Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*

Inge Seim[1,2,*], Xiaodong Fang[3,4,*], Zhiqiang Xiong[3], Alexey V. Lobanov[1], Zhiyong Huang[3], Siming Ma[1], Yue Feng[3], Anton A. Turanov[1], Yabing Zhu[3], Tobias L. Lenz[1], Maxim V. Gerashchenko[1,5], Dingding Fan[3], Sun Hee Yim[1], Xiaoming Yao[3], Daniel Jordan[1], Yingqi Xiong[3], Yong Ma[3], Andrey N. Lyapunov[6], Guanxing Chen[3], Oksana I. Kulakova[7], Yudong Sun[3], Sang-Goo Lee[2], Roderick T. Bronson[8], Alexey A. Moskalev[7,9,10], Shamil R. Sunyaev[1], Guojie Zhang[3], Anders Krogh[4], Jun Wang[3,4,11,#], Vadim N. Gladyshev[1,2,#]

[1] Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA

[2] Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, South Korea

[3] BGI-Shenzhen, Shenzhen, 518083, China

[4] Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark

[5] University of Nebraska, Lincoln, NE, 68588, USA

[6] Kirov State Center for Distance Education of Children, Kirov, 610006, Russia

[7] Institute of Biology, Komi Science Center, Russian Academy of Sciences, Syktyvkar, 167982, Russia

[8] Rodent Histopathology Laboratory, Harvard Medical School, Boston, MA 02115, USA

[9] Syktyvkar State University, Syktyvkar, 167001, Russia

[10] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, 141700, Russia

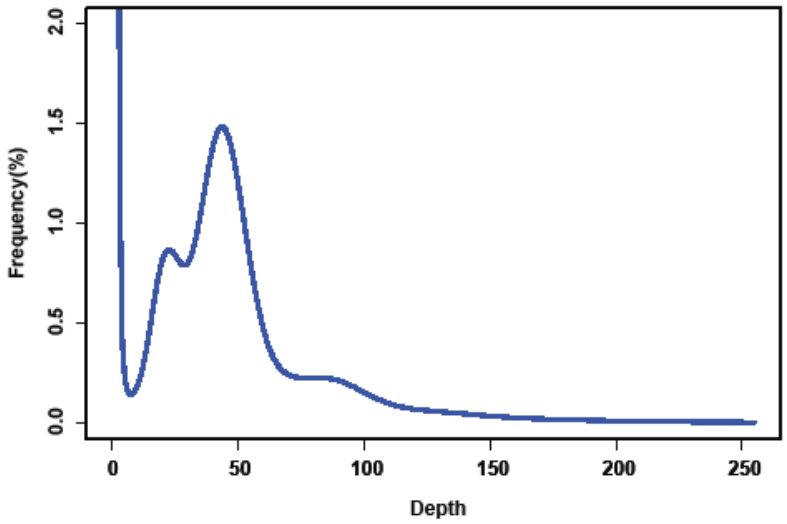
[11] Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark

[*] equal contribution

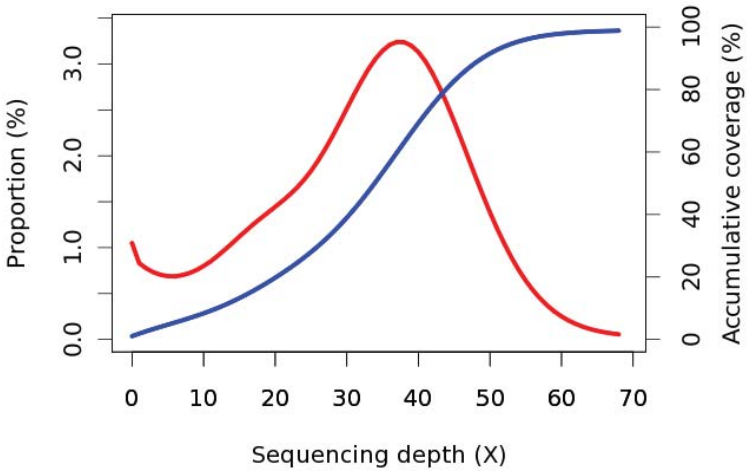
corresponding authors. Contact information: vgladyshev@rics.bwh.harvard.edu

SUPPLEMENTARY FIGURES

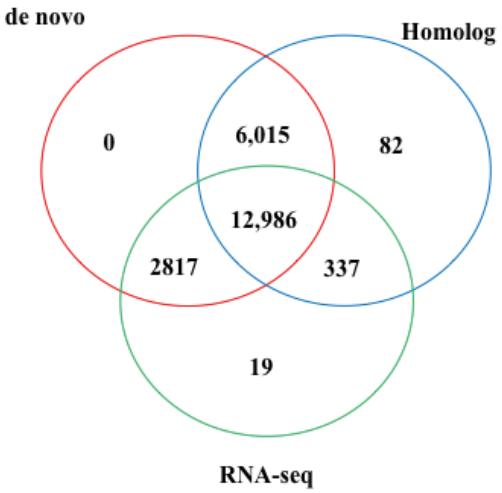
Supplementary Figure S1 | Hibernating *M. brandtii* at sampling site, Obvalnaya cave, Russia.



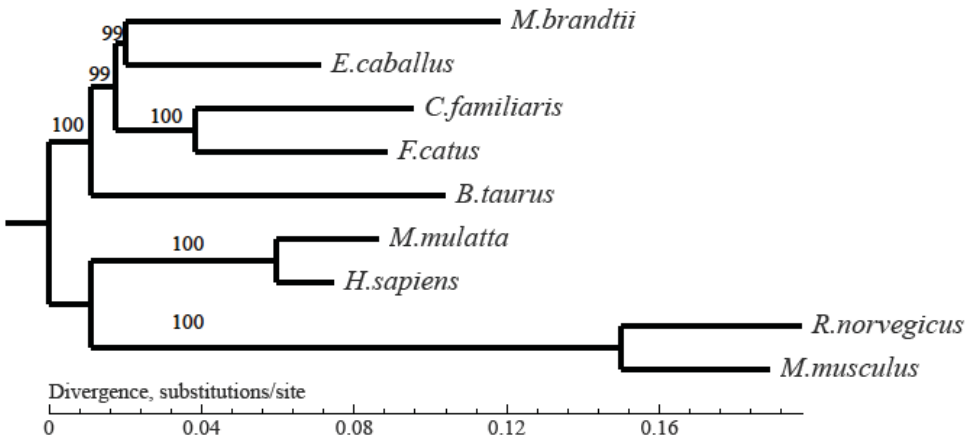
Supplementary Figure S2 | 17-k-mer estimation of genome size. Genome size of *M. brandtii* was estimated to be 1.9 Gb based on reads from short insert size libraries.



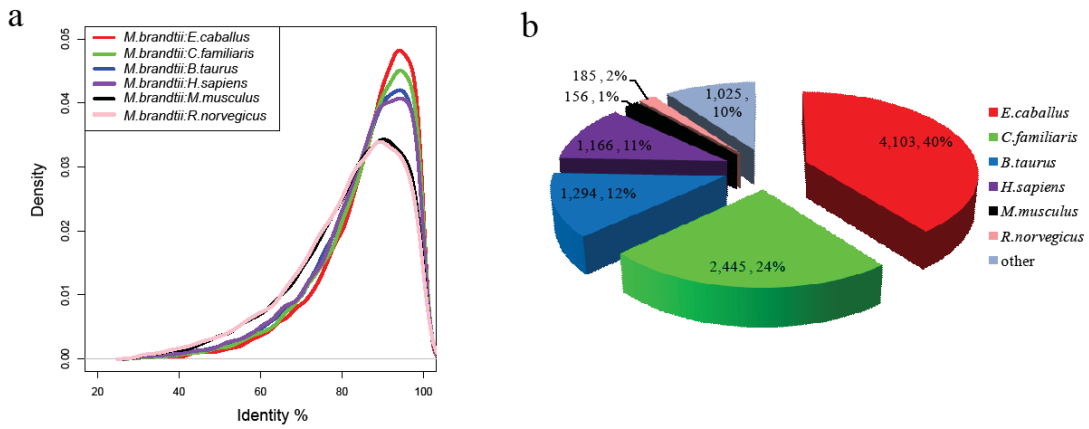
Supplementary Figure S3 | Sequencing depth distribution of *M. brandtii* genome. High quality short insert size reads were mapped to the associated genome with an average depth of 34, and approximately 91.64% of the genome was covered by more than 10 reads. The red curve denotes the proportion of the genome in a given sequencing depth, and the blue curve shows the cumulative coverage of the genome.



