window were identified as candidates for positive selection within the *L. cinereus* branch in this updated assessment.
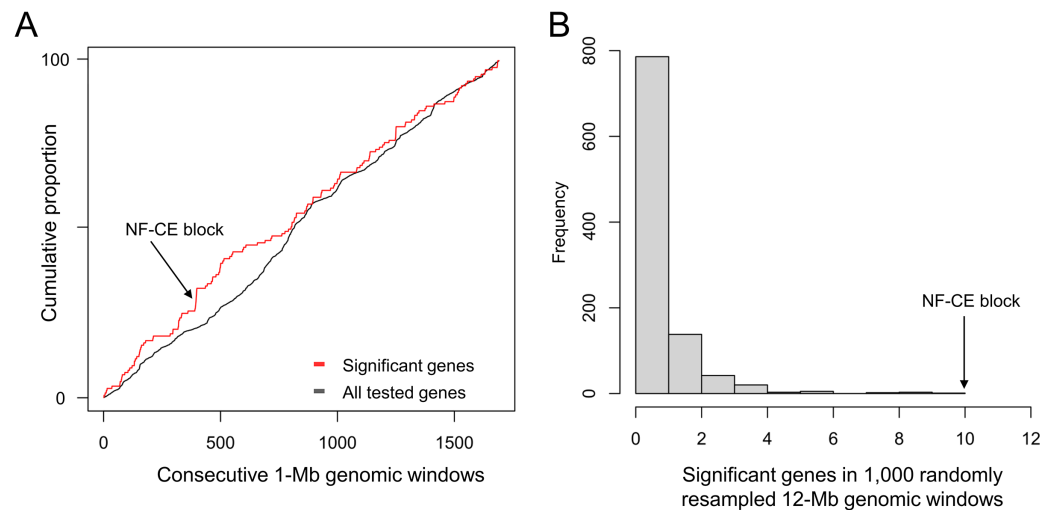
Figure 1 A pronounced cluster of positively selected genes occurs in the chromosome-level assembly of hoary bat (*Lasiurus cinereus*). The genomic region containing the cluster of selection candidates is labeled "NF-CE block", see text for details. (A) Cumulative proportion of all single-copy orthologs tested for positive selection in a previous study (see text for details) compared with the cumulative proportion of genes with significant test results. (B) Histogram of the number of positive selection candidates in windows of comparable size to the 12-Mb region identified in panel A. The observed value for the NF-CE block is ten. Full-size ☑ DOI: 10.7717/peerj.17482/fig-1
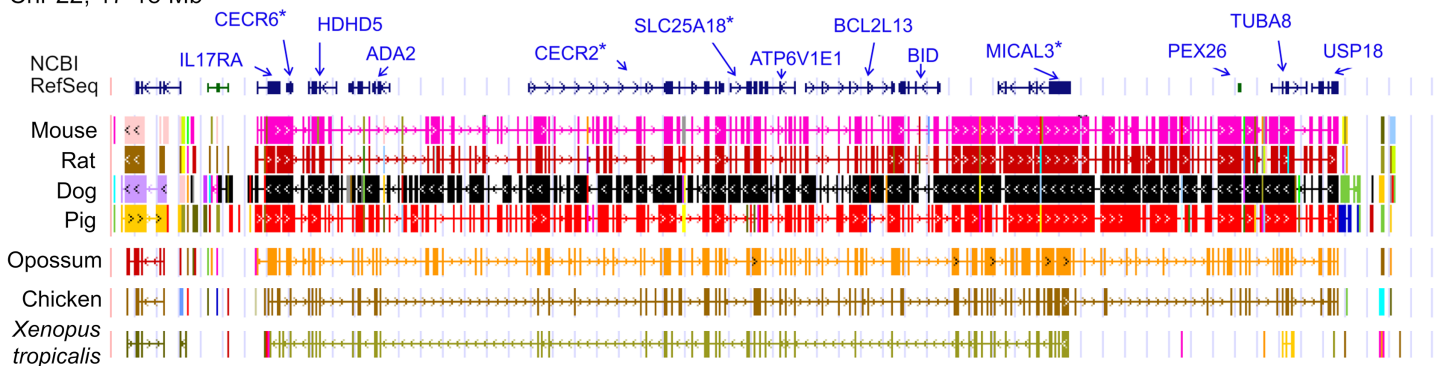

Figure 2 Demarcation of the cat-eye (CE) synteny block based on the gene order in human. Each row is a genomic data track derived from the University of California Santa Cruz (UCSC) Genome Browser, with the approximate location in the human genome indicated in the upper left. The top track shows ideograms of the human RefSeq genes curated by the National Center for Biotechnology Information (NCBI), indicating exon structure and orientation, labeled with the gene symbols used in the text. Asterisks indicate genes with evidence of positive selection within the *Lasiurus cinereus* branch of the tested phylogeny. The subsequent tracks identify blocks of conserved sequence in representative Tetrapoda of increasing evolutionary distance to human. Within each species, alignments on the same chromosome share a common color and are linked by flow lines if contiguous. Regions that are approximately uniform in color and contiguous within each species are syntenic. See Methods for data track sources. Full-size ☑ DOI: 10.7717/peerj.17482/fig-2

None of these 12 alignments were significant for positive selection within bats generally (*i.e.*, testing a site-level model with no branch-specific parameters) after FDR correction (File S2). However, the *Trpc4* alignment was observed to be highly variable in the C-terminal third of the coding sequence within bats overall as well as divergent from the other tetrapod clades in the same region (Fig. 4). In contrast, other tetrapod orders did not
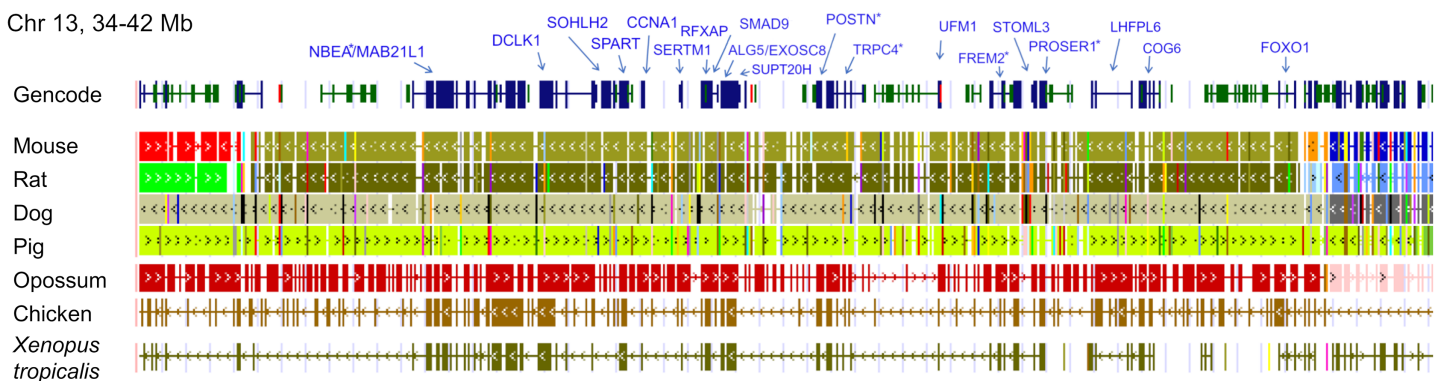
Figure 3 Demarcation of the NF synteny block based on gene order in human. Each row is a genomic data track derived from the University of California Santa Cruz (UCSC) Genome Browser, with the approximate location in the human genome indicated in the upper left. The top track shows ideograms of human genes curated by the Gencode consortium, indicating exon-intron structure and orientation, labeled with the gene symbols used in the text. Asterisks indicate genes with evidence of positive selection within the *Lasiurus cinereus* branch of the tested phylogeny. The subsequent tracks identify blocks of conserved sequence in representative Tetrapoda of increasing evolutionary distance to human. Within each species, alignments on the same chromosome share a common color and are linked by flow lines if contiguous. Regions that are approximately uniform in color and contiguous within each species are syntenic. Not the human gene *Ccdc169*, which lies between *Sohlh2* and *Spart*, is not conserved across mammals and therefore not labeled here. See Methods for data track sources.   Full-size 🖾 DOI: 10.7717/peerj.17482/fig-3

Table 1 Branch-specific tests of evolutionary rate for previously identified positive-selection candidates clustered in the hoary bat genome. Significant *P*-values are bolded. See text for details.

Gene	Synteny block	lnL: Model = 0, NSsites = 0	lnL: Model = 2, NSsites = 0	Adjusted *P*-value	Background ω	Foreground ω
Amer3		−11,606.9213	−11,599.1666	**0.0003**	0.3127	0.7241
Cecr2	CE	−17,895.0499	−17,887.2513	**0.0003**	0.2194	0.4779
Cecr6	CE	−6,567.5161	−6,557.1220	**0.0000**	0.1386	0.4092
Fgf9		−1,467.9643	−1,462.3278	**0.0021**	0.0030	0.3742
Frem2	NF	−41,738.7078	−41,719.1637	**0.0000**	0.1449	0.3532
Mical3	CE	−22,368.7987	−22,362.1366	**0.0008**	0.1672	0.3137
Periostin	NF	−8,688.5241	−8,679.9892	**0.0002**	0.0584	0.2468
Proser1	NF	−11,209.2924	−11,201.2627	**0.0002**	0.2019	0.6130
Sacs		−47,422.3021	−47,411.1321	**0.0000**	0.0832	0.1817
Slc25a18	CE	−2,305.1623	−2,296.6956	**0.0002**	0.1605	2.2338

show increased variation in this region either within or among clades. *Trpc4* encodes a Ca^{2+} transmembrane channel subunit involved in diverse processes, the C-terminal region of which is intracellular and believed to interact with inositol triphosphate receptors (ITPRs) as well as calmodulin (*Tang et al., 2001*). Given the diverse hibernation strategies of bats, it is noteworthy that *Trpc4* is required in mice for heat detection and subsequent thermoregulation by warm-sensing neurons (*Zhou et al. 2023*). I therefore performed additional *post-hoc* tests of positive diversifying selection on *Trpc4* within bats relative to a representative mammalian outgroup, Carnivora. With *Trpc4* sequences from carnivores designated as background and bat sequences designated as foreground, PAML analysis revealed strong evidence of diversifying selection within bats for the entire coding sequence (1.12E-7) and for the ITPR-binding domain alone (3.36E-7). Remarkably,

Gene	Synteny block	lnL: Model = 0, NSsites = 0	lnL: Model = 2, NSsites = 0	Adjusted *P*-value	Background ω	Foreground ω
Alg5	NF	−3,167.7468	−3,167.7380	0.9541	0.1248	0.1395
Atp6v1e1	CE	−1,851.9959	−1,850.9754	0.2882	0.0496	0.0001
Bcl2l13	CE	−4,528.8733	−4,528.8732	0.9922	0.2803	0.2792
Ccna1	NF	−5,422.2964	−5,421.3659	0.3067	0.1379	0.2666
Cog6	NF	−6,466.0667	−6,465.4972	0.3977	0.0531	0.0973
Dclk1	NF	−6,276.2568	−6,274.3244	0.1052	0.0180	0.0547
Exosc8	NF	−2,588.5123	−2,585.7166	0.0413	0.0762	0.3830
Foxo1	NF	−6,126.2541	−6,125.6486	0.3944	0.0572	0.0994
Hdhd5	CE	−4,230.8031	−4,230.7411	0.7997	0.1590	0.1912
Il17ra	CE	−10,733.4234	−10,729.4470	0.0118	0.2544	0.4890
Mab21l1	NF	−1,957.4403	−1,957.3366	0.7415	0.0476	0.0302
Lhfpl6	NF	−2,999.7726	−2,999.5977	0.7094	0.0021	0.0001
Nbea	NF	−28,137.9727	−28,125.8943	**0.0000**	0.0332	0.1234
Pex26	CE	−4,080.0169	−4,079.8760	0.7330	0.1967	0.1444
Nhlrc3	NF	−3,773.5075	−3,773.3928	0.7415	0.3949	0.3239
Rfxap	NF	−2,809.1775	−2,808.5338	0.3909	0.1045	0.1798
Smad9	NF	−4,432.7381	−4,432.0203	0.3694	0.0257	0.0353
Spart	NF	−8,726.5589	−8,725.2287	0.2058	0.1288	0.2327
Supt20h	NF	−7,073.7274	−7,073.2486	0.4371	0.1244	0.1845
Trpc4	NF	−9,860.3126	−9,851.2214	**0.0001**	0.0656	0.2388
Tuba8	CE	−3,897.1940	−3,897.1896	0.9553	0.0144	0.0160
Usp18	CE	−3,896.5612	−3,895.6903	0.3148	0.2748	0.5061

neither BUSTED nor aBSREL supported positive selection in those same alignments. This discrepancy might be attributable to the greater evolutionary distance between bats and carnivores, which could complicate the estimation of background evolutionary rates due to factors such as rate variability and mutational saturation (reviewed in *Joy et al., 2016*). Alternatively, the ITPR region may simply be evolving neutrally within bats, yet this seems unlikely given that truncations and frameshifts are not seen and radical substitutions are infrequent (Fig. 4). Factors such as gene conversion or unrecognized paralogy also do not appear to contribute to the observed *Trpc4* diversity, as BLASTP searches with either the full *P. kuhli* predicted protein or the ITPR region alone did not match more strongly to any other Trpc homolog (File S2). Alterative splicing can also be excluded as a confounding factor as the predicted exon structures are similar to orthologs in other taxa and an outgroup sequence from cat aligned only to predicted exons in *P. kuhlii* (see Fig. S3).

Structural evolution of the CE and NF synteny blocks in bats and tetrapod outgroups

Ten tetrapod taxa including a bird, non-placental mammals, and placental mammals were surveyed to evaluate the ancestral organization of the clustered positive selection
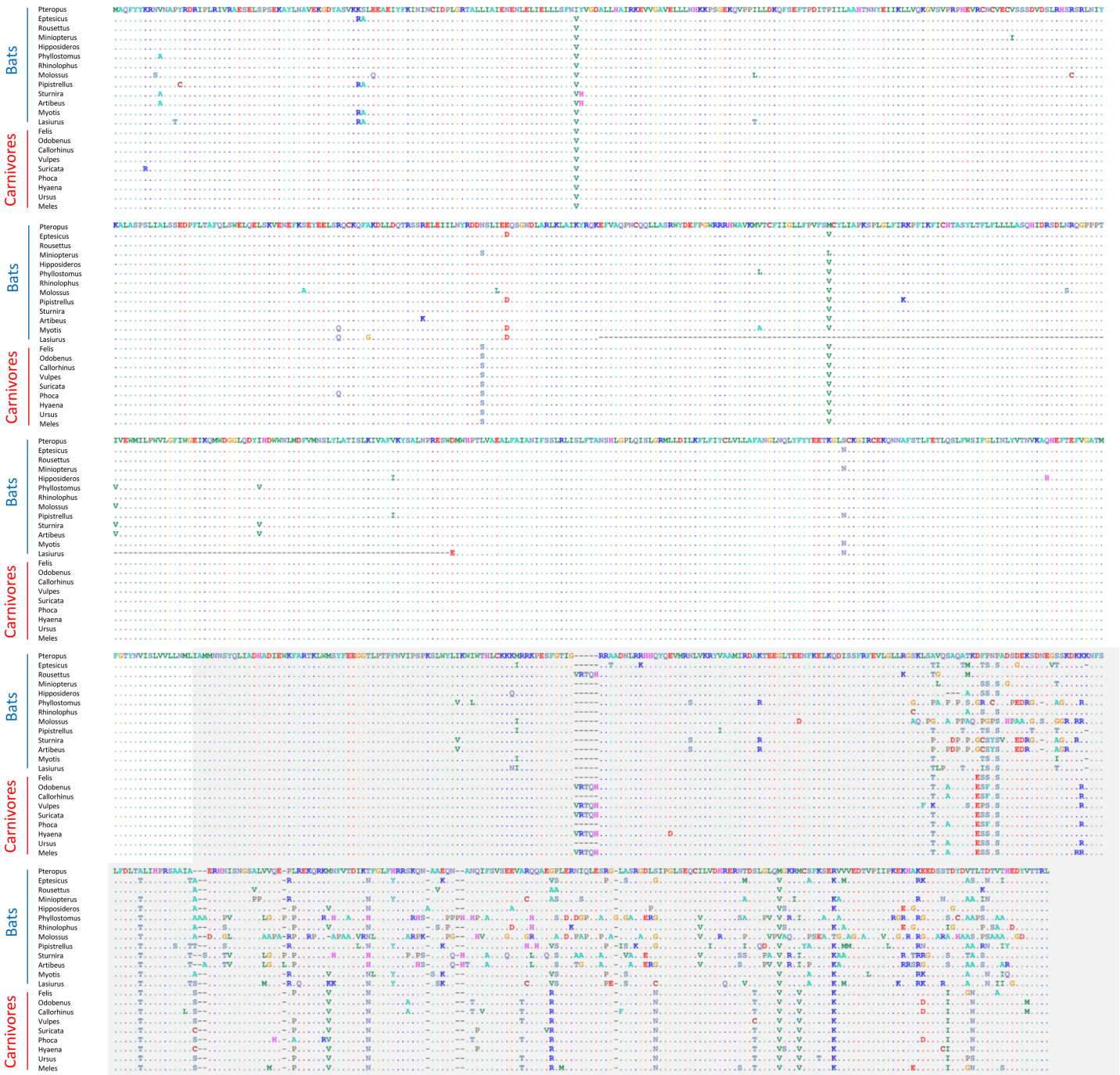
Figure 4 Alignment of predicted protein sequences of the Trpc4 gene in representative Carnivora and Chiroptera. The gray-shaded C-terminal region corresponds to the inositol triphosphate receptor (ITPR) binding region annotated in the human protein and discussed in the text. Residues that are unchanged from the first sequence in the alignment are represented by a dot to better highlight variable positions. Dashes indicate missing sequence. The alignment wraps to each row of the figure, position numbers are not shown for clarity.   Full-size ◫ DOI: 10.7717/peerj.17482/fig-4

candidates and the tempo of structural change in their vicinity during tetrapod evolution (File S3), a representative subset of which is shown in Fig. 5. Some structural variation in the CE block is seen in the *B. taurus* and *O. anatinus* genomes and individual gene

deletions have occurred (*e.g.*, *Ada2* in *M. musculus* and *Slc25a18* in *S. scofra* and *B. taurus*). However, altered gene order or orientation was not observed in the NF block nor is it linked with the CE block in any outgroup genome. The bats *R. aegypticus* and *R. ferrumiquinum*, representatives of suborder Yinpterochiroptera (*Agnarsson et al., 2011*; *Amador et al., 2018*), have genomic architectures (Figs. 5 and 6) similar to the other tetrapod clades, but differ in that *Lhflp6* and *Cog6* are far removed from other NF genes on the same linkage group. The positively selected gene *Amer3* does not occur near either the NF or CE synteny blocks in other tetrapods but is linked to CE genes in all bats examined. The positively selected genes *Sacs* and *Fgf9* have a conserved relative orientation in outgroups and while usually linked to the NF synteny block they never lie within either synteny block (unlike in bats).

Within Yangochiroptera (Fig. 6 and File S3), multiple distinct arrangements are evident within and among the CE and NF synteny blocks, as well as the positively selected genes *Sacs*, *Fgf9*, and *Amer3*. A common arrangement is shared between *P. discolor* and *A. jamaicensis*, of family Phyllostomidae (Yangochiroptera), in which the NF and CE blocks are tightly linked, the NF block is further rearranged in gene order, *Amer3* lies between *Il17ra* and *Nhlrc3*, and *Fgf9* is effectively unlinked from *Sacs* and genes of both synteny blocks. Organization of these gene regions is more variable within the four vesper bats examined, such that a single unambiguous sequence of chromosomal rearrangements is not apparent without homoplasy (Fig. 7 illustrates one possible reconstruction of evolutionary events, inferred by inspection). The vespertilionid species share at least partial linkage of NF and CE genes (Fig. 6, File S3), but in arrangements distinct from phyllostomids. In *E. fuscus*, only the CE genes *Il17ra*, *Cecr6*, and *Hdhd5* are linked with NF, with the remainder on a separate linkage group, apparently due to a chromosomal fission. *Amer3* again lies between the two synteny blocks, but in two novel arrangements in vesper bats. *M. myotis* exhibits a unique integration of the *Il17ra-Cecr2-Hdhd5* trio and *Amer3* within the NF block. *P. kuhlii* and *L. cinereus* are more similar in gene order compared with other vesper bats examined, but differ in that *Ufm1* of the NF synteny block is on a different scaffold in *L. cinereus* (it also has a distinct location in *E. fuscus*) and *Amer3* is in the same relative order but inverted in *L. cinereus*. Note that the ideograms in Figs. 5 and 6 are oriented with the NF block first and the CE block second for consistency, since the plus-strand designation of each linkage group is arbitrary (hence *P. kuhlii* and *L. cinereus* coordinates are descending and ascending, respectively, on the main linkage groups; see File S3 for genomic accessions and coordinates).

The genomic organization in *M. molossus*, representative of the free-tailed bat family (Molossidae), reveals several changes from Yinpterochiroptera that are shared with phyllostomid and vespertilionid bats and may be basal to Yangochiroptera: 1) the insertion of *Sacs* within the NF synteny block; 2) deletion of *Stoml3*; and 3) a rearrangement of the NF genes *Lhfpl6-Foxo1*. However, the consensus view that Molossidae is within the superfamily Vespertilionoidea (*Agnarsson et al., 2011*; *Amador et al., 2018*) and shares a more recent common ancestor with vespertilionids than with phyllostomids is incongruent with aspects of Fig. 6. Most notably, this phylogenetic position implies that either the NF and CE blocks merged independently in vespertilionid and phyllostomid
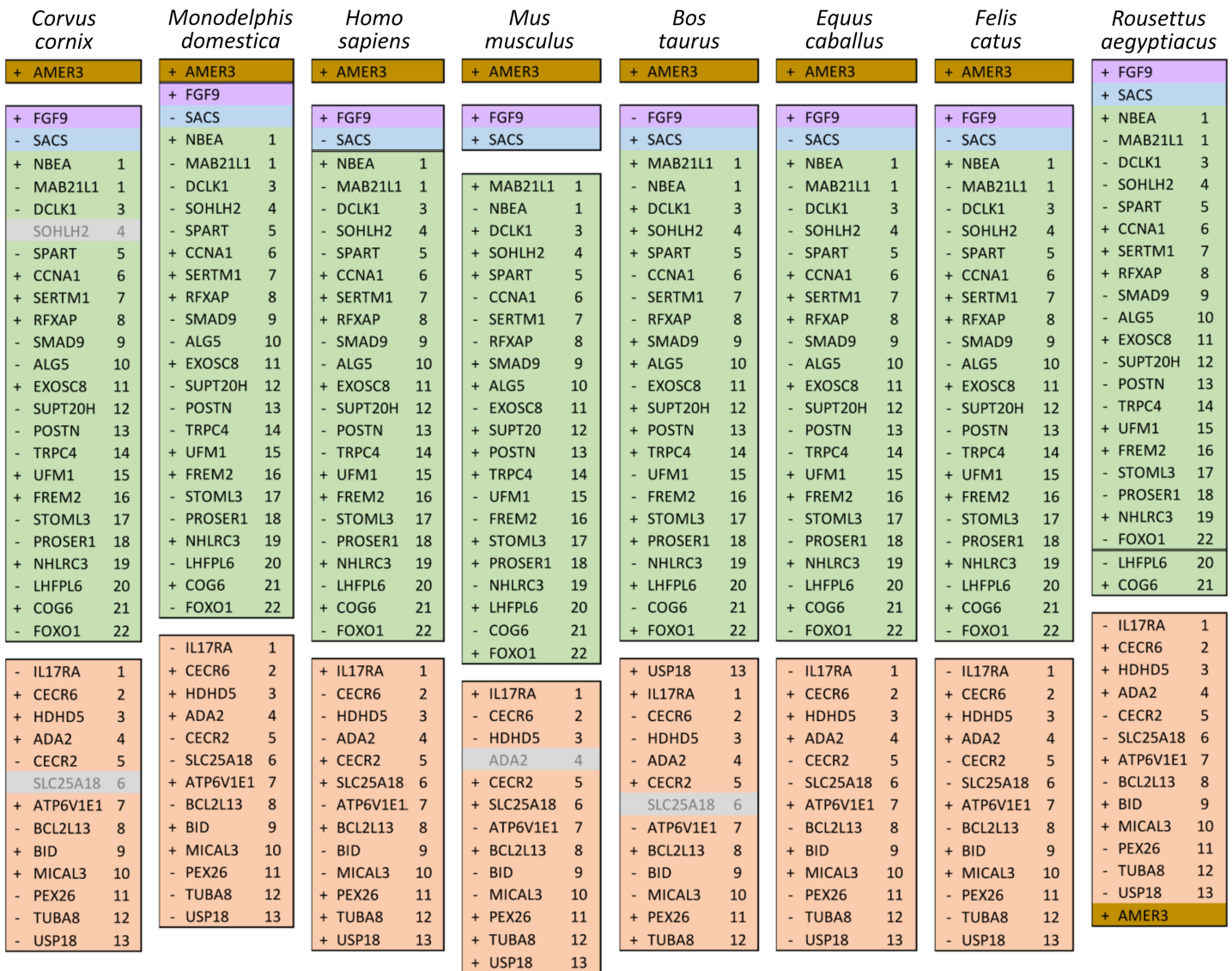
Figure 5. Relative positions in tetrapod genomes of synteny blocks and individual genes analyzed in this study.