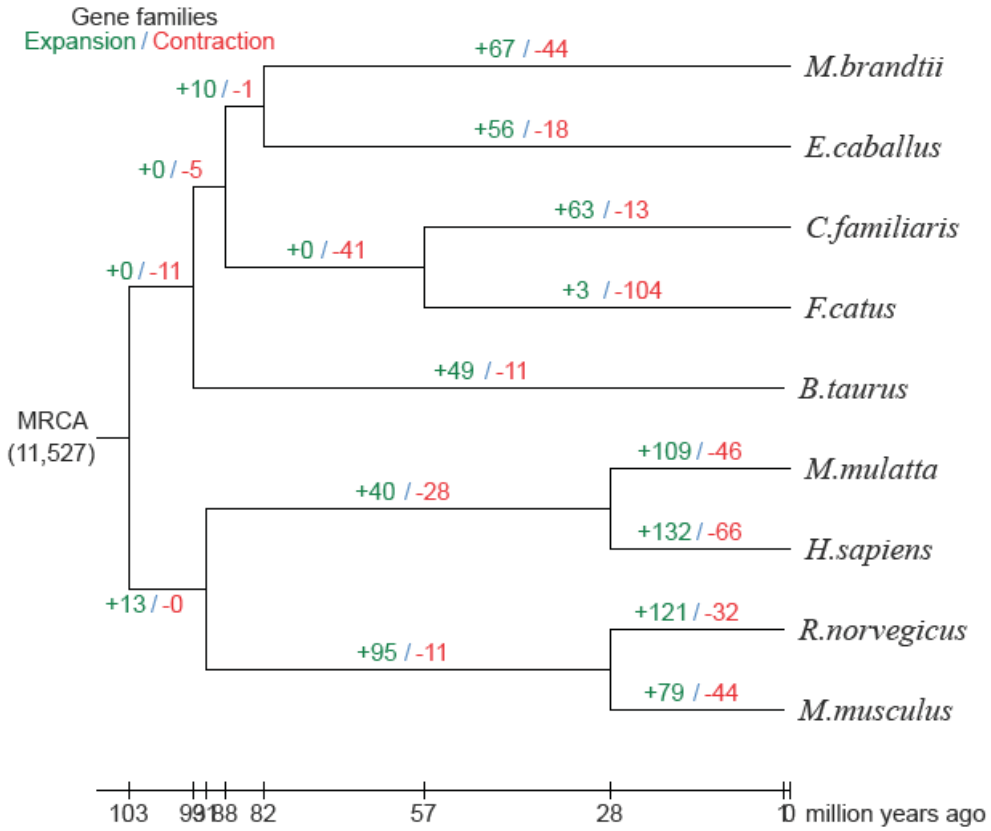
Supplementary Figure S4 | Evidence for gene models from three types of gene prediction sources. If a gene model predicted by different methods showed more than 50% overlap, multiple evidence support was called. Red, blue and green circles represent *de novo*-, homology-, and evidence-based (transcriptomic) gene predictions, respectively.



Supplementary Figure S5 | Phylogenetic tree of 9 mammalian species. Trees were constructed with PhyML under GTR +gamma model based on total CDS. The bootstrap values were calculated based on 1,000 replications.



Supplementary Figure S6 | *M. brandtii* proteins compared to orthologs in related mammals. **(a)** Identity distribution of *M. brandtii* proteins to other species. Bat/Horse has the highest protein similarity. **(b)** Pie diagram shows the number and proportion of genes that share the highest protein identity between *M. brandtii* and related mammals, and horse has the highest gene number and proportion (1:1 ortholog were used for this analysis, "other" means more than one species share the highest identity with *M. brandtii*).

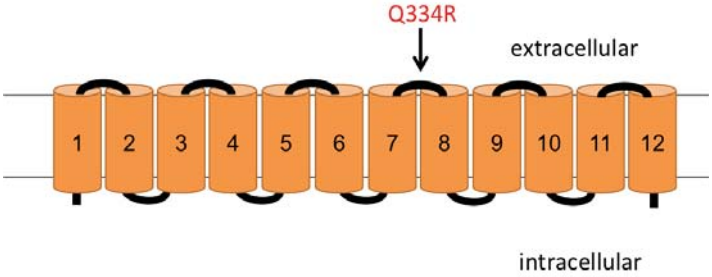


Supplementary Figure S7 | Expanded and contracted gene families. Gene family expansion events are shown in green, and gene family contraction events in red. MRCA refers to the gene family number of the most recent common ancestor.

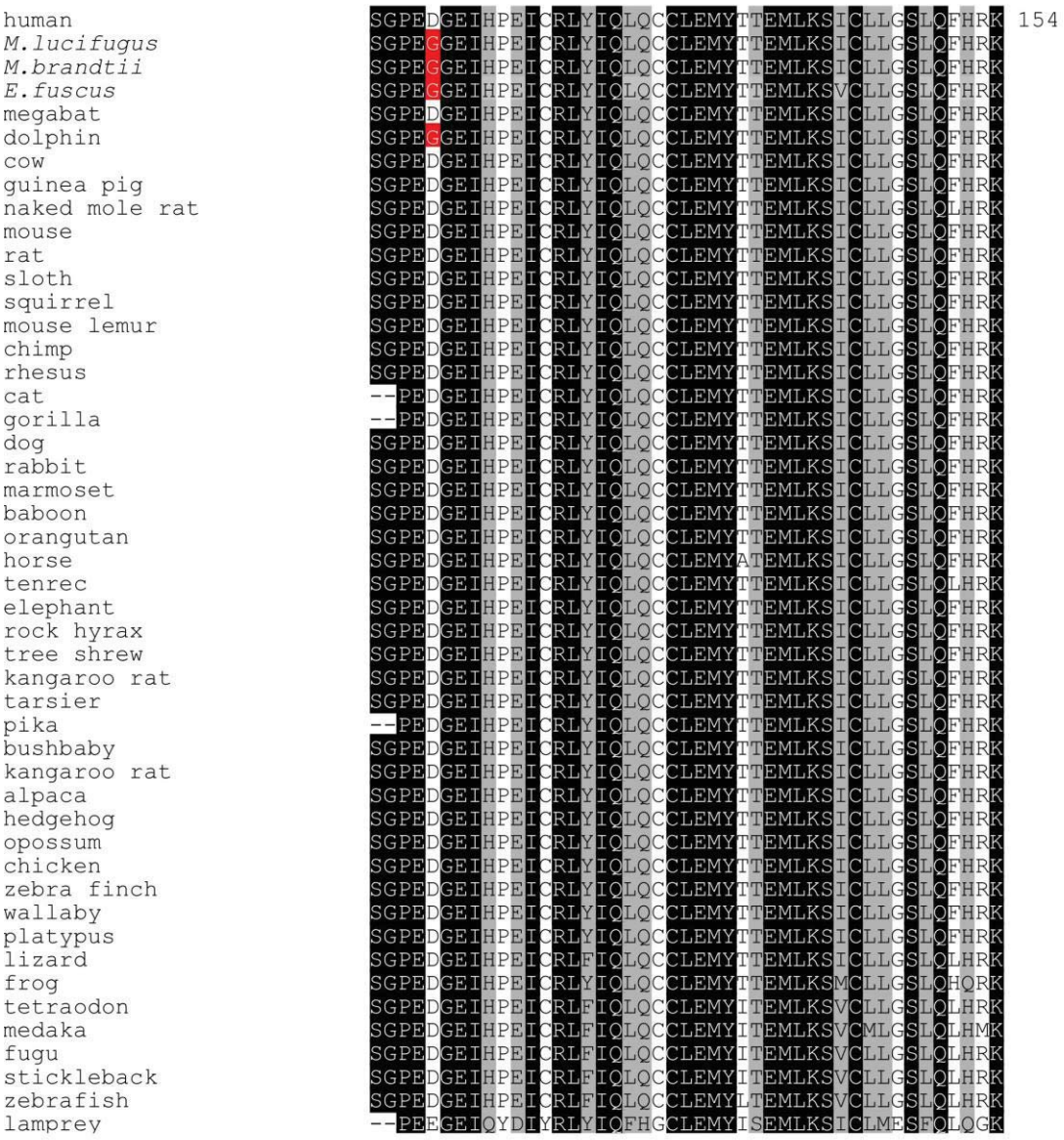
Supplementary Figure S8 | *GULO* is a pseudogene in *M. brandtii*. Nucleotide sequences of exon 4 of *GULO* in *M. lucifugus* (GenBank accession code JX259509) and *M. brandtii*. Identical nucleotide residues are shaded in black and premature stop codons in red. Corresponding amino acid residues are shown above and below their respective alignments.

a

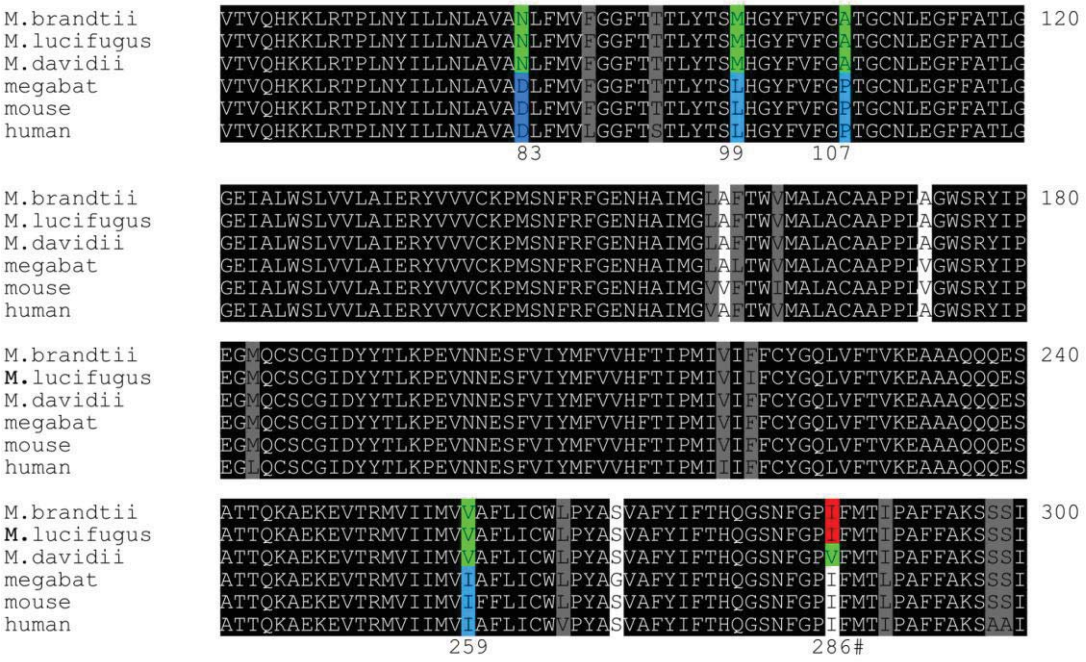
b



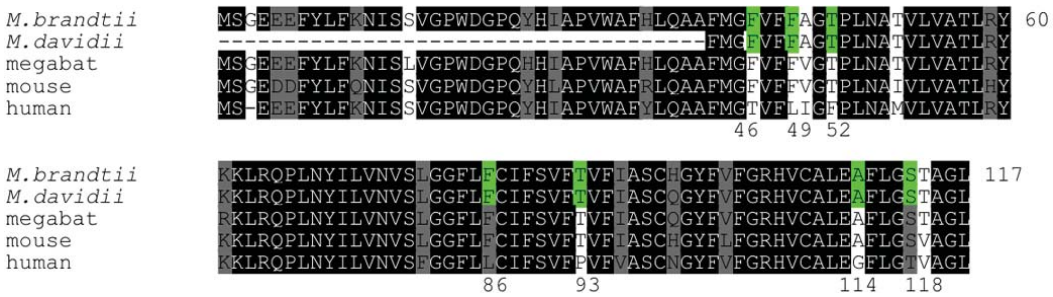
Supplementary Figure S9 | Shared unique residue in SLC45A2 of echolocating bats and dolphin. (a) Alignment of vertebrate SLC45A2 protein sequences. Transmembrane domains are highlighted in orange and extracellular regions in black. Unique amino acids in echolocating bats and the Atlantic bottlenose dolphin are highlighted in red. **(b)** Model of SLC45A2 subunit showing the approximate position of the novel amino acid (in red). Transmembrane sequences are shown in orange.



Supplementary Figure S10 | Shared unique residue in RGS7BP of echolocating bats and dolphin. Alignment of vertebrate RGS7BP-gene encoded sequences. Novel amino acid changes are shown in red, and conserved residues are highlighted in black.

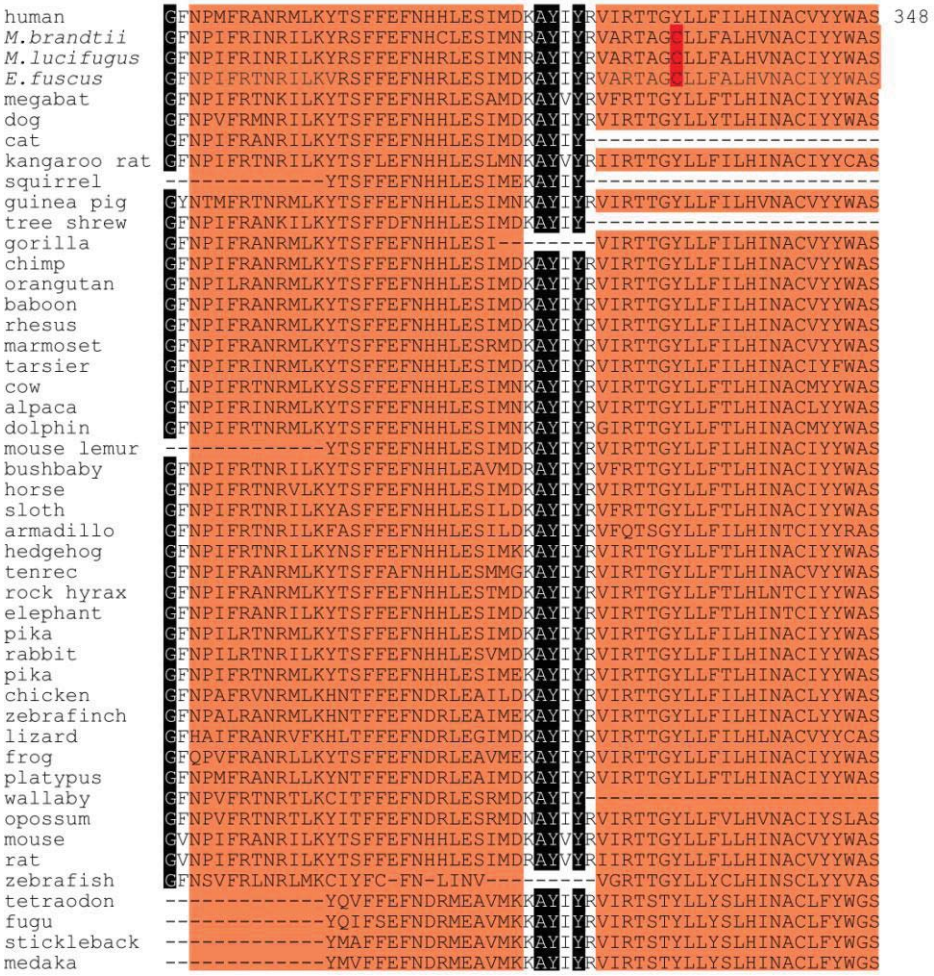


Supplementary Figure S11 | Comparison of bat, mouse and human Rhodopsin/*RHO* sequences. Residues corresponding to a previous study[2] that examined Rhodopsin in the bat genus *Myotis* are highlighted in green, while conflicting residues are shown in red. # indicates a residue previously thought to be uniquely changed in *Myotis* by analyzing *M. laniger*, *M. davidii* and *M. ricketti* sequences[2].

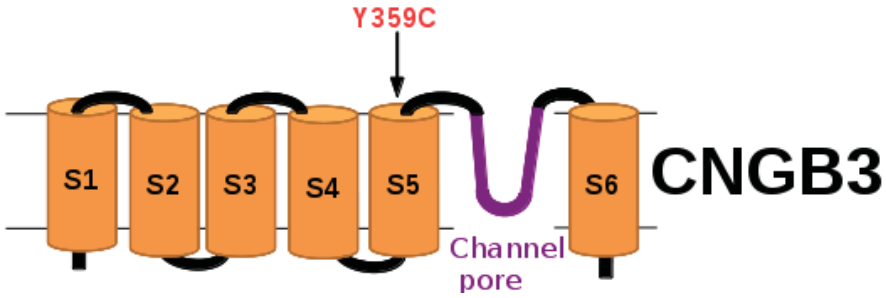


Supplementary Figure S12 | Comparison of bat, mouse and human S opsin/*OPN1SW* sequences. Residues that alter the tuning of short-wavelength sensitive opsin/*OPN1SW* (S opsin) are numbered and the corresponding *Myotis* residues are highlighted in green.