Corvus cornix
	Gene	#
+	AMER3	
+	FGF9	
−	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
−	LHFPL6	20
+	COG6	21
−	FOXO1	22
−	IL17RA	1
+	CECR6	2
+	HDHD5	3
+	ADA2	4
−	CECR2	5
	SLC25A18	6
+	ATP6V1E1	7
−	BCL2L13	8
+	BID	9
+	MICAL3	10
−	PEX26	11
−	TUBA8	12
−	USP18	13

Monodelphis domestica
	Gene	#
+	AMER3	
+	FGF9	
−	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
−	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
−	LHFPL6	20
+	COG6	21
−	FOXO1	22
−	IL17RA	1
+	CECR6	2
+	HDHD5	3
+	ADA2	4
−	CECR2	5
−	SLC25A18	6
+	ATP6V1E1	7
−	BCL2L13	8
+	BID	9
+	MICAL3	10
−	PEX26	11
−	TUBA8	12
−	USP18	13

Homo sapiens
	Gene	#
+	AMER3	
+	FGF9	
−	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
−	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
−	LHFPL6	20
+	COG6	21
−	FOXO1	22
+	IL17RA	1
−	CECR6	2
−	HDHD5	3
−	ADA2	4
+	CECR2	5
+	SLC25A18	6
−	ATP6V1E1	7
+	BCL2L13	8
−	BID	9
−	MICAL3	10
+	PEX26	11
+	TUBA8	12
+	USP18	13

Mus musculus
	Gene	#
+	AMER3	
+	FGF9	
+	SACS	
+	MAB21L1	1
−	NBEA	1
+	DCLK1	3
+	SOHLH2	4
+	SPART	5
−	CCNA1	6
−	SERTM1	7
−	RFXAP	8
+	SMAD9	9
+	ALG5	10
−	EXOSC8	11
+	SUPT20	12
+	POSTN	13
+	TRPC4	14
−	UFM1	15
−	FREM2	16
+	STOML3	17
+	PROSER1	18
−	NHLRC3	19
+	LHFPL6	20
−	COG6	21
+	FOXO1	22
+	IL17RA	1
−	CECR6	2
−	HDHD5	3
	ADA2	4
+	CECR2	5
+	SLC25A18	6
−	ATP6V1E1	7
+	BCL2L13	8
−	BID	9
−	MICAL3	10
−	PEX26	11
+	TUBA8	12
+	USP18	13

Bos taurus
	Gene	#
+	AMER3	
−	FGF9	
+	SACS	
+	MAB21L1	1
−	NBEA	1
+	DCLK1	3
+	SOHLH2	4
+	SPART	5
−	CCNA1	6
−	SERTM1	7
−	RFXAP	8
+	SMAD9	9
+	ALG5	10
−	EXOSC8	11
+	SUPT20H	12
+	POSTN	13
+	TRPC4	14
−	UFM1	15
−	FREM2	16
+	STOML3	17
+	PROSER1	18
−	NHLRC3	19
+	LHFPL6	20
−	COG6	21
+	FOXO1	22
+	USP18	13
+	IL17RA	1
−	CECR6	2
−	HDHD5	3
−	ADA2	4
+	CECR2	5
	SLC25A18	6
−	ATP6V1E1	7
+	BCL2L13	8
−	BID	9
−	MICAL3	10
+	PEX26	11
+	TUBA8	12

Equus caballus
	Gene	#
+	AMER3	
+	FGF9	
−	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
−	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
+	LHFPL6	20
+	COG6	21
−	FOXO1	22
−	IL17RA	1
+	CECR6	2
+	HDHD5	3
+	ADA2	4
−	CECR2	5
−	SLC25A18	6
+	ATP6V1E1	7
−	BCL2L13	8
+	BID	9
+	MICAL3	10
−	PEX26	11
−	TUBA8	12
−	USP18	13

Felis catus
	Gene	#
+	AMER3	
+	FGF9	
−	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
−	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
−	LHFPL6	20
+	COG6	21
−	FOXO1	22
−	IL17RA	1
+	CECR6	2
+	HDHD5	3
+	ADA2	4
−	CECR2	5
−	SLC25A18	6
+	ATP6V1E1	7
−	BCL2L13	8
+	BID	9
+	MICAL3	10
−	PEX26	11
−	TUBA8	12
−	USP18	13

Rousettus aegyptiacus
	Gene	#
+	FGF9	
+	SACS	
+	NBEA	1
−	MAB21L1	1
−	DCLK1	3
−	SOHLH2	4
−	SPART	5
+	CCNA1	6
+	SERTM1	7
+	RFXAP	8
−	SMAD9	9
−	ALG5	10
+	EXOSC8	11
−	SUPT20H	12
−	POSTN	13
−	TRPC4	14
+	UFM1	15
+	FREM2	16
−	STOML3	17
−	PROSER1	18
+	NHLRC3	19
−	FOXO1	22
−	LHFPL6	20
+	COG6	21
−	IL17RA	1
+	CECR6	2
+	HDHD5	3
+	ADA2	4
−	CECR2	5
−	SLC25A18	6
+	ATP6V1E1	7
−	BCL2L13	8
+	BID	9
+	MICAL3	10
−	PEX26	11
−	TUBA8	12
−	USP18	13
+	AMER3	

Figure 5 Relative positions in tetrapod genomes of synteny blocks and individual genes analyzed in this study. Each species diagram consists of one gene per row, represented by the human gene symbol for the orthologous group. A plus or minus sign indicates the orientation of the gene on the reference sequence. Gaps between gene blocks indicate they are on different linkage groups. A double line between genes on the same linkage group indicates a gap greater than 10 Mb. The *Nbea-Foxo1* (NF) synteny block is colored green and the cat-eye (CE) synteny block is colored orange (see text for definitions of these blocks). Within each block, genes are numbered according to their order in the human genome as a reference. The genes *Sacs*, *Fgf9*, and *Amer3* are not numbered because they are not considered part of either synteny block; rather, they are shown because they were identified as positive selection candidates closely linked to the two synteny blocks in *Lasiurus cinereus*. These three genes are colored blue, purple, and brown, respectively. Gene symbols are in unitalicized upper case for legibility. Genes that are unannotated and presumed absent in a species are grayed. Full-size ☑ DOI: 10.7717/peerj.17482/fig-5

ancestors, or the merged NF and CE blocks subsequently split in molossids (a reversal). Further complicating the inferred order of chromosomal changes, the molossid and phyllostomid taxa also share a very pronounced increase in GC content within the coding sequences of the analyzed genes (Fig. S4). This shift in background nucleotide composition may relate to the extreme subtelomeric location of NF and CE genes in some of these taxa. For example, the most 3' gene coordinate of the merged NF+CE block in *A. jamaicensis* is

Figure 6

Gene synteny diagram across twelve tetrapod genomes (species columns shown left to right). Each entry lists gene orientation (+/−) and the human-genome order number.

Rousettus aegyptiacus
+ FGF9
+ SACS
+ NBEA 1
− MAB21L1 1
− DCLK1 3
− SOHLH2 4
− SPART 5
+ CCNA1 6
+ SERTM1 7
+ RFXAP 8
− SMAD9 9
− ALG5 10
+ EXOSC8 11
− SUPT20H 12
− POSTN 13
− TRPC4 14
+ UFM1 15
+ FREM2 16
− STOML3 17
− PROSER1 18
+ NHLRC3 19
+ FOXO1 22
− LHFPL6 20
+ COG6 21

− IL17RA 1
+ CECR6 2
+ HDHD5 3
+ ADA2 4
− CECR2 5
− SLC25A18 6
+ ATP6V1E1 7
− BCL2L13 8
+ BID 9
+ MICAL3 10
− PEX26 11
− TUBA8 12
− USP18 13
+ AMER3

Rhinolophus ferrumequinum
+ FGF9
− SACS
+ MAB21L1 1
− NBEA 1
+ DCLK1 3
+ SOHLH2 4
+ SPART 5
+ CCNA1 6
− SERTM1 7
− RFXAP 8
+ SMAD9 9
+ ALG5 10
− EXOSC8 11
+ SUPT20H 12
+ POSTN 13
+ TRPC4 14
− UFM1 15
+ FREM2 16
+ STOML3 17
+ PROSER1 18
− NHLRC3 19
− FOXO1 22
+ LHFPL6 20
− COG6 21

+ IL17RA 1
− CECR6 2
− HDHD5 3
+ ADA2 4
+ CECR2 5
+ SLC25A18 6
− ATP6V1E1 7
+ BCL2L13 8
− BID 9
− MICAL3 10
+ PEX26 11
+ TUBA8 12
+ USP18 13
+ AMER3

Artibeus jamaicensis
+ FGF9

+ FOXO1 22
− COG6 21
+ LHFPL6 20
+ SACS
+ NBEA 1
− MAB21L1 1
− DCLK1 3
− SOHLH2 4
− SPART 5
+ CCNA1 6
+ SERTM1 7
+ RFXAP 8
− SMAD9 9
− ALG5 10
+ EXOSC8 11
− SUPT20H 12
− POSTN 13
− TRPC4 14
+ UFM1 15
+ FREM2 16
  STOML3 17
− PROSER1 18
+ NHLRC3 19
− AMER3
+ IL17RA 1
− CECR6 2
− HDHD5 3
− ADA2 4
+ CECR2 5
+ SLC25A18 6
− ATP6V1E1 7
+ BCL2L13 8
− BID 9
− MICAL3 10
+ PEX26 11
+ TUBA8 12
+ USP18 13

Phyllostomus discolor
+ FGF9

− FOXO1 22
+ COG6 21
− LHFPL6 20
− SACS
+ MAB21L1 1
+ NBEA 1
+ DCLK1 3
+ SOHLH2 4
+ SPART 5
− CCNA1 6
− SERTM1 7
+ RFXAP 8
+ SMAD9 9
+ ALG5 10
− EXOSC8 11
+ SUPT20H 12
+ POSTN 13
+ TRPC4 14
− UFM1 15
− FREM2 16
  STOML3 17
+ PROSER1 18
− NHLRC3 19
+ AMER3
− IL17RA 1
+ CECR6 2
+ HDHD5 3
− ADA2 4
− CECR2 5
− SLC25A18 6
+ ATP6V1E1 7
+ BCL2L13 8
+ BID 9
+ MICAL3 10
− PEX26 11
− TUBA8 12
− USP18 13

Molossus molossus
+ FOXO1 22
− COG6 21
+ LHFPL6 20
+ SACS
+ NBEA 1
− MAB21L1 1
− DCLK1 3
− SOHLH2 4
− SPART 5
+ CCNA1 6
+ SERTM1 7
+ RFXAP 8
− SMAD9 9
− NHLRC3 19
+ PROSER1 18
  STOML3 17
− FREM2 16
− UFM1 15
+ TRPC4 14
+ POSTN 13
+ SUPT20H 12
− EXOSC8 11
+ ALG5 10

− AMER3
+ IL17RA 1
− CECR6 2
− HDHD5 3
− ADA2 4
+ CECR2 5
+ SLC25A18 6
− ATP6V1E1 7
+ BCL2L13 8
− BID 9
− MICAL3 10
+ PEX26 11
+ TUBA8 12
+ USP18 13
− FGF9

Myotis myotis
+ FOXO1 22
− COG6 21
+ LHFPL6 20
+ SACS
+ NBEA 1
− MAB21L1 1
− DCLK1 3
+ HDHD5 3
+ CECR6 2
− IL17RA 1
+ AMER3
+ ALG5 10
+ EXOSC8 11
− SUPT20H 12
− POSTN 13
− TRPC4 14
+ UFM1 15
+ FREM2 16
− PROSER1 18
+ NHLRC3 19
+ SMAD9 9
− RFXAP 8
  SERTM1 7
− CCNA1 6
+ SPART 5
+ SOHLH2 4
− USP18 13
− TUBA8 12
− PEX26 11
+ MICAL3 10
+ BID 9
− BCL2L13 8
+ ATP6V1E1 7
− SLC25A18 6
− CECR2 5
  ADA2 4
− FGF9

Eptesicus fuscus
− UFM1 15

− FOXO1 22
+ COG6 21
− LHFPL6 20
− SACS
+ MAB21L1 1
+ NBEA 1
+ DCLK1 3
+ SOHLH2 4
+ SPART 5
− CCNA1 6
  SERTM1 7
+ RFXAP 8
+ SMAD9 9
+ NHLRC3 19
− PROSER1 18
  STOML3 17
+ FREM2 16
− TRPC4 14
− POSTN 13
+ SUPT20H 12
+ EXOSC8 11
− ALG5 10
+ AMER3
− HDHD5 3
− CECR6 2
+ IL17RA 1

− FGF9
+ ADA2 4
− CECR2 5
− SLC25A18 6
+ ATP6V1E1 7
− BCL2L13 8
+ BID 9
+ MICAL3 10
+ USP18 13
+ PEX26 11
+ TUBA8 12

Pipistrellus kuhlii
− FOXO1 22
+ COG6 21
− LHFPL6 20
− SACS
+ MAB21L1 1
− NBEA 1
+ DCLK1 3
+ SOHLH2 4
+ SPART 5
− CCNA1 6
  RFXAP 8
+ SMAD9 9
+ NHLRC3 19
− PROSER1 18
  STOML3 17
+ FREM2 16
+ UFM1 15
− TRPC4 14
− POSTN 13
− SUPT20H 12
+ EXOSC8 11
− ALG5 10
+ AMER3
− HDHD5 3
− CECR6 2
+ IL17RA 1
− PEX26 11
− TUBA8 12
− USP18 13
− MICAL3 10
− BID 9
+ BCL2L13 8
− ATP6V1E1 7
+ SLC25A18 6
+ CECR2 5
  ADA2 4
+ FGF9