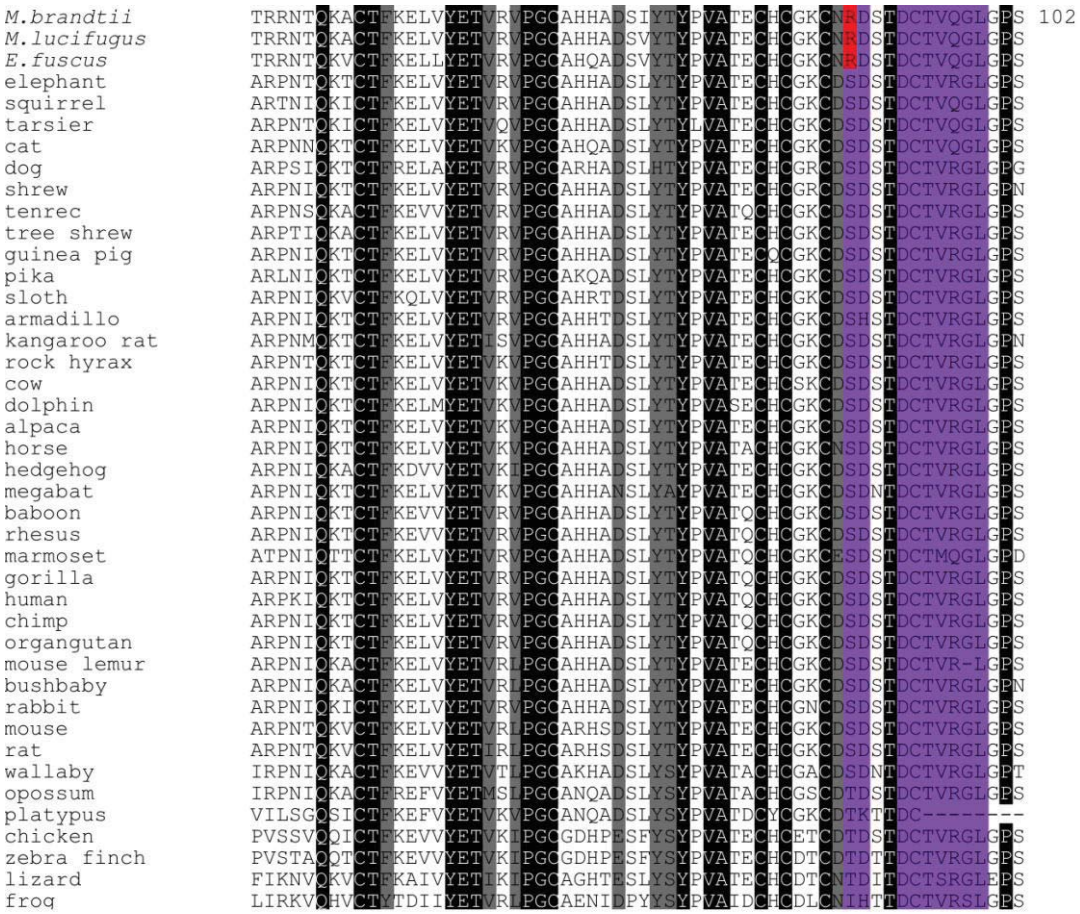
a



b

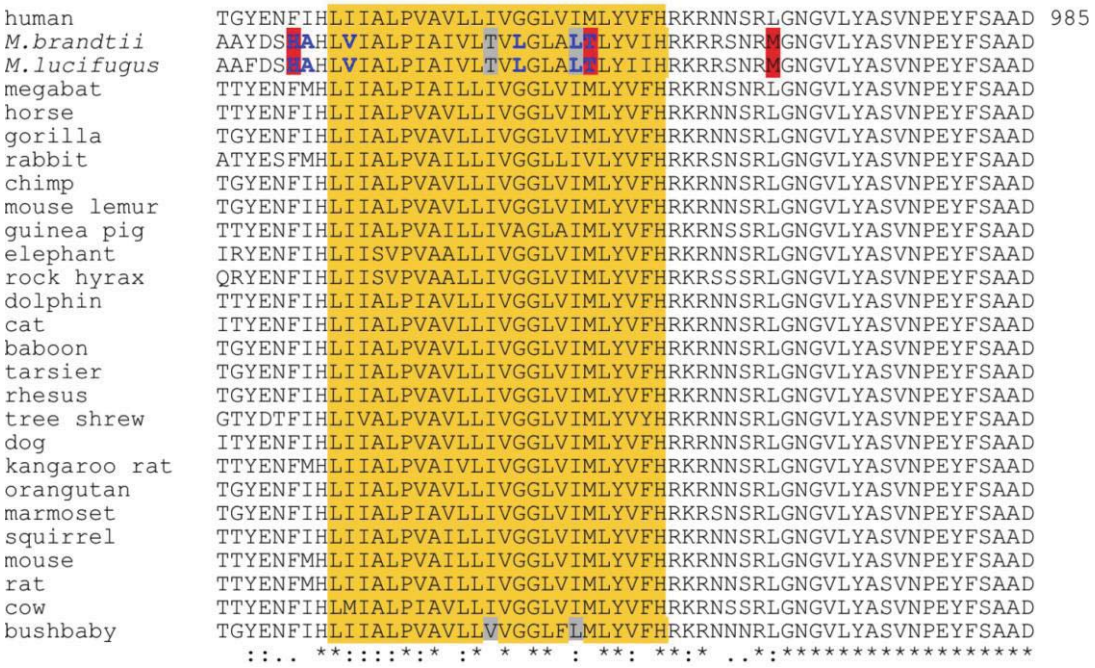


Supplementary Figure S13 | Unique amino acid change in CNGB3 of echolocating bats. (a) Alignment of vertebrate CNGB3 protein sequences. Transmembrane regions are highlighted in orange and unique amino acids in echolocating bats in red. **(b)** Model of CNG subunit showing the approximate position of the novel CNGB3 amino acids (in red). The channel pore is shown in purple and transmembrane domains 1 to 6 (S1-S6) in orange.

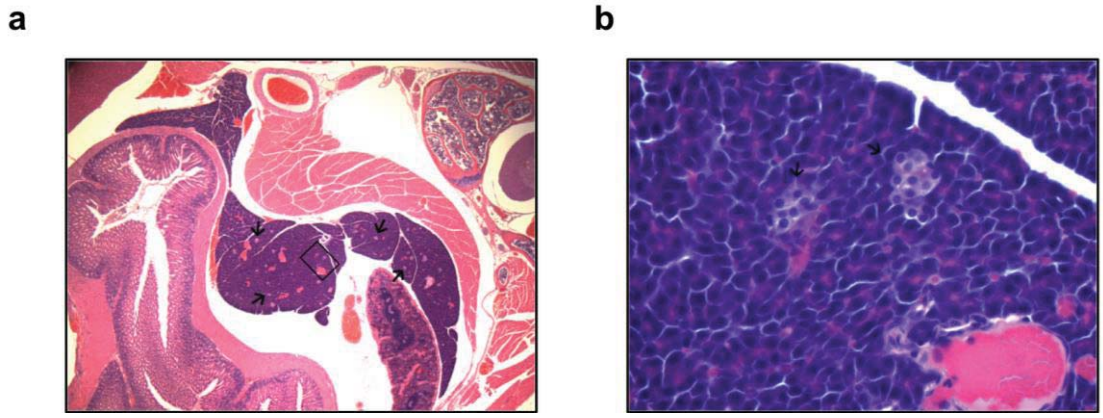


Supplementary Figure S14 | Alignment of vertebrate FSHβ sequences. Alignment of vertebrate FSHB-gene encoded sequences. FSHβ residues buried at the FSHR receptor-ligand interface are highlighted in purple, while radical amino acid substitutions within the interface in bats with delayed ovulation are shown in red. Amino acid numbering corresponds to the mature FSHβ peptide.

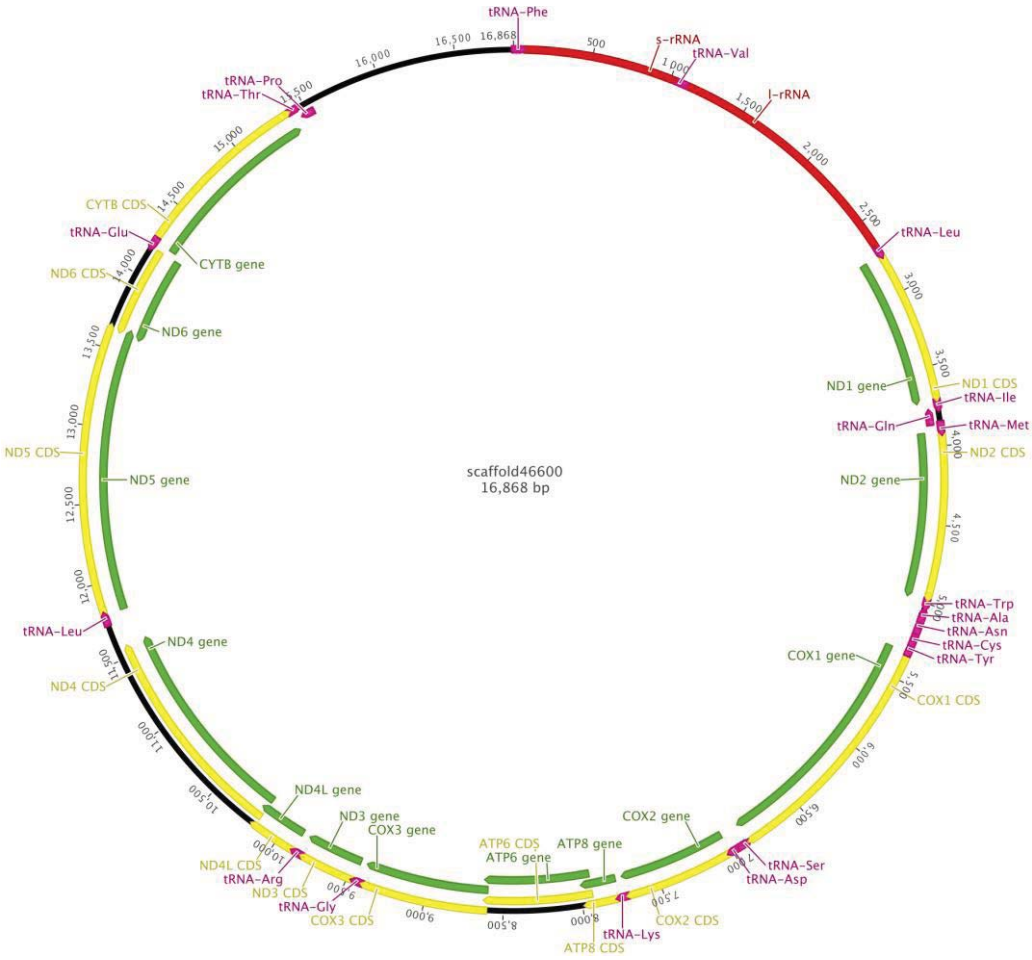
Supplementary Figure S15 | Unique amino acid deletion in Vespertilionoidea bat GHR. Alignment of vertebrate *GHR*-encoded sequences. An amino acid residue deleted in the superfamily Vespertilionoidea is shown in red.



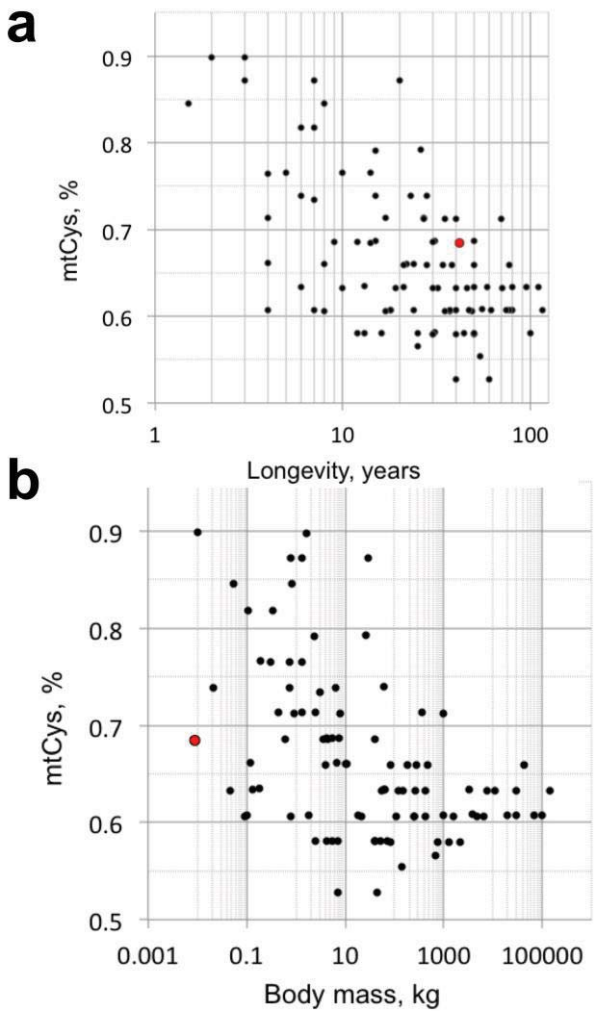
Supplementary Figure S16 | Unique amino acid residues in Vespertilionoidea bat IGF1R. Alignment of mammalian *IGF1R*-encoded sequences is shown. Transmembrane domains are highlighted in orange and unique amino acids in *Myotis* in red. Blue font indicates amino acid sites under positive selection in bats in the genus *Myotis* ($P < 0.05$ and with a Bayes empirical bayes (BEB) probability >95% under the branch-site model in CodeML (n=27) with a guide tree predicted by PRANK and *M. brandtii* and *M. lucifugus* examined separately as foreground branches). A star (*) indicates a highly conserved amino acid. Residues with similar charge properties are indicated by a colon (:). Amino acids unique to the *Myotis* and conserved in all other mammals except the candidate low-coverage bushbaby genome (1.9X) ortholog are highlighted in grey.



Supplementary Figure S17 | Diminished size of islets of Langerhans in *M. brandtii.* The pancreas (arrows) stains more deeply with Hematoxylin than do pancreata from other animals since there are fewer xymogen granules in exocrine cells. The islets are tiny and cannot be seen at this magnification **(a,** x4**).** At higher magnification **(b,** x60**)** tiny islets with 5-6 cells can be seen.



Supplementary Figure S18 | Physical map of *M. brandtii* mitochondrial DNA. Protein-coding genes and coding regions are shown in green and yellow, respectively. Transfer RNA (tRNA) genes are shown in pink.



Supplementary Figure S19 | Low cysteine content of *M. brandtii* mt-encoded proteins. Cysteine content of proteins encoded by mitochondrial DNA (mtCys, in % of all amino acids encoded by mitochondrial DNA) in 95 mammals is plotted against their maximum lifespan **(a)** and body mass **(b)**. Data were obtained from AnAge. Brandt's bat is shown by a red dot. Negative correlations between mtCys and lifespan, and between mtCys and body mass, suggest that the Brandt's bat's lifespan is slightly above that expected based on its mtCys content, whereas its body mass is substantially lower than that expected from its mtCys content.

SUPPLEMENTARY TABLES

Supplementary Table S1 | Data production.

Pair-end libraries	Insert Size	Total Data (Gb)	Reads Length (bp)	Sequence coverage (X)*	Physical coverage (X)
Illumina Reads	170bp	54.39	91	27.20	25.54
	500bp	22.84	93	11.42	30.69
	800bp	24.01	91	12.01	53.07
	2kb	20.69	49	10.35	211.17
	5kb	17.07	49	8.54	435.49
	10kb	12.35	49	6.17	629.95
	20kb	4.37	49	2.18	445.45
Total	-----	155.72	-----	77.86	1,831.36

*assuming 2 Gb genome size

Supplementary Table S2 | Statistics of the assembly.

Index	Contig		Scaffold	
	Length (bp)	**Number**	**Length (bp)**	**Number**
N90	2,339	121,024	223,054	965
N80	7,295	75,302	962,962	563
N70	11,635	53,238	1,597,504	387
N60	16,175	38,330	2,305,849	273
N50	21,110	27,229	3,111,371	192
Total length	2,052,216,787		2,177,979,242	
Max. length	217,926		22,122,726	
Number>100bp		844,285		677,946
Number>2000bp		126,755		4,835

Supplementary Table S3 | Statistics of high quality short insert size read mapping.

Sample	Genome size (M)	Effective size (M)	Raw Data		Mapped Data					depth	Coverage (M)	%
			Reads (M)	Base (M)	Reads(M)	%	Base (M)	%	PE%*			
Myotis brandtii	2,178	2,052	799	72,876	769	96.28	70,168	96.28	94.39	34.19	2,031	98.97

*PE% indicates the proportion of reads that were proper paired mapping to genome with the reasonable insert size associated with the library we built.

Supplementary Table S4 | Heterozygosity ratio of *M. brandtii*.