Lasiurus cinereus
+ UFM1 15

+ FOXO1 22
+ COG6 21
+ LHFPL6 20
+ SACS
− MAB21L1 1
+ NBEA 1
− DCLK1 3
− SOHLH2 4
− SPART 5
+ CCNA1 6
  SERTM1 7
+ RFXAP 8
− SMAD9 9
− NHLRC3 19
+ PROSER1 18
  STOML3 17
− FREM2 16
+ TRPC4 14
+ POSTN 13
− SUPT20H 12
+ EXOSC8 11
+ ALG5 10
+ AMER3
+ HDHD5 3
+ CECR6 2
− IL17RA 1
+ PEX26 11
+ TUBA8 12
+ USP18 13
+ MICAL3 10
+ BID* 9
− BCL2L13 8
+ ATP6V1E1 7
− SLC25A18 6
− CECR2 5
  ADA2 4
− FGF9

Figure 6 Relative positions in twelve tetrapod genomes of synteny blocks and individual genes analyzed in this study. Each species diagram consists of one gene per row, represented by the human gene symbol for the orthologous group. A plus or minus sign indicates the orientation of the gene on the reference sequence. Gaps between genes indicate they are on different linkage groups. A double line between genes on the same linkage group indicates a physical distance greater than 10 Mb. The *Nbea-Foxo1* (NF) synteny block is colored green and the cat-eye (CE) synteny block is colored orange (see text for definitions of these blocks). Within each block, genes are numbered according to their order in the human genome as a reference. The genes *Sacs*, *Fgf9* and *Amer3* are not numbered because they are not considered part of either synteny block; rather, they are shown because they were identified as positive selection candidates closely linked to the two synteny blocks in *Lasiurus cinereus*. These three genes are colored blue, purple, and brown, respectively. Gene symbols are in unitalicized upper case for legibility. Genes that are unannotated and presumed absent in a species are grayed. Species are grouped by phylogenetic position using the same color scheme as in Fig. 7. The asterisk denotes uncertainty as to whether the gene *Bid* is functional in *L. cinereus*. Full-size 🖾 DOI: 10.7717/peerj.17482/fig-6

only six bases from the end of the linkage group (File S3). Changes in genomic location relative to chromosome ends are known to influence background nucleotide composition in mammals, and these compositional shifts can occur rapidly (*Montoya-Burgos, Boursot & Galtier, 2003*).

While Figs. 5 and 6 illustrate the changing relative positions of the positive selection candidates *Amer3*, *Fgf9*, and *Sacs* in tetrapod genomes, it should not be inferred that those genes have necessarily moved individually. Identifying and plotting conserved landmark genes that are adjacent to those positive selection candidates in outgroup and bat genomes demonstrates that each of the three selection candidates has moved as part of larger
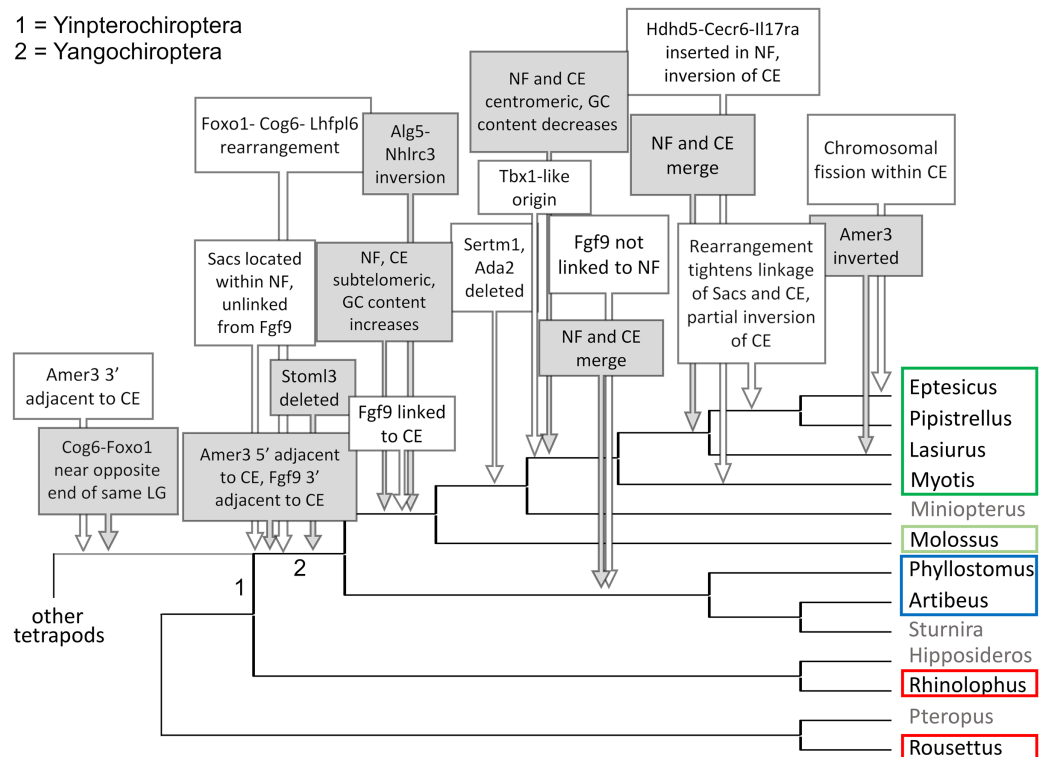
Figure 7 Hypothesized sequence of structural changes in gene organization in bats. Taxa for which chromosome-scale assemblies were available for this analysis are marked by colored boxes, which correspond to the colors used in Fig. 6. Grayed taxa were not analyzed for synteny because the relevant genes were not on large linkage groups. Genes are identified by their gene symbols, whereas NF and CE denote synteny blocks described in the text. GC denotes G + C content of gene coding sequences.

Full-size 🔄 DOI: 10.7717/peerj.17482/fig-7

multigene blocks (Figs. S5 and S6). *Amer3* likely moved as part of an ancestral block of eight genes (Fig. S5), whereas the genes that flank them in other mammals, represented by the human gene order in Fig. S5B, were not lost in bats but are instead located distant from *Amer3* on the same chromosome or on a different chromosome. Additional inversions of these discrete gene blocks have subsequently occurred during bat evolution, giving rise to diverse relative orientations (Fig. S5). Linkage of the *Amer3* block to the CE and NF genes has also been maintained despite subsequent chromosomal breakage (*e.g.*, the genes are telomeric in *A. jamaicensis* and *M. molossus* as noted above but occur in the middle of large linkage groups in the other taxa shown in Fig. S5).

The selection candidates *Fgf9* and *Sacs* are tightly linked in all outgroup genomes examined (Fig. S6) but based on conserved landmark genes have split into two distinct, rearranged blocks in bats. (Note other, less conserved genes are also present in the vicinity but are not shown for clarity.) These gene blocks are maintained in approximately the same order in outgroups, with the exception of an inversion in mouse, and lie several Mb from the NF block. In bats, the *Fgf9* and *Sacs* blocks are separated from each other by several Mb or are on separate linkage groups. In molossid and vespertilionid bats, the Fgf9 block is tightly linked to the NF block and has remained so through subsequent

rearrangements in the region. The *Fgf9* block has also remained largely intact in all bat lineages examined, whereas most genes of the *Sacs* block present in Yinpterochiroptera are found elsewhere in the genomes of Yangochiroptera.

It remains to be determined whether the CE gene *Bid* is intact in *L. cinereus*. BLASTN alignments of *Bid* orthologs from other Vespertilionidae identified only a single contiguous match of ~200 nt within the *L. cinereus* assembly (for comparison, the coding sequence is ~700 nt in *E. fuscus*). Gene pseudogenization and loss are themselves functionally significant events that may relate to adaptive divergence as well, and indeed three genes of the NF and CE blocks (*Ada2*, *Stoml3*, and *Sertm1*) were not found in multiple vespertilionid bats. Pseudogenization of at least three genes has also substantially altered the *Amer3* block subsequent to being linked to the NF and CE blocks in bats, although *Gpr148* has been lost in other mammals as well (Fig. S5A). However, some presumed gene losses could simply reflect assembly errors, such that transcriptomic data may be needed to confirm their sequence and functionality. A summary of gene order and orientation of all gene blocks discussed here, including gene loss events, is depicted for bats and outgroups in Fig. S7.

Functional roles of positively selected genes and synteny blocks

Eight of the 12 positive selection candidates in *L. cinereus* have peak expression in brain tissue in either human or mouse based on NCBI Gene data, whereas nine have enriched expression in a brain tissue category according to the DAVID "Up_tissue" annotation table (Table 3). Similarly biased expression is seen in the DAVID data when all genes in the NF and CE blocks are included regardless of selection test result. For the 38 genes for which DAVID annotation information was available, 29 (76.3%) had elevated expression in human brain tissue (Table 3). However, a smaller proportion of genes (13 of 37 genes with available data, or 35.1%) had peak expression in brain or central nervous system (CNS) in either human or mouse based on NCBI Gene data. No significant gene ontology enrichment was found for either the selected genes or for all NF and CE genes. I conclude that while the positive selection candidates have relatively high expression in brain or CNS, as do the NF and CE genes as a group, neither gene set is over-represented in annotated pathways or biological processes. Furthermore, a recent protein-protein interaction map of the mouse brain uncovered no direct pairwise interactions between proteins encoded by the genes listed in Fig. 5 (see Table S3 of *Pourhaghighi et al., 2020*).

For five selection candidates, deleterious mutations are associated with mild to severe defects of organogenesis or embryonic neural development, as indicated by clinical variants in human or by experiments in animal models. *Cecr2* deletion causes anencephaly, a severe defect of cranial and neural development (*Banting et al., 2005*; *Fairbridge et al., 2010*; *Dicipulo et al., 2021*). *Frem2* loss of function underlies Fraser Syndrome, characteristics of which include cryptophthalmy and syndactyly, both of which can be recapitulated in a mouse model (*Jadeja et al., 2005*; *Timmer et al., 2005*). A frameshift in *Proser1* has been associated with craniofacial dysmorphy and genitourinary developmental defects in human (*Salah et al., 2022*). Missense mutations in *Fgf9* are associated with multiple syntoses syndrome, which is characterized by joint fusions of the hand and

Table 3 Tissue-specific expression patterns for positive-selection candidates as well as other genes of the NF and CE blocks (see text for block definition). The sources of tissue-specific expression are described in the text. Tissues that are part of the brain or central nervous system are bolded. Tissue labels that are specific to human disease states were excluded. Positive selection candidates are denoted by an asterisk.

Gene symbol	Gene name	DAVID tissue enrichment	NCBI tissue of peak expression (human/mouse)
Ada2	Adenosine deaminase 2	**Brain**, Liver, Thymus, Trachea, Uterus	Spleen/NA
Alg5	Dolichyl-phosphate beta-glucosyltransferase	**Hypothalamus**, Liver, Prostate, Umbilical cord blood	Thyroid/placenta adult
Amer3*	APC membrane recruitment protein 3	**Brain, Cerebellum**	**Brain/CNS E18**
Atp6v1e1	ATPase H+ transporting V1 subunit E1	**Amygdala**, **Brain**, **Cajal-Retzius cell**, **Cerebellum**, Liver, Lung	**Brain/cortex adult**
Bcl2l13	BCL2 like 13	**Amygdala**, **Brain**, Eye, Human skeletal muscle, Liver, Skin, Synovial membrane tissue, Testis, Trachea	Fat/heart adult
Bid	BH3 interacting domain death agonist	**Brain**, Embryonic kidney, Fetal liver, Liver, Normal colorectal tissue, Skin	Bone marrow/kidney adult
Ccna1	Cyclin A1	**Brain**, Myeloid, Testis	Testis/testis adult
Cecr2*	Histone acetyl-lysine reader	**Brain**, Liver, Skeletal muscle	**Brain**/testis adult
Cecr6*	Transmembrane protein 121B	**Brain**	**Brain**/NA
Cog6	Component of oligomeric golgi complex 6	**Amygdala**, Aorta, **Brain**, Fetal skin	Testis/testis adult
Dclk1	Doublecortin like kinase 1	**Brain, Fetal brain, Hippocampus**	**Brain/frontal lobe adult**
Exosc8	Exosome component 8	**Brain**, Uterus	Testis/**CNS E11.5**
Fgf9*	Fibroblast growth factor 9	Foreskin, Kidney	Kidney/**cerebellum adult**
Foxo1	Forkhead box O1	Lymph, Placenta	Ovary/ovary adult
Frem2*	FRAS1 related extracellular matrix 2	Fetal kidney, Plasma, Tongue	Kidney/bladder adult
Hdhd5	Haloacid dehalogenase like hydrolase domain containing 5	**Brain**, Embryo, Lymph, Testis	Duodenum/thymus adult
Il17ra	Interleukin 17 receptor A	Placenta, T-cell, Uterus	Bone marrow/thymus adult
Lhfpl6	LHFPL tetraspan subfamily member 6	**Amygdala**, Lung	Fat/lung adult
Mab21l1	Mab-21 like 1	**Brain**	NA/NA
Mical3*	Microtubule associated monooxygenase, calponin and LIM domain containing 3	**Brain**, Liver, Lymph, Pancreas, Testis	Testis/testis adult
Nbea*	Neurobeachin	**Brain**, Embryonic head, Spleen, Testis	**Brain/CNS E18**
Nhlrc3	NHL repeat containing 3	Heart, Placenta, Testis	Thyroid/placenta adult
Pex26	Peroxisomal biogenesis factor 26	**Brain**, Colon, Fibroblast, Ileal mucosa, Uterus	Colon/ovary adult
Postn*	Periostin	Liver, Periodontal ligament, Placenta, Plasma, Thyroid gland	Skin/limb E14.5
Proser1*	Proline and serine rich 1	**Amygdala, Brain**, Colon endothelium, Peripheral nervous system	Placenta/thymus adult
Rfxap	Regulatory factor X associated protein	Lymphoblast, Testis, **Thalamus**	Testis/ovary adult
Sacs*	Sacsin	**Astrocyte, Brain**, Fetal liver, Uterine endothelium	**Brain/CNS E18**
Sertm1	Serine rich and transmembrane domain containing 1	**Amygdala, Brain**	**Brain/CNS E18**
Slc25a18*	Solute carrier family 25 member 18	**Brain**, Liver	**Brain/cortex adult**
Smad9	SMAD family member 9	**Brain**, Eye, **Fetal brain**	Thyroid/adrenal adult
Sohlh2	Spermatogenesis and oogenesis specific basic helix-loop-helix 2	Testis	Testis/testis adult
Spart	Spartin	**Brain**, Placenta	Ovary/limb E14.5