Sample	Genome size (Mbp)	Effective size (removing gaps, Mbp)	Sequencing depth (X)	SNP		InDel	
				#Hete	%Hete	#Hete	%Hete
Myotis brandtii	2,178	2,052	34.19	7,312,602	0.371	1,271,833	0.065

Supplementary Table S5 | Comparison of transposable elements in mammalian genomes.

	Myotis brandtii		Human		Macaque		Mouse		Rat		Dog	
	Length (Mp)	% genome	Length (Mp)	% genome	Length (Mp)	% genome	Length (Mp)	% genome	Length (Mp)	% genome	Length (Mp)	% genome
DNA	87.78	4.03	102.4	3.3	80.1	2.58	63.7	2.3	66.8	2.5	68.3	2.7
LINE	250.97	11.52	543	17.5	395.4	12.77	495.3	18.2	529.5	19.5	418.6	16.54
LTR	100.53	4.62	257.2	8.3	192.7	6.22	283.9	10.4	220.8	8.1	114.7	4.5
SINE	32.36	1.49	349.4	11.3	279.2	9.02	166.7	6.1	144.6	5.3	226.7	8.9
Other	0.01	0	26.4	0.9	5.6	0.18	7.6	0.3	6.8	0.2	0.008	0
Unknown	14.75	0.68	4.8	0.2	2.7	0.09	43.7	1.6	50.7	1.9	1.1	0.04
Total	483.06	22.17	1,257.70	40.5	955.7	30.86	1,017	37.4	978	36	822.8	32.5

Supplementary Table S6 | Statistics of gene parameters.

Gene set	Number	complete ORF	%	Single Exon gene	%	Average gene length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
M.brandtii	22,256	21,199	95.25	3,348	15.04	32,464	1,501	8.50	177	4,243
E.caballus	20,383	10,096	49.53	3,323	16.30	32,268	1,532	9.28	165	3,713
C.familiaris	19,281	8,985	46.60	1,208	6.27	30,994	1,578	9.90	159	3,305
M.musculus	22,837	21,214	92.89	4,198	18.38	36,630	1,515	8.56	177	4,646
R.norvegicus	22,841	16,745	73.31	3,552	15.55	30,892	1,452	8.59	169	3,879
B.taurus	19,970	14,629	73.25	2,459	12.31	35,523	1,599	9.59	167	3,949
H.sapiens	22,389	20,098	89.77	3,318	14.82	44,855	1,560	8.96	174	5,436
H.glaber	22,561	19,137	84.82	3,930	17.42	32,533	1,439	8.05	179	4,410

Supplementary Table S7 | Functional classification of *M. brandtii* genes.

		Number	Percent (%)
Total		22,256	100
Annotated	SWISS-PROT	20,725	93.13
	KEGG	9,158	41.15
	InterPro	19,665	88.36
	GO	15,883	71.37
Unannotated		1,512	6.79

Supplementary Table S8 | Orthologous relationship between *M. brandtii* and mammals.

Type	Ortholog number
M.brandtii: E.caballus	14,140
M.brandtii:C.familiaris	13,997
M.brandtii: B.taurus	14,345
M.brandtii: H.sapiens	14,212
M.brandtii: M.musculus	14,095
M.brandtii: R.norvegicus	13,420
1:1 ortholog	10,374

Supplementary Table S9 | *Myotis brandtii* gained genes compared to human.

GeneID	Gene Symbol	Gene Name
Myotis_brandtii_GOT2_10010193	*GOT2*	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
Myotis_brandtii_GOT2_10010194	*GOT2*	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)
Myotis_brandtii_Armc5_10025112	*ARMC5*	armadillo repeat containing 5
Myotis_brandtii_ATP8B3_10024502	*ATP8B3*	probable phospholipid-transporting ATPase IK-like
Myotis_brandtii_ATP8B3_10024504	*ATP8B3*	probable phospholipid-transporting ATPase IK-like
Myotis_brandtii_ATP8B3_10024505	*ATP8B3*	probable phospholipid-transporting ATPase IK-like
Myotis_brandtii_BAG6_10034716	*BAG6*	BCL2-associated athanogene 6
Myotis_brandtii_Atg14_10033117	*ATG14*	autophagy related 14
Myotis_brandtii_Atg14_10033118	*ATG14*	autophagy related 14
Myotis_brandtii_CASP8AP2_10032158	*CASP8AP2*	caspase 8 associated protein 2
Myotis_brandtii_CADM1_10028129	*CADM1*	cell adhesion molecule 1
Myotis_brandtii_CDAN1_10035881	*CDAN1*	congenital dyserythropoietic anemia, type I
Myotis_brandtii_Cdan1_10035882	*CDAN1*	congenital dyserythropoietic anemia, type I (human)
Myotis_brandtii_Chchd5_10032912	*CHCHD5*	coiled-coil-helix-coiled-coil-helix domain containing 5
Myotis_brandtii_CHIA_10030030	*CHIA*	chitinase, acidic
Myotis_brandtii_CKAP2_10031257	*CKAP2*	cytoskeleton associated protein 2
Myotis_brandtii_NA_10028549	*NA*	NA
Myotis_brandtii_CLDND1_10030172	*CLDND1*	claudin domain containing 1
Myotis_brandtii_C14orf43_10010219	*C14ORF43*	chromosome 14 open reading frame 43
Myotis_brandtii_Nudt21_10029674	*NUDT21*	nudix (nucleoside diphosphate linked moiety X)-type motif 21
Myotis_brandtii_SLC7A3_10035698	*SLC7A3*	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3
Myotis_brandtii_Cul1_10034861	*CUL1*	cullin 1
Myotis_brandtii_Cwc25_10032337	*CWC25*	CWC25 spliceosome-associated protein homolog (S. cerevisiae)
Myotis_brandtii_CWC25_10032338	*CWC25*	similar to putative spliceosomal complex component
Myotis_brandtii_CYCS_10032985	*CYCS*	cytochrome c, somatic
Myotis_brandtii_DAZAP2_10018817	*DAZAP2*	DAZ associated protein 2
Myotis_brandtii_DDX5_10025247	*DDX5*	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
Myotis_brandtii_DHX8_10026498	*DHX8*	DEAH (Asp-Glu-Ala-His) box polypeptide 8
Myotis_brandtii_DHX8_10026499	*DHX8*	DEAH (Asp-Glu-Ala-His) box polypeptide 8
Myotis_brandtii_NA_10009523	*NA*	NA
Myotis_brandtii_ECAT1_10027163	*ECAT1*	ES cell associated transcript 1
Myotis_brandtii_ECE2_10035161	*ECE2*	family M13 non-peptidase homologue (M13 family)
Myotis_brandtii_eef1a_10009184	*EEF1A*	elongation factor 1-alpha
Myotis_brandtii_EEF1A1_10009185	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1
Myotis_brandtii_Eef1a1_10020724	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1
Myotis_brandtii_EEF1A1_10020725	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1
Myotis_brandtii_EEF1A1_10029169	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1
Myotis_brandtii_ETF1_10017428	*ETF1*	eukaryotic translation termination factor 1
Myotis_brandtii_FAM48A_10010229	*FAM48A*	family with sequence similarity 48, member A
Myotis_brandtii_FBXO31_10011222	*FBXO31*	F-box protein 31
Myotis_brandtii_FBXO31_10023009	*FBXO31*	F-box protein 31
Myotis_brandtii_FBXO31_10024405	*FBXO31*	F-box protein 31
Myotis_brandtii_FBXO31_10029175	*FBXO31*	F-box protein 31
Myotis_brandtii_FBXO31_10034120	*FBXO31*	F-box protein 31
Myotis_brandtii_FOP_10032407	*FOP*	NA
Myotis_brandtii_PTGFRN_10014937	*PTGFRN*	prostaglandin F2 receptor negative regulator-like