(Continued)

Table 3 (continued)			
Gene symbol	Gene name	DAVID tissue enrichment	NCBI tissue of peak expression (human/mouse)
Stoml3	Stomatin like 3	Lung, Trachea	Lung/**frontal lobe adult**
Supt20h	SPT20 homolog, SAGA complex component	Kidney, Prostate, Testis, Trachea	Testis/testis adult
*Trpc4**	Transient receptor potential cation channel subfamily C member 4	Embryonic kidney, Kidney, **Thalamus**	Endometrium/**frontal lobe adult**
Tuba8	Tubulin alpha 8	**Amygdala, Brain**, **Caudate nucleus**, Skeletal muscle	Heart/testis adult
Ufm1	Ubiquitin fold modifier 1	Bone marrow, **Brain**, Kidney	Thyroid/placenta adult
Usp18	Ubiquitin specific peptidase 18	**Brain**, Ovary, Uterus	Fat/liver E18

cranium as well as craniofacial dysmorphy (*Wu et al., 2009*; *Rodriguez-Zabala et al., 2017*); these phenotypes can be recapitulated in a mouse model (*Tang et al., 2017*). Missense mutations, microdeletions, and reciprocal translocations in *Nbea* are associated with neurodevelopmental disease, including autism and epilepsy (*Mulhern et al., 2018*). While no developmental defect has been reported for *Amer3* specifically, the homolog to which *Amer3* binds, *Amer1*, does have an association with craniofacial dysmorphy and abnormal organogenesis in human (*Mi et al., 2020*). However, not all deleterious phenotypes of selection candidates appear early in development: *Sacs* mutations underlie an adult-onset neurodegenerative syndrome characterized by spastic ataxia (*Bagaria, Bagyinszky & An, 2022*).

In addition to these gene-specific associations, diverse structural aberrations in human involving the CE block, such as supernumerary chromosomes and microdeletions, have overlapping phenotypes with a craniofacial component (reviewed in *Glaeser et al., 2021*). A spontaneous deletion of the majority of NF genes has been associated with impaired neurological development and craniofacial dysmorphy in an isolated clinical report (*Miura et al., 2020*). Recurrent deletion or duplication of five genes that include *Amer3* (also known as *Fam123C*) is associated with clinically diagnosed behavioral problems, epilepsy, and cranial dysmorphy (*Dharmadhikari et al., 2012*); the deletion spans the genes *Gpr148* to *Plekthb2* shown in Fig. S5A.

The phenotypes of deleterious mutations reveal aspects of gene function, but may be completely unrelated to positively selected phenotypic variation. Nonetheless, bats are well known for extreme cranial divergence that underpins ecological traits. Cranial morphology has been shown to evolve *via* allometric processes such as heterochrony (*Camacho et al., 2020*), the underlying mechanisms of which are beginning to be revealed (*Camacho et al., 2019*). *Arbour, Curtis & Santana (2021)* identified distinct modules of cranial and mandibular development that have diversified across the bat phylogeny, particularly with respect to oral-echolocating, nasal-echolocating, and non-echolocating taxa.

The biomechanics of feeding also strongly shapes cranial evolution at both deep and shallow divergence times (*Hedrick & Dumont, 2018*; *Camacho et al., 2019*). For example,

cranial morphology was found to parallel dietary similarity in some Mediterranean *Myotis* species (*Evin et al., 2008*).

Birth and death of a Tbx1-like gene family in Vespertilionidae

Cornman & Cryan (2022) identified a variable gene family with partial homology to the DNA-binding transcription factor *Tbx1*, the latter having many important roles in embryonic development (*Baldini, Fulcoli & Illingworth, 2017*). The gene family was initially identified because one member lies within the CE block in *L. cinereus* but homologs were not detected in other tetrapod orders. In addition to multiexon gene models, single-exon genes and partial genes were identified, suggesting that at least some of these '*Tbx1*-like' family members were retrogenes or pseudogenes. To further investigate the origins and functionality of this gene-family expansion, I performed a TBLASTN search with the homologous sequence XP_027987819.1 from *E. fuscus* against the genomes of *P. kuhlii* and *M. myotis*. Multiple unannotated *Tbx1*-like homologs are present in both of the latter genomes (File S4), at least some of which appear to have the minimum complement of gene features (a multiexon example with a predicted core promoter is shown in File S5). No Tbx1-like sequences were identified in non-vespertilionid bat genomes.

High-scoring BLAST matches in *P. kuhlii* are also consistently supported by low-level RNA-Seq coverage at those sites (Fig. S8B). Furthermore, the well conserved portion of the *Tbx1*-like family retains the key arginine residue that in the TBX1 protein binds the major groove of double-stranded DNA (*El Omari et al., 2012*), and the N-terminal protein regions have secondary structures very similar to that of human *Tbx1* even after all primary sequence homology is lost (Fig. S8A). In contrast, the C-terminal portions of the predicted proteins appears more disordered and lack the second and third alpha helices and DNA interaction residues found in TBX1. This pattern of sequence divergence also characterizes the *Tbx* gene family generally (*El Omari et al., 2012*), in that the conserved protein domain that defines the *Tbx* family encompasses the same region that *Tbx1* shares with *Tbx1*-like, whereas *Tbx* homologs show increased divergence at the same point that *Tbx1*-like diverges from *Tbx1*, which encompasses the dimerization domain.

Evidence of transcription and conservation of functionally important domains suggests that at least some *Tbx1*-like genes are functional, albeit with potentially high evolutionary turnover or pseudogenization rates. The latter possibility is illustrated by a comparison of the *P. kuhlii Tbx1*-like gene proposed in File S5 to its closest homologs in three other vespertilionids, including *L. cinereus* (Fig. S8C). All three homologs in these other taxa show clear signs of pseudogenization, including internal stop codons, frameshifts, and an unconserved start codon. I conclude that the burst of *Tbx1*-like sequences began early in the Vespertilionid lineage if not earlier, likely with continued birth and death in subsequent lineages.

While pseudogenes are ubiquitous in complex genomes (and often disregarded on non-scientific grounds (*Cheetham, Faulkner & Dinger, 2020*), the apparent burst in *Tbx1*-like duplications within the same lineage exhibiting positive selection on genes affecting cranial morphology would be a remarkable coincidence if functionally unrelated. This is

because *Tbx1* mutation, duplication, and haploinsufficiency are all associated with craniofacial dysmorphy syndrome in human (DiGeorge Syndrome), which can be recapitulated in a mouse model (*Lindsay et al., 2001*; *Liao et al., 2004*; *McDermid & Morrow, 2002*). An expansive literature has since established *Tbx1* as a key regulator of cranial development, governing aspects of the differentiation and migration of cells arising from the pharyngeal arches and neural crest (reviewed in *Baldini, Fulcoli & Illingworth, 2017*). The human phenotypes associated with *Tbx1* dosage variation overlap with cat-eye syndrome and both syndromes are attributable in part to defects in pharyngeal arch development mediated by abnormal cell migration (*Tan et al., 2010*, *Ivins & Scambler, 2022*). Indeed, *Tbx1* is tightly linked to the CE synteny block in human and deletions of either constitute similar subclasses of 22q11 microdeletion syndrome (*Tan et al., 2010*), although *Tbx1* is not located near either synteny block in mammals generally (see *Tbx1* gene pages for taxa in Fig. 5 and File S3). One hypothesis suggested by the advent of *Tbx1*-like sequences in vesper bats is that they have been at least transiently functional and have affected aspects of *Tbx1*-regulatory networks relevant to positively selected phenotypes in the hoary bat lineage, perhaps by modulating or inhibiting TBX1 dimerization at particular binding sites. Tbx1 is believed to regulate gene expression by binding C-rich DNA motifs and then promoting histone methylation near transcription start sites in a dosage-dependent fashion (*Fulcoli et al., 2016*). This hypothesis does not require that TBX1 directly regulate selection candidates or other genes with which they are linked, although 3 of 12 selection candidates (*Frem2*, *Proser1*, and *Slc25a18*) were differential expressed in a mouse *Tbx1* haploinsufficiency model (*Fulcoli et al., 2016*), compared with 1,992 of 22,807 total genes (8.7%) tested in that study.

CAVEATS AND CONCLUSIONS

This analysis further established a cluster of 12 genes exhibiting signatures of diversifying selection within the hoary bat lineage, which lie within a 12-Mb window of the hoary bat genome. This genomic hotspot is dominated by re-assorted elements of two distinct synteny blocks that are conserved and unlinked in other tetrapods. In fact, structural changes within this region during bat evolution appears to require parallel occurrences or subsequent reversals in different bat lineages. The selected genes specifically and the synteny blocks generally are associated with cranial and neural development, based on expression patterns, disease associations, and functional studies in model organisms. The analysis also confirmed previously reported *Tbx1*-like duplications within vesper bats, although the timing remains difficult to determine since the sequences are largely unannotated in those genomes and turnover appears rapid. Nonetheless, the strict conservation of a critical functional region despite high divergence elsewhere in the predicted proteins (see also *Cornman & Cryan, 2022*) implies purifying selection. I conclude that this genomic region is a hotspot of adaptive evolution in the hoary bat lineage that likely relates to cranial and neurological traits underlying ecological diversification.

Different tests of selection fit conceptually distinct, parameter-rich models to quantify excess nonsynonymous substitutions in an alignment, a phenomenon that is likely both

transient and conservative as a metric of positive selection (see *Seo, Kishino & Thorne, 2004*). Not surprisingly, different algorithms and data sets give different results. For example, positive rates were higher with the PAML package than with the HyPhy package, and the PAML algorithm is known to be subject to false positives if the assumption of rate homogeneity on background lineages is unreasonable (*Smith et al., 2015*). Nonetheless, results from the two packages were broadly similar for longer alignments, indicating convergence of model results as the information content of alignments increased. For example, using either tree 1 or tree 2, six of nine alignments exceeding 2,000 analyzed positions were supported by the BUSTED method at an FDR-corrected *P*-value of less than 0.05 (File S2). For tree 3, seven of nine were concordant. Incomplete concordance between the two methods may also arise if diversification had begun prior to the divergence of *Lasiurus*, as suggested by the fact that the aBSREL method was generally not significant for the same comparisons. Methodological congruence might therefore be greater for alternative designations of foreground branches. Yet the diversification of *Trpc4* within bats (Fig. 4) is a striking example of how different approaches to quantifying episodic selection can produce highly discordant interpretations of the same alignment.

Another caveat is that evolutionary rate estimation may be sensitive to errors in phylogeny or reconstructed states at unsampled nodes (*Feng et al., 2020*) and factors such as undetected paralogy, gene conversion, nucleotide composition bias, and recombination can lead to false positives (*Anisimova, Nielsen & Yang, 2003*; *Galtier & Duret, 2007*; *Ratnakumar et al., 2010*). For example, two genes identified as positive selection candidates in this region by *Cornman & Cryan (2022)*, *Rps13* and *Necap1*, were found to be paralogous retrogenes and removed from this analysis for this reason. Phylogenetic uncertainty also exists for the studied species, particularly with respect to vespertilionid taxa. This uncertainty was addressed by evaluating alternative guide trees, which produced qualitatively similar PAML results. Pruning compositionally skewed taxa (tree 4) generally increased *P*-values of the tests such that fewer were significant at the same alpha, yet most remained significant at an FDR-corrected *P*-value of 0.05. Furthermore, the closest outgroups of *L. cinereus* are very similar in nucleotide composition of tested genes (Fig. S3) and thus unlikely to bias the *L. cinereus* branch test. Moreover, the shift in background composition affected whole genomic regions yet only a minority of tested genes within them had signatures of positive selection.