Myotis_brandtii_NA_10008666	*NA*	NA
Myotis_brandtii_FTH1_10032500	*FTH1*	potential iron transporter similar to S. cerevisiae FTH1 (YBR207W)
Myotis_brandtii_FTH1_10034417	*FTH1*	potential iron transporter similar to S. cerevisiae FTH1 (YBR207W)
Myotis_brandtii_FTL_10019016	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10023160	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10027963	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10032313	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10033417	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10033419	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10035612	*FTL*	ferritin, light polypeptide
Myotis_brandtii_FTL_10035765	*FTL*	ferritin, light polypeptide
Myotis_brandtii_Ftl1_10018904	*FTL1*	ferritin light chain 1
Myotis_brandtii_FUNDC2_10012321	*FUNDC2*	FUN14 domain containing 2
Myotis_brandtii_GAPDH_10022408	*GAPDH*	hypothetical protein
Myotis_brandtii_GAPDH_10022409	*GAPDH*	hypothetical protein
Myotis_brandtii_GAPDH_10022547	*GAPDH*	hypothetical protein
Myotis_brandtii_GAPDH_10023219	*GAPDH*	hypothetical protein
Myotis_brandtii_GALNT11_10009460	*GALNT11*	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)
Myotis_brandtii_Galnt11_10009461	*GALNT11*	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11 (GalNAc-T11)
Myotis_brandtii_GLTP_10026468	*GLTP*	glycolipid transfer protein
Myotis_brandtii_GLTP_10026469	*GLTP*	glycolipid transfer protein
Myotis_brandtii_GLTP_10028392	*GLTP*	glycolipid transfer protein
Myotis_brandtii_GOLIM4_10035910	*GOLIM4*	golgi integral membrane protein 4
Myotis_brandtii_Gpbp1_10012594	*GPBP1*	GC-rich promoter binding protein 1
Myotis_brandtii_Hmgb2_10018437	*HMGB2*	high mobility group box 2
Myotis_brandtii_Hmgb2_10029741	*HMGB2*	high mobility group box 2
Myotis_brandtii_HMGN1_10032322	*HMGN1*	uncharacterized LOC709908
Myotis_brandtii_HMGN4_10023816	*HMGN4*	non-histone chromosomal protein HMG-17-like
Myotis_brandtii_HNRNPH2_10028411	*HNRNPH2*	heterogeneous nuclear ribonucleoprotein H2-like
Myotis_brandtii_Ift122_10023372	*IFT122*	intraflagellar transport 122 homolog (Chlamydomonas)
Myotis_brandtii_IGF2BP2_10033173	*IGF2BP2*	insulin-like growth factor 2 mRNA binding protein 2
Myotis_brandtii_EIF2S3_10023015	*EIF2S3*	eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa
Myotis_brandtii_Eif4a3_10031650	*EIF4A3*	eukaryotic translation initiation factor 4A, isoform 3
Myotis_brandtii_EIF5_10035634	*EIF5*	eukaryotic translation initiation factor 5
Myotis_brandtii_IFT20_10028766	*IFT20*	intraflagellar transport protein 20
Myotis_brandtii_ILF2_10018503	*ILF2*	interleukin enhancer binding factor 2, 45kDa
Myotis_brandtii_ILF2_10019357	*ILF2*	interleukin enhancer binding factor 2, 45kDa
Myotis_brandtii_ILF2_10026032	*ILF2*	interleukin enhancer binding factor 2, 45kDa
Myotis_brandtii_KPNA2_10012583	*KPNA2*	NA
Myotis_brandtii_Kpna2_10018010	*KPNA2*	karyopherin (importin) alpha 2
Myotis_brandtii_Ino80c_10034875	*INO80C*	INO80 complex subunit C
Myotis_brandtii_Ing1_10032312	*ING1*	inhibitor of growth family, member 1
Myotis_brandtii_IPO4_10023804	*IPO4*	importin-4-like
Myotis_brandtii_KIAA0174_10022418	*KIAA0174*	KIAA0174
Myotis_brandtii_Kif15_10035048	*KIF15*	kinesin family member 15
Myotis_brandtii_LSM14A_10008869	*LSM14A*	LSM14A, SCD6 homolog A (S. cerevisiae)
Myotis_brandtii_LSM7_10028402	*LSM7*	small nuclear riboprotein, U6 sm-like protein
Myotis_brandtii_MARK3_10026910	*MARK3*	MAP/microtubule affinity-regulating kinase 3-like
Myotis_brandtii_MAST2_10035163	*MAST2*	microtubule associated serine/threonine kinase 2

Myotis_brandtii_ORF61_10015816	*ORF61*	uracil-DNA glycosylase
Myotis_brandtii_MAD2L1_10018893	*MAD2L1*	MAD2 mitotic arrest deficient-like 1 (yeast)
Myotis_brandtii_Med22_10028621	*MED22*	mediator complex subunit 22
Myotis_brandtii_Chchd4_10019427	*CHCHD4*	coiled-coil-helix-coiled-coil-helix domain containing 4
Myotis_brandtii_MDK_10033218	*MDK*	midkine (neurite growth-promoting factor 2)
Myotis_brandtii_NA_10020689	*NA*	NA
Myotis_brandtii_MMGT1_10033995	*MMGT1*	membrane magnesium transporter 1
Myotis_brandtii_Morf4l1_10018151	*MORF4L1*	mortality factor 4 like 1
Myotis_brandtii_Morf4l1_10018152	*MORF4L1*	mortality factor 4 like 1
Myotis_brandtii_CDC25A_10034525	*CDC25A*	cell division cycle 25 homolog A (S. pombe)
Myotis_brandtii_NA_10002569	*NA*	NA
Myotis_brandtii_NA_10002743	*NA*	NA
Myotis_brandtii_NA_10002888	*NA*	NA
Myotis_brandtii_NA_10004088	*NA*	NA
Myotis_brandtii_NA_10004402	*NA*	NA
Myotis_brandtii_NA_10004641	*NA*	NA
Myotis_brandtii_NA_10004981	*NA*	NA
Myotis_brandtii_NA_10005083	*NA*	NA
Myotis_brandtii_NA_10005157	*NA*	NA
Myotis_brandtii_NA_10005875	*NA*	NA
Myotis_brandtii_NA_10006509	*NA*	NA
Myotis_brandtii_NA_10006856	*NA*	NA
Myotis_brandtii_NA_10006871	*NA*	NA
Myotis_brandtii_NA_10007122	*NA*	NA
Myotis_brandtii_NA_10007161	*NA*	NA
Myotis_brandtii_NA_10007227	*NA*	NA
Myotis_brandtii_NA_10007256	*NA*	NA
Myotis_brandtii_NA_10007652	*NA*	NA
Myotis_brandtii_NA_10008258	*NA*	NA
Myotis_brandtii_NA_10009448	*NA*	NA
Myotis_brandtii_NA_10009703	*NA*	NA
Myotis_brandtii_NA_10010161	*NA*	NA
Myotis_brandtii_NA_10010310	*NA*	NA
Myotis_brandtii_NA_10010447	*NA*	NA
Myotis_brandtii_NA_10011326	*NA*	NA
Myotis_brandtii_NA_10011610	*NA*	NA
Myotis_brandtii_NA_10011705	*NA*	NA
Myotis_brandtii_NA_10011708	*NA*	NA
Myotis_brandtii_NA_10012326	*NA*	NA
Myotis_brandtii_NA_10012473	*NA*	NA
Myotis_brandtii_NA_10012953	*NA*	NA
Myotis_brandtii_NA_10013740	*NA*	NA
Myotis_brandtii_NA_10013744	*NA*	NA
Myotis_brandtii_NA_10013840	*NA*	NA
Myotis_brandtii_NA_10013967	*NA*	NA
Myotis_brandtii_NA_10014175	*NA*	NA
Myotis_brandtii_NA_10014178	*NA*	NA
Myotis_brandtii_NA_10014372	*NA*	NA
Myotis_brandtii_NA_10014558	*NA*	NA
Myotis_brandtii_NA_10014607	*NA*	NA
Myotis_brandtii_NA_10015233	*NA*	NA