Despite these important caveats, it bears emphasizing that the conclusions of this study are not predicated on any specific gene undergoing episodic positive selection. They are instead based on the tight genomic clustering of numerous positive selection tests concomitant with a high number of structural changes within synteny blocks that show few changes in other tetrapod orders. These observations hold regardless of any methodological sensitivity for a specific gene. This study also strengthens the genome-wide conclusions of *Cornman & Cryan (2022)*, given that all ten positive selection tests repeated here with additional data produced the same result at an alpha of 0.01. That study also identified several positively selected genes affecting cranial development elsewhere in the *L. cinereus* genome.

This study examined the order and evolutionary rates of protein-coding genes only. Purifying selection on the expression context of long noncoding RNA (lncRNA) genes has been proposed to constrain coding-gene synteny in some cases (*Hufton et al., 2009*), and lncRNA genes occur within these synteny blocks in many taxa. For example, the human genes *Cecr1*, *Cecr7*, and *Fam230D* are all lncRNAs within the CE block defined by Fig. 1. However, *Cecr1* and *Cecr7* have been shown not to be conserved with other mammals (*e.g.*, *Bridgland et al., 2003*; *Footz et al., 2001*) and both *Cecr7* and *Fam230D* appear to be recently arisen within the primate lineage (*Bridgland et al., 2003*; *Delihas, 2018*). lncRNA genes present in the human NF synteny block (*e.g.*, *LINC02343* and *LINC00571*) also do not have conserved orthologs listed in the UCSC genome browser. While the annotation of lncRNAs remains poorly developed compared with protein-coding genes, conservation of orthologous lncRNAs *per se* does not appear to drive the maintenance of the NF and CE blocks across tetrapod orders. Yet regulation of chromosomal 'neighborhoods' (*sensu Nora, Dekker & Heard, 2013*) by lncRNAs could still be a mechanism by which synteny is maintained even if the lncRNAs themselves turn over or no longer retain evidence of orthology (*Engreitz et al., 2016*; *Quinn et al., 2016*).

Regardless of the contribution of lncRNAs, co-regulation *via* shared *cis* regulatory elements and chromatin neighborhood effects remains an important hypothesis for the maintenance of these synteny blocks during tetrapod evolution as well as the clustering of positive selection candidates in *L. cinereus*. For example, 'transcriptionally associated chromosome domains' are recognized facets of hierarchical chromatin folding that contribute to correlated transcription on scales from kilobases to megabases (*Nora, Dekker & Heard, 2013*), distances that are very relevant to the synteny blocks studied here. Moreover, the critical link between chromatin state and co-expression may be manifested in only a few cell types or developmental stages (cf. *Eckalbar et al., 2016*), and thus not apparent in aggregate measures of gene expression such as given in Table 3.

Future directions could include refining estimates of the timing and magnitude of episodic selection within clades as more genomes become available. The hypothesis that structural changes within synteny blocks alter gene expression profiles, chromatin modifications, or chromatin topological domains could potentially be tested with RNA-Seq, ChIP-Seq, and HiC data, although the fact that many of these genes act during embryonic development is constraining (but see *Eckalbar et al., 2016* for an example in bats). Fine-scale analysis of genotype-phenotype associations for these genes, *e.g.*, by focusing on highly diverse genera such as *Myotis*, could suggest ecological drivers of diversifying selection. Further comparative genomic study of the orthologous genes in bats could also serve as a case study of how genic and karyotypic evolution interact to drive phenotypic divergence, given that karyotypic evolution is considered an important mode of adaptation and species diversification (*Damas, Corbo & Lewin, 2021*).

ACKNOWLEDGEMENTS

of *Bickham (1987)* is reproduced in Fig. S2 with permission. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

• Robert S. Cornman conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
DNAZoo genome assembly of the hoary bat, available at https://www.dnazoo.org/assemblies/aeorestes_cinereus

The short read data, PRJNA559902, were used to further evaluate the genome assembly.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.17482#supplemental-information.

REFERENCES

Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. 2011. A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLOS Currents Tree of Life* **3**:RRN1212.

Aguileta G, Refregier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution* **9(4)**:656–670 DOI 10.1016/j.meegid.2009.03.010.

Amador LI, Moyers Arevalo RL, Almeida FC, Catalano SA, Giannini NP. 2018. Bat systematics in the light of unconstrained analyses of a comprehensive molecular supermatrix. *Journal of Mammalian Evolution* **25(1)**:37–70 DOI 10.1007/s10914-016-9363-8.

Anisimova M, Nielsen R, Yang Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164(3)**:1229–1236 DOI 10.1093/genetics/164.3.1229.

Arbour J, Curtis A, Santana S. 2021. Sensory adaptations reshaped intrinsic factors underlying morphological diversification in bats. *BMC Biology* **19(1)**:1–13 DOI 10.1186/s12915-021-01022-3.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* **25(1)**:25–29 DOI 10.1038/75556.

Bagaria J, Bagyinszky E, An SSA. 2022. Genetics of autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) and role of sacsin in neurodegeneration. *International Journal of Molecular Sciences* **23**:552 DOI 10.3390/ijms23010552.

Baldini A, Fulcoli FG, Illingworth E. 2017. Tbx1: transcriptional and developmental functions. *Current Topics in Developmental Biology* **122(5)**:223–243 DOI 10.1016/bs.ctdb.2016.08.002.

Banting GS, Barak O, Ames TM, Burnham AC, Kardel MD, Cooch NS, Davidson CE, Godbout R, McDermid HE, Shiekhattar R. 2005. CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Human Molecular Genetics* **14**:513–524 DOI 10.1093/hmg/ddi048.

Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58(3)**:268–276 DOI 10.1016/j.ymeth.2012.05.001.

Bickham JW. 1987. Chromosomal variation among seven species of lasiurine bats (Chiroptera: Vespertilionidae). *Journal of Mammalogy* **68(4)**:837–842 DOI 10.2307/1381561.

Bridgland L, Footz TK, Kardel MD, Riazi AM, McDermid HE. 2003. Three duplicons form a novel chimeric transcription unit in the pericentromeric region of chromosome 22q11. *Human Genetics* **112(1)**:57–61 DOI 10.1007/s00439-002-0827-y.

Camacho J, Heyde A, Bhullar B-AS, Haelewaters D, Simmons NB, Abzhanov A. 2019. Peramorphosis, an evolutionary developmental mechanism in neotropical bat skull diversity. *Developmental Dynamics* **248(11)**:1129–1143 DOI 10.1002/dvdy.90.

Camacho J, Moon R, Smith SK, Lin JD, Randolph C, Rasweiler JJ, Behringer RR, Abzhanov A. 2020. Differential cellular proliferation underlies heterochronic generation of cranial diversity in phyllostomid bats. *EvoDevo* **11(1)**:11 DOI 10.1186/s13227-020-00156-9.

Cheetham SW, Faulkner GJ, Dinger ME. 2020. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics* **21(3)**:191–201 DOI 10.1038/s41576-019-0196-1.

Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, Hlavina W, Kapustin Y, Meric P, Maglott D, Birtle Z, Marques AC, Graves T, Zhou S, Teague B, Potamousis K, Churas C, Place M, Herschleb J, Runnheim R, Forrest D, Amos-Landgraf J, Schwartz DC, Cheng Z, Lindblad-Toh K, Eichler EE, Ponting CP, Mouse Genome Sequencing Consortium. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLOS Biology* **7(5)**:e1000112 DOI 10.1371/journal.pbio.1000112.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6(2)**:80–92 DOI 10.4161/fly.19695.

Cornman RS. 2024. DNA sequences used to analyze evolutionary rates of genes in bats: U.S. Geological Survey data release. *Available at https://doi.org/10.5066/P1YKF5BJ*.

Cornman RS, Cryan PM. 2022. Positively selected genes in the hoary bat (*Lasiurus cinereus*) lineage: prominence of thymus expression, immune and metabolic function, and regions of ancient synteny. *PeerJ* **10(12)**:e13130 DOI 10.7717/peerj.13130.

Cornman RS, Fike JA, Oyler-McCance SJ, Cryan PM. 2021. Historical effective population size of North American hoary bat (*Lasiurus cinereus*) and challenges to estimating trends in contemporary effective breeding population size from archived samples. *PeerJ* **9(2)**:e11285 DOI 10.7717/peerj.11285.

Damas J, Corbo M, Lewin HA. 2021. Vertebrate chromosome evolution. *Annual Review of Animal Biosciences* **9(1)**:1–27 DOI 10.1146/annurev-animal-020518-114924.

Delihas N. 2018. Formation of a family of long intergenic noncoding RNA genes with an embedded translocation breakpoint motif in human chromosomal low copy repeats of 22q11. 2 —some surprises and questions. *Non-Coding RNA* **4(3)**:16 DOI 10.3390/ncrna4030016.

Dharmadhikari AV, Kang S-HL, Szafranski P, Person RE, Sampath S, Prakash SK, Bader PI, Phillips III JA, Hannig V, Williams M, Vinson SS, Wilfong AA, Reimschisel TE, Craigen WJ, Patel A, Bi W, Lupski JR, Belmont J, Cheung SW, Stankiewicz P. 2012. Small rare recurrent deletions and reciprocal duplications in 2q21. 1, including brain-specific ARHGEF4 and GPR148. *Human Molecular Genetics* **21(15)**:3345–3355 DOI 10.1093/hmg/dds166.

Dicipulo R, Norton KA, Fairbridge NA, Kibalnyk Y, Fox SC, Hornberger LK, McDermid HE. 2021. Cecr2 mutant mice as a model for human cat eye syndrome. *Scientific Reports* **11**:3111 DOI 10.1038/s41598-021-82556-y.

Dong D, Lei M, Hua P, Pan YH, Mu S, Zheng G, Pang E, Lin K, Zhang S. 2016. The genomes of two bat species with long constant frequency echolocation calls. *Molecular Biology and Evolution* **34(1)**:20–34 DOI 10.1093/molbev/msw231.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research* **43(W1)**:W389-94 DOI 10.1093/nar/gkv332.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356(6333)**:92–95 DOI 10.1126/science.aal3327.

Eckalbar WL, Schlebusch SA, Mason MK, Gill Z, Parker AV, Booker BM, Nishizaki S, Muswamba-Nday C, Terhune E, Nevonen KA, Makki N, Friedrich T, VanderMeer JE, Pollard KS, Carbone L, Wall JD, Illing N, Ahituv N. 2016. Transcriptomic and epigenomic characterization of the developing bat wing. *Nature Genetics* **48(5)**:528–536 DOI 10.1038/ng.3537.

El Omari K, De Mesmaeker J, Karia D, Ginn H, Bhattacharya S, Mancini EJ. 2012. Structure of the DNA-bound T-box domain of human TBX1, a transcription factor associated with the DiGeorge syndrome. *Proteins: Structure, Function, and Bioinformatics* **80(2)**:655–660 DOI 10.1002/prot.23208.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539(7629)**:452–455 DOI 10.1038/nature20149.

Evin A, Baylac M, Ruedi M, Mucedda M, Pons J-M. 2008. Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): a geometric morphometric appraisal. *Biological Journal of the Linnean Society* **95(3)**:529–538 DOI 10.1111/j.1095-8312.2008.01076.x.

Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjöstedt E, Lundberg E, Szigyarto CA-K, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson A, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlén M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics: MCP* **13**:397–406 DOI 10.1074/mcp.M113.035600.

Fairbridge NA, Dawe CE, Niri FH, Kooistra MK, King-Jones K, McDermid HE. 2010. Cecr2 mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Research Part A: Clinical and Molecular Teratology* **88**:619–625 DOI 10.1002/bdra.20695.

Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth BC, Petersen B, Wang Z, Zhou Q, Diekhans M, Chen W, Andreu-Sánchez S, Margaryan A, Travis Howard J, Parent C, Pacheco G, Sinding MHS, Puetz L, Cavill E, Ribeiro AM, Eckhart L, Fjeldså J, Hosner PA, Brumfield RT, Christidis L, Bertelsen MF, Sicheritz-Ponten T, Thomas Tietze D, Robertson BC, Song G, Borgia G, Claramunt S, Lovette IJ, Cowen SJ, Njoroge P, Dumbacher JP, Ryder OA, Fuchs J, Bunce M, Burt DW, Cracraft J, Meng G, Hackett SJ, Ryan PG, Jønsson KA, Jamieson IG, da Fonseca RR, Braun EL, Houde P, Mirarab S, Suh A, Hansson B, Ponnikas S, Sigeman H, Stervander M, Frandsen PB, van der Zwan H, van der Sluis R, Visser C, Balakrishnan CN, Clark AG, Fitzpatrick JW, Bowman R, Chen N, Cloutier A, Sackton TB, Edwards SV, Foote DJ, Shakya SB, Sheldon FH, Vignal A, Soares AER, Shapiro B, González-Solís J, Ferrer-Obiol J, Rozas J, Riutort M, Tigano A, Friesen V, Dalén L, Urrutia AO, Székely T, Liu Y, Campana MG, Corvelo A, Fleischer RC, Rutherford KM, Gemmell NJ, Dussex M, Mouritsen H, Thiele N, Delmore K, Liedvogel M, Franke A, Hoeppner MP, Krone O, Fudickar AM, Milá B, Ketterson ED, Fidler AE, Friis G, Parody-Merino AM, Battley PF, Cox MP, Costa Barroso Lima NC, Prosdocimi F, Lee Parchman TL, Schlinger BA, Loiselle BA, Blake JG, Chuan Lim HC, Day LB, Fuxjager MJ, Baldwin MW, Braun MJ, Wirthlin M, Dikow RB, Ryder TB, Camenisch G, Keller LF, DaCosta JM, Hauber ME, Louder MIM, Witt CC, McGuire JA, Mudge J, Megna LC, Carling MD, Wang B, Taylor SA, Del-Rio G, Aleixo A, Tereza Ribeiro Vasconcelos AT, Mello CV, Weir JT, Haussler D, Li Q, Yang H, Wang J, Lei F, Rahbek C, Gilbert MTP, Graves GR, Jarvis ED, Paten B, Zhang G. 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**:252–257 DOI 10.1038/s41586-020-2873-9.

Footz TK, Brinkman-Mills P, Banting GS, Maier SA, Riazi MA, Bridgland L, Hu S, Birren B, Minoshima S, Shimizu N, Pan H, Nguyen T, Fang F, Fu Y, Ray L, Wu H, Shaull S, Phan S, Yao Z, Chen F, Huan A, Hu P, Wang Q, Loh P, Qi S, Roe BA, McDermid HE. 2001. Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: a search for candidate genes at or near the human chromosome 22 pericentromere. *Genome Research* **11(6)**:1053–1070 DOI 10.1101/gr.154901.

Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland J E, Mudge JM, Sisu C, Wright JC, Armstrong J, Barnes I, Berry A, Bignell A, Boix C, Carbonell Sala S, Cunningham F, Di Domenico Tás, Donaldson S, Fiddes I T, García Girón C, Gonzalez JM, Grego T, Hardy M, Hourlier T, Howe KL, Hunt T, Izuogu OG, Johnson R, Martin FJ, Martínez L, Mohanan S, Muir P, Navarro FCP, Parker A, Pei B, Pozo F, Riera FC, Ruffier M, Schmitt BM, Stapleton E, Suner M-M, Sycheva I, Uszczynska-Ratajczak B, Wolf MY, Xu J, Yang Y T, Yates A, Zerbino D, Zhang Y, Choudhary J S, Gerstein M, Guigó R, Hubbard TJP,

Kellis M, Paten B, Tress ML, Flicek P. 2021. GENCODE 2021. *Nucleic Acids Research* **49(D1)**:D916–D923 DOI 10.1093/nar/gkaa1087.

Fulcoli FG, Franzese M, Liu X, Zhang Z, Angelini C, Baldini A. 2016. Rebalancing gene haploinsufficiency in vivo by targeting chromatin. *Nature Communications* **7(1)**:11688 DOI 10.1038/ncomms11688.

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *TRENDS in Genetics* **23(6)**:273–277 DOI 10.1016/j.tig.2007.03.011.

Glaeser AB, Diniz BL, Santos AS, Guaraná BB, Muniz VF, Carlotto BS, Everling EM, Noguchi PY, Garcia AR, Miola J, Riegel M, Mergener R, Gazzola Zen PR, Machado Rosa RF. 2021. A child with cat-eye syndrome and oculo-auriculo-vertebral spectrum phenotype: a discussion around molecular cytogenetic findings. *European Journal of Medical Genetics* **64(11)**:104319 DOI 10.1016/j.ejmg.2021.104319.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.

Hawkins JA, Kaczmarek ME, Müller MA, Drosten C, Press WH, Sawyer SL. 2019. A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species. *Proceedings of the National Academy of Sciences of the United States of America* **116(23)**:11351–11360 DOI 10.1073/pnas.1814995116.

Hedrick BP, Dumont ER. 2018. Putting the leaf-nosed bats in context: a geometric morphometric analysis of three of the largest families of bats. *Journal of Mammalogy* **99(5)**:1042–1054 DOI 10.1093/jmammal/gyy101.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Research* **17(12)**:1837–1849 DOI 10.1101/gr.6249707.

Hufton AL, Mathia S, Braun H, Georgi U, Lehrach H, Vingron M, Poustka AJ, Panopoulou G. 2009. Deeply conserved chordate noncoding sequences preserve genome synteny but do not drive gene duplicate retention. *Genome Research* **19(11)**:2036–2051 DOI 10.1101/gr.093237.109.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409(6822)**:860–921 DOI 10.1038/35057062.

Ivins S, Scambler P. 2022. Embryonic development in 22q11.2 deletion syndrome. In: McDonald-McGinn DM, ed. *The Chromosome 22q11.2 Deletion Syndrome*. Amsterdam, Netherlands: Elsevier, 54–76.

Jadeja S, Smyth I, Pitera JE, Taylor MS, Van Haelst M, Bentley E, McGregor L, Hopkins J, Chalepakis G, Philip N, Perez Aytes A, Watt FM, Darling SM, Jackson I, Woolf AS, Scrambler PJ. 2005. Identification of a new gene mutated in Fraser syndrome and mouse myelencephalic blebs. *Nature Genetics* **37(5)**:520–525 DOI 10.1038/ng1549.

Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, Winkler S, Jermiin LS, Skirmuntt EC, Katzourakis A, Burkitt-Gray L, Ray DA, Sullivan KAM, Roscito JG, Kirilenko BM, Dávalos LM, Corthals AP, Power ML, Jones G, Ransome RD, Dechmann DKN, Locatelli AG, Puechmaille SJ, Fedrigo O, Jarvis ED, Hiller M, Vernes SC, Myers EW, Teeling EC. 2020. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583(7817)**:578–584 DOI 10.1038/s41586-020-2486-3.

Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AF. 2016. Ancestral reconstruction. *PLOS Computational Biology* **12(7)**:e1004763 DOI 10.1371/journal.pcbi.1004763.

Kalbfleisch TS, Rice ES, DePriest MS Jr., Walenz BP, Hestand MS, Vermeesch JR, O′ Connell BL, Fiddes IT, Vershinina AO, Saremi NF, Petersen JL, Finno CJ, Bellone RR, McCue ME, Brooks SA, Bailey E, Orlando L, Green RE, Miller DC, Antczak DF, MacLeod JN. 2018. Improved reference genome for the domestic horse increases assembly

contiguity and composition. *Communications Biology* **1(1)**:197
DOI 10.1038/s42003-018-0199-z.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple
sequence alignment based on fast fourier transform. *Nucleic Acids Research* **30**:3059–3066
DOI 10.1093/nar/gkf436.

**Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD,
Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW, Muse SV. 2020.**
HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies.
Molecular Biology and Evolution **37(1)**:295–299 DOI 10.1093/molbev/msz197.

Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008.
Patterns of positive selection in six mammalian genomes. *PLOS Genetics* **4**:e1000144
DOI 10.1371/journal.pgen.1000144.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*
9:357–359 DOI 10.1038/nmeth.1923.

**Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, Hinrichs AS, Lee BT,
Nassar LR, Powell CC, Raney BJ, Rosenbloom KR, Schmelter D, Speir ML, Zweig AS,
Haussler D, Haeussler M, Kuhn RM, Kent WJ. 2020.** UCSC Genome Browser enters 20th year.
Nucleic Acids Research **48**:D756–D761 DOI 10.1093/nar/gkz1012.

**Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R,
Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IV, Hackett KJ, Haussler D,
Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, van Sluys M-A, Soltis PS,
Xu X, Yang H, Zhang G. 2018.** Earth biogenome project: sequencing life for the future of life.
Proceedings of the National Academy of Sciences of the United States of America
115(17):4325–4333 DOI 10.1073/pnas.1720115115.

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
1000 Genome Project Data Processing Subgroup. 2009.** The sequence alignment/map format
and SAMtools. *Bioinformatics* **25(16)**:2078–2079 DOI 10.1093/bioinformatics/btp352.

**Liao J, Kochilas L, Nowotschin S, Arnold JS, Aggarwal VS, Epstein JA, Brown MC, Adams J,
Morrow BE. 2004.** Full spectrum of malformations in velo-cardio-facial syndrome/DiGeorge
syndrome mouse models by altering Tbx1 dosage. *Human Molecular Genetics*
13(15):1577–1585 DOI 10.1093/hmg/ddh176.

**Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J,
Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alföldi J,
Beal K, Chang J, Clawson H, Cuff J, Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ,
Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I,
Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Broad
Institute Sequencing Platform and Whole Genome Assembly Team, Worley KC, Kovar CL,
Muzny DM, Gibbs RA, Baylor College of Medicine Human Genome Sequencing Center
Sequencing Team, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Genome Institute
at Washington University, Birney E, Margulies EH, Herrero J, Green ED, Haussler D,
Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M. 2011.** A high-resolution
map of human evolutionary constraint using 29 mammals. *Nature* **478(7370)**:476–482
DOI 10.1038/nature10530.

**Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, Jurecic V, Ogunrinu G,
Sutherland HF, Scrambler PJ, Bradley A, Baldini A. 2001.** Tbx1 haploinsufficiency in the
DiGeorge syndrome region causes aortic arch defects in mice. *Nature* **410**:97–101
DOI 10.1038/35065105.

McDermid HE, Morrow BE. 2002. Genomic disorders on 22q11. *The American Journal of Human Genetics* **70(5)**:1077–1088 DOI 10.1086/340363.

Mi J, Parthasarathy P, Halliday BJ, Morgan T, Dean J, Nowaczyk MJ, Markie D, Robertson SP, Wade EM. 2020. Deletion of exon 1 in *AMER1* in osteopathia striata with cranial sclerosis. *Genes* **11(12)**:1439 DOI 10.3390/genes11121439.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, BAzter MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Marshall Grave JS, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447(7141)**:167–177 DOI 10.1038/nature05805.

Miura M, Ishiyama A, Nakagawa E, Sasaki M, Kurosawa K, Inoue K, Goto Y. 2020. 13q13.3 microdeletion associated with apparently balanced translocation of 46,XX, t(13) suggests *NBEA* involvement. *Brain and Development* **42(8)**:581–586 DOI 10.1016/j.braindev.2020.05.006.

Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends in Genetics* **19(3)**:128–130 DOI 10.1016/S0168-9525(03)00021-0.

Mulhern MS, Stumpel C, Stong N, Brunner HG, Bier L, Lippa N, Riviello J, Rouhl RP, Kempers M, Pfundt R, Stegmann AP, Kukolich MK, Telegrafi A, Lehman A, Lopez-Rangel E, Houcinat N, Barth M, den Hollander N, Hoffer MJV, Weckhuysen S, Roovers J, Djemie T, Darca D, Ceulemans B, Craiu D, Lemke JR, Korff C, Mefford HC, Meyers CT, Siegler Z, Hiatt SM, Cooper GM, Bebin M, Snijders Blok L, Veenstra-Lnol HE, Baugh EH, Brilstra EH, Volker-Touw CML, van Binsbergen E, Revah-Politi A, Pereira E, McBrian D, Pacault M, Isidor B, La Caignec C, Gilbert-Dussardier B, Bilan F, Heinzen EL, Goldstein DB, Stevens SJC, Sands TT. 2018. *NBEA*: developmental disease gene with early generalized epilepsy phenotypes. *Annals of Neurology* **84(5)**:788–795 DOI 10.1002/ana.25350.

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* **11**:715–724 DOI 10.1093/oxfordjournals.molbev.a040152.

Nora EP, Dekker J, Heard E. 2013. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* **35(9)**:818–828 DOI 10.1002/bies.201300040.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44(D1)**:D733–D745 DOI 10.1093/nar/gkv1189.

Paulat NS, Storer JM, Moreno-Santillan D, Osmanski AB, Sullivan KAM, Grimshaw JR, Korstian JM, Halsey M, Garcia CJ, Crookshanks C, Roberts J, Smit AFA, Hubley R, Rosen J,

Teeling EC, Vernes SC, Hiller M, Myers E, Pippel M, Brown T, Rojas D, Davalos LM, Ray DA. 2023. *Eptesicus fuscus* isolate TK198812, whole genome shotgun sequencing project. *Available at* https://www.ncbi.nlm.nih.gov/nuccore/JANZWY000000000.1.

Pinzari CA, Kang L, Michalak P, Jermiin LS, Price DK, Bonaccorso FJ. 2020. Analysis of genomic sequence data reveals the origin and evolutionary separation of Hawaiian hoary bat populations. *Genome Biology and Evolution* **12(9)**:1504–1514 DOI 10.1093/gbe/evaa137.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JB. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344(6190)**:1410–1414 DOI 10.1126/science.1253226.

Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B, Volfovsky N, Schäffer AA, Agarwala R, Narfström K, Murphy WJ, Giger U, Roca AL, Antunes A, Menotti-Raymond M, Yuhki N, Pecon-Slattery J, Johnson WE, Bourque G, Tesler G, NISC Comparative Sequencing Program, O'Brien SJ. 2007. Initial sequence and comparative analysis of the cat genome. *Genome Research* **17(11)**:1675–1689 DOI 10.1101/gr.6380007.

Pourhaghighi R, Ash PE, Phanse S, Goebels F, Hu LZ, Chen S, Zhang Y, Wierbowski SD, Boudeau S, Moutaoufik MT, Malty RH, Malolepsza E, Tsafou K, Nathan A, Cromar G, Guo H, Al Abdullatif A, Appico DJ, Becker LA, Gitler AD, Pulst SM, Youssef A, Hekman R, Havugimana PC, White CA, Blum BC, Ratti A, Bryant CD, Parkinson J, Lage K, Babu M, Yu H, Bader GD, Wolozin B, Emili A. 2020. BraInMap elucidates the macromolecular connectivity landscape of mammalian brain. *Cell Systems* **10(4)**:333–350 DOI 10.1016/j.cels.2020.03.003.

Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. 2016. Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes & Development* **30(2)**:191–207 DOI 10.1101/gad.272187.115.

R Core Team R. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. *Available at* https://www.r-project.org.

Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, Webster MT. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365(1552)**:2571–2580 DOI 10.1098/rstb.2010.0007.

Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Computers & Chemistry* **26(1)**:51–56 DOI 10.1016/S0097-8485(01)00099-7.

Rodriguez-Zabala M, Aza-Carmona M, Rivera-Pedroza CI, Belinchón A, Guerrero-Zapata I, Barraza-García J, Vallespin E, Lu M, del Pozo A, Glucksman MJ, Santos-Simarro F, Heath KE. 2017. FGF9 mutation causes craniosynostosis along with multiple synostoses. *Human Mutation* **38(11)**:1471–1476 DOI 10.1002/humu.23292.

Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, Rowan TN, Low WY, Zimin A, Couldrey C, Hall R, Li W, Rhie A, Ghurye J, McKay SD, Thibaud-Nissen F, Hoffman J, Murdoch BM, Snelling WM, McDaneld TG, Hammond JA, Schwartz JC, Nandolo W, Hagen DE, Dreischer C, Schultheiss SJ, Schroeder SG, Phillippy AM, Cole JB, Van Tassell CP, Liu G, Smith TPL, Medrano JF. 2020. *De novo* assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9(3)**:giaa021 DOI 10.1093/gigascience/giaa021.

Salah A, Almannai M, Al Ojaimi M, Radefeldt M, Gulati N, Iqbal M, Alawbathani S, Al-Ali R, Beetz C, El-Hattab AW. 2022. A homozygous frame-shift variant in *PROSER1* is associated with developmental delay, hypotonia, genitourinary malformations, and distinctive facial features. *Clinical Genetics* **101**:565–570 DOI 10.1111/cge.14126.

Seo TK, Kishino H, Thorne JL. 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Molecular Biology and Evolution* **21(7)**:1201–1213 DOI 10.1093/molbev/msh088.

Shen Y-Y, Liang L, Zhu Z-H, Zhou W-P, Irwin DM, Zhang Y-P. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proceedings of the National Academy of Sciences of the United States of America* **107(19)**:8666–8671 DOI 10.1073/pnas.0912613107.

Sjodin BM, Galbreath KE, Lanier HC, Russello MA. 2021. Chromosome-level reference genome assembly for the American pika (*Ochotona princeps*). *Journal of Heredity* **112(6)**:549–557 DOI 10.1093/jhered/esab031.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* **32(5)**:1342–1353 DOI 10.1093/molbev/msv022.

Tan T, Gordon C, Amor D, Farlie P. 2010. Developmental perspectives on copy number abnormalities of the 22q11.2 region. *Clinical Genetics* **78(3)**:201–218 DOI 10.1111/j.1399-0004.2010.01456.x.

Tang J, Lin Y, Zhang Z, Tikunova S, Birnbaumer L, Zhu MX. 2001. Identification of common binding sites for calmodulin and inositol 1,5-trisphosphate receptors on the carboxyl termini of Trp channels. *Journal of Biological Chemistry* **276(24)**:21303–21310 DOI 10.1074/jbc.M102316200.

Tang L, Wu X, Zhang H, Lu S, Wu M, Shen C, Chen X, Wang Y, Wang W, Shen Y, Gu M, Ding X, Jin X, Fei J, Wang Z. 2017. A point mutation in Fgf9 impedes joint interzone formation leading to multiple synostoses syndrome. *Human Molecular Genetics* **26(7)**:1280–1293 DOI 10.1093/hmg/ddx029.

The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, Hill DP, Lee R, Mi M, Moxon S, Mungall CJ, Muruganugan A, Mushayahama T, Sternberg PW, Thomas PD, Van Auken K, Ramsey J, Siegele DA, Chisholm RL, Fey P, Aspromonte MC, Nugnes MV, Quaglia F, Tosatto S, Giglio M, Nadendla S, Antonazzo G, Attrill H, dos Santos G, Marygold S, Strelets V, Tabone CJ, Thurmond J, Zhou P, Ahmed SH, Asanitthong P, Luna Buitrago D, Erdol MN, Gage MC, Ali Kadhum M, Li KYC, Long M, Michalak A, Pesala A, Pritazahra A, Saverimuttu SCC, Su R, Thurlow KE, Lovering RC, Logie C, Oliferenko S, Blake J, Christie K, Corbani L, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Smith C, Cuzick A, Seager J, Cooper L, Elser J, Jaiswal P, Gupta P, Jaiswal P, Naithani S, Lera-Ramirez M, Rutherford K, Wood V, De Pons JL, Dwinell MR, Hayman GT, Kaldunski ML, Kwitek AE, Laulederkind SJF, Tutaj MA, Vedi M, Wang SJ, D'Eustachio P, Aimo L, Axelsen K, Bridge A, Hyka-Nouspikel N, Morgat A, Aleksander SA, Cherry JM, Engel SR, Karra K, Miyasato SR, Nash RS, Skrzypek MS, Weng S, Wong ED, Bakker E, Berardini TZ, Reiser L, Auchincloss A, Axelsen K, Argoud-Puy G, Blatter MC, Boutet E, Breuza L, Bridge A, Casals-Casas C, Coudert E, Estreicher A, Livia Famiglietti M, Feuermann M, Gos A, Gruaz-Gumowski N, Hulo C, Hyka-Nouspikel N, Jungo F, Le Mercier P, Lieberherr D, Masson P, Morgat A, Pedruzzi I, Pourcel L, Poux S, Rivoire C, Sundaram S, Bateman A, Bowler-Barnett E, Bye-A-Jee H, Denny P, Ignatchenko A, Ishtiaq R, Lock A, Lussi Y, Magrane M, Martin MJ,

Orchard S, Raposo P, Speretta E, Tyagi N, Warner K, Zaru R, Diehl AD, Lee R, Chan J, Diamantakis S, Raciti D, Zarowiecki M, Fisher M, James-Zorn C, Ponferrada V, Zorn A, Ramachandran S, Ruzicka L, Westerfield M. 2023. The gene ontology knowledgebase in 2023. *Genetics* **224(1)**:iyad031 DOI 10.1093/genetics/iyad031.

Timmer JR, Mak TW, Manova K, Anderson KV, Niswander L. 2005. Tissue morphogenesis and vascular stability require the Frem2 protein, product of the mouse myelencephalic blebs gene. *Proceedings of the National Academy of Sciences of the United States of America* **102(33)**:11746–11750 DOI 10.1073/pnas.0505404102.

Wang K, Tian S, Galindo-González J, Dávalos LM, Zhang Y, Zhao H. 2020. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Molecular Ecology* **29(22)**:4366–4381 DOI 10.1111/mec.15542.

Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, Giron CG, Griffin DK, Hall R, Hannum G, Hourlier T, Howe K, Hume DA, Izuogu O, Kim K, Koren S, Liu H, Manchanda N, Martin FJ, Nonneman DJ, O'Connor RE, Philippy AM, Rohrer GA, Rosen BD, Rund LA, Sargent CA, Schook LB, Schroeder SG, Schwartz AS, Skinner BM, Talkbot R, Tseng E, Tuggle CK, Watson M, Smith TPL, Archibald AL. 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **9(6)**:giaa051 DOI 10.1093/gigascience/giaa051.

Wei L, Liu Y, Dubchak I, Shon J, Park J. 2002. Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics* **35(2)**:142–150 DOI 10.1016/S1532-0464(02)00506-3.

Wu X-L, Gu M-M, Huang L, Liu X-S, Zhang H-X, Ding X-Y, Xu J-Q, Cui B, Wang L, Lu S-Y, Chen X-Y, Zhang H-G, Huang W, Yuan W-T, Yang J-M, Gu Q, Fei J, Chen Z, Yuan Z-M, Wang Z-G. 2009. Multiple synostoses syndrome is due to a missense mutation in exon 2 of *FGF9* gene. *The American Journal of Human Genetics* **85(1)**:53–63 DOI 10.1016/j.ajhg.2009.06.007.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21(9)**:1859–1875 DOI 10.1093/bioinformatics/bti310.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24(8)**:1586–1591 DOI 10.1093/molbev/msm088.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**:32–43 DOI 10.1093/oxfordjournals.molbev.a026236.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, Shen Y, Pervouchine DD, Djebali S, Thurman RE, Kaul R, Rynes E, Kirilusha A, Marinov GK, Williams BA, Trout D, Amrhein H, Fisher-Aylor K, Antoshechkin I, DeSalvo G, See L-H, Fastuca M, Drenkow J, Zaleski C, Dobin A, Prieto P, Lagarde J, Bussotti G, Tanzer A, Denas O, Li K, Bender MA, Zhang M, Byron R, Groudine MT, McCleary D, Pham L, Ye Z, Kuan S, Edsall L, Wu Y-C, Rasmussen MD, Bansal MS, Kellis M, Keller CA, Morrissey CS, Mishra T, Jain D, Dogan N, Harris RS, Cayting P, Kawli T, Boyle AP, Euskirchen G, Kundaje A, Lin S, Lin Y, Jansen C, Malladi VS, Cline MS, Erickson DT, Kirkup VM, Learned K, Sloan CA, Rosenbloom KR, Lacerda de Sousa B, Beal K, Pignatelli M, Flicek P, Lian J, Kahveci T, Lee D, Kent WJ, Ramalho Santos M, Herrero J, Notredame C, Johnson A, Vong S, Lee K, Bates D, Neri F, Diegel M, Canfield T, Sabo PJ, Wilken MS, Reh TA, Giste E, Shafer A, Kutyavin T, Haugen E, Dunn D, Reynolds AP, Neph S, Humbert R, Hansen RS, De Bruijn M, Selleri L, Rudensky A, Josefowicz S, Samstein R, Eichler EE, Orkin SH, Levasseur D, Papayannopoulou T, Chang K-H, Skoultchi A, Gosh S, Disteche C, Treuting P, Wang Y, Weiss MJ, Blobel GA,

Cao X, Zhong S, Wang T, Good PJ, Lowdon RF, Adams LB, Zhou X-Q, Pazin MJ, Feingold EA, WOld B, Taylor J, Mortazavi A, Weissman SM, Stamatoyannopolous JA, Snyder MP, Guigo R, Gingeras TR, Gilbert DM, Hardison RC, BEer MA, Ren B, The Mouse ENCODE Consortium. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515(7527)**:355–364 DOI 10.1038/nature13992.

Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, Wynne JW, Xiong Z, Baker ML, Zhao W, Tachedjian M, Zhu Y, Zhour P, Jiang X, Ng J, Yang L, Wu L, Xiao J, Feng Y, Chen Y, Sun X, Zhang Y, Marsh GA, Crameri G, Broder CC, Frey KG, Wang L-F, Wang J. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339(6118)**:456–460 DOI 10.1126/science.1230835.

Zhou Q, Fu X, Xu J, Dong S, Liu C, Cheng D, Gao C, Huang M, Liu Z, Ni X, Hua R, Tu H, Sun H, Shen Q, Chen B, Zhang J, Zhang L, Yang H, Hu J, Yang W, Pei W, Yao Q, Sheng X, Zhang J, Yang WZ, Shen WL. 2023. Hypothalamic warm-sensitive neurons require *TRPC4* channel for detecting internal warmth and regulating body temperature in mice. *Neuron* **111**:387–404 DOI 10.1016/j.neuron.2022.11.008.

Zhou Y, Shearwin-Whyatt L, Li J, Song Z, Hayakawa T, Stevens D, Fenelon JC, Peel E, Cheng Y, Pajpach F, Bradley N. 2021. Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* **592(7856)**:756–762 DOI 10.1038/s41586-020-03039-0.