Myotis_brandtii_NA_10015949	*NA*	NA
Myotis_brandtii_NA_10016154	*NA*	NA
Myotis_brandtii_NA_10016161	*NA*	NA
Myotis_brandtii_NA_10016917	*NA*	NA
Myotis_brandtii_NA_10017018	*NA*	NA
Myotis_brandtii_NA_10017047	*NA*	NA
Myotis_brandtii_NA_10017128	*NA*	NA
Myotis_brandtii_NA_10017586	*NA*	NA
Myotis_brandtii_NA_10017665	*NA*	NA
Myotis_brandtii_NA_10017748	*NA*	NA
Myotis_brandtii_NA_10017811	*NA*	NA
Myotis_brandtii_NA_10017938	*NA*	NA
Myotis_brandtii_NA_10017970	*NA*	NA
Myotis_brandtii_NA_10018308	*NA*	NA
Myotis_brandtii_NA_10018566	*NA*	NA
Myotis_brandtii_NA_10019025	*NA*	NA
Myotis_brandtii_NA_10019152	*NA*	NA
Myotis_brandtii_NA_10019168	*NA*	NA
Myotis_brandtii_NA_10019212	*NA*	NA
Myotis_brandtii_NA_10020562	*NA*	NA
Myotis_brandtii_NA_10020821	*NA*	NA
Myotis_brandtii_NA_10020847	*NA*	NA
Myotis_brandtii_NA_10021344	*NA*	NA
Myotis_brandtii_NA_10021367	*NA*	NA
Myotis_brandtii_NA_10021372	*NA*	NA
Myotis_brandtii_NA_10021549	*NA*	NA
Myotis_brandtii_NA_10021553	*NA*	NA
Myotis_brandtii_NA_10021775	*NA*	NA
Myotis_brandtii_NA_10021854	*NA*	NA
Myotis_brandtii_NA_10021970	*NA*	NA
Myotis_brandtii_NA_10022983	*NA*	NA
Myotis_brandtii_NA_10023189	*NA*	NA
Myotis_brandtii_NA_10023338	*NA*	NA
Myotis_brandtii_NA_10023440	*NA*	NA
Myotis_brandtii_NA_10023468	*NA*	NA
Myotis_brandtii_NA_10023644	*NA*	NA
Myotis_brandtii_NA_10023905	*NA*	NA
Myotis_brandtii_NA_10023910	*NA*	NA
Myotis_brandtii_NA_10024174	*NA*	NA
Myotis_brandtii_NA_10024175	*NA*	NA
Myotis_brandtii_NA_10024182	*NA*	NA
Myotis_brandtii_NA_10024576	*NA*	NA
Myotis_brandtii_NA_10024999	*NA*	NA
Myotis_brandtii_NA_10026089	*NA*	NA
Myotis_brandtii_NA_10026266	*NA*	NA
Myotis_brandtii_NA_10026296	*NA*	NA
Myotis_brandtii_NA_10027166	*NA*	NA
Myotis_brandtii_NA_10027436	*NA*	NA
Myotis_brandtii_NA_10027464	*NA*	NA
Myotis_brandtii_NA_10027566	*NA*	NA
Myotis_brandtii_NA_10027642	*NA*	NA

Myotis_brandtii_NA_10028011	*NA*	NA
Myotis_brandtii_NA_10028399	*NA*	NA
Myotis_brandtii_NA_10028515	*NA*	NA
Myotis_brandtii_NA_10028663	*NA*	NA
Myotis_brandtii_NA_10029400	*NA*	NA
Myotis_brandtii_NA_10029618	*NA*	NA
Myotis_brandtii_NA_10029906	*NA*	NA
Myotis_brandtii_NA_10029949	*NA*	NA
Myotis_brandtii_NA_10029996	*NA*	NA
Myotis_brandtii_NA_10030171	*NA*	NA
Myotis_brandtii_NA_10030289	*NA*	NA
Myotis_brandtii_NA_10030295	*NA*	NA
Myotis_brandtii_NA_10030357	*NA*	NA
Myotis_brandtii_NA_10030731	*NA*	NA
Myotis_brandtii_NA_10030833	*NA*	NA
Myotis_brandtii_NA_10030848	*NA*	NA
Myotis_brandtii_NA_10030940	*NA*	NA
Myotis_brandtii_NA_10031024	*NA*	NA
Myotis_brandtii_NA_10031025	*NA*	NA
Myotis_brandtii_NA_10031156	*NA*	NA
Myotis_brandtii_NA_10031616	*NA*	NA
Myotis_brandtii_NA_10031853	*NA*	NA
Myotis_brandtii_NA_10031924	*NA*	NA
Myotis_brandtii_NA_10032193	*NA*	NA
Myotis_brandtii_NA_10032522	*NA*	NA
Myotis_brandtii_NA_10032738	*NA*	NA
Myotis_brandtii_NA_10033363	*NA*	NA
Myotis_brandtii_NA_10033981	*NA*	NA
Myotis_brandtii_NA_10034004	*NA*	NA
Myotis_brandtii_NA_10034437	*NA*	NA
Myotis_brandtii_NA_10034440	*NA*	NA
Myotis_brandtii_NA_10035064	*NA*	NA
Myotis_brandtii_NA_10035067	*NA*	NA
Myotis_brandtii_NA_10035085	*NA*	NA
Myotis_brandtii_NA_10035319	*NA*	NA
Myotis_brandtii_NA_10035336	*NA*	NA
Myotis_brandtii_NA_10035360	*NA*	NA
Myotis_brandtii_NA_10035375	*NA*	NA
Myotis_brandtii_NA_10035382	*NA*	NA
Myotis_brandtii_NA_10035470	*NA*	NA
Myotis_brandtii_NA_10035638	*NA*	NA
Myotis_brandtii_NA_10035697	*NA*	NA
Myotis_brandtii_NA_10035791	*NA*	NA
Myotis_brandtii_NA_10035973	*NA*	NA
Myotis_brandtii_NGRN_10022900	*NGRN*	neugrin, neurite outgrowth associated
Myotis_brandtii_Nhp2l1_10020694	*NHP2L1*	NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)
Myotis_brandtii_Nkap_10018116	*NKAP*	NFKB activating protein
Myotis_brandtii_Nono_10029908	*NONO*	non-POU-domain-containing, octamer binding protein
Myotis_brandtii_Nop10_10027797	*NOP10*	NOP10 ribonucleoprotein homolog (yeast)
Myotis_brandtii_NOP14_10018438	*NOP14*	similar to S. cerevisiae NOP14 (YDL148C) U3 snoRNP/small subunit rRNA processome component

Myotis_brandtii_NSF_10022249	*NSF*	vesicle-fusing ATPase
Myotis_brandtii_NA_10006876	*NA*	NA
Myotis_brandtii_PIK3R2_10024461	*PIK3R2*	phosphoinositide-3-kinase, regulatory subunit 2 (beta)
Myotis_brandtii_PABPC4_10023698	*PABPC4*	poly(A) binding protein, cytoplasmic 4 (inducible form)
Myotis_brandtii_PABPC4_10023699	*PABPC4*	poly(A) binding protein, cytoplasmic 4 (inducible form)
Myotis_brandtii_PCID2_10033784	*PCID2*	PCI domain-containing protein 2
Myotis_brandtii_Pcid2_10033785	*PCID2*	PCI domain containing 2
Myotis_brandtii_Pdcl3_10028014	*PDCL3*	phosducin-like 3
Myotis_brandtii_PEX5_10026123	*PEX5*	similar to S. cerevisiae PEX5 (YDR244W) receptor for peroxisomal targeting sequence 1
Myotis_brandtii_PPP3CC_10028538	*PPP3CC*	protein phosphatase 3, catalytic subunit, gamma isozyme
Myotis_brandtii_PRDX3_10023933	*PRDX3*	peroxiredoxin 3
Myotis_brandtii_PRDX5_10026214	*PRDX5*	peroxiredoxin 5
Myotis_brandtii_PTMA_10019120	*PTMA*	prothymosin, alpha (gene sequence 28)
Myotis_brandtii_ATIC_10023696	*ATIC*	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase
Myotis_brandtii_Ranbp1_10017823	*RANBP1*	RAN binding protein 1
Myotis_brandtii_Rpl12_10003701	*RPL12*	ribosomal protein L12
Myotis_brandtii_RPL12_10026992	*RPL12*	similar to S. cerevisiae RPL12B (YDR418W) and RPL12A (YEL054C) genes for cytosolic large subunit ribosomal protein L12
Myotis_brandtii_RPL13_10028769	*RPL13*	Ribosomal protein L13, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL17_10016236	*RPL17*	Ribosomal protein L17, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL18_10015319	*RPL18*	similar to C terminus of S. cerevisiae RPL18A (YOL120C) and RPL18B (YOL120C); intron and 5' exon upstream
Myotis_brandtii_Rpl19_10021565	*RPL19*	ribosomal protein L19
Myotis_brandtii_Rpl19_10021566	*RPL19*	ribosomal protein L19
Myotis_brandtii_RPL21_10025405	*RPL21*	similar to C terminus of S. cerevisiae RPL21A (YBR191W) and RPL21B (YPL079W) genes for cytosolic ribosomal protein large subunit; short 5' exon upstream of intron
Myotis_brandtii_Rpl26_10028664	*RPL26*	ribosomal protein L26
Myotis_brandtii_Rpl27_10034731	*RPL27*	ribosomal protein L27
Myotis_brandtii_Rpl27a_10020321	*RPL27A*	ribosomal protein L27A
Myotis_brandtii_RPL29_10008202	*RPL29*	Rpl29p
Myotis_brandtii_RPL29_10012324	*RPL29*	Rpl29p
Myotis_brandtii_Rpl31_10034841	*RPL31*	ribosomal protein L31
Myotis_brandtii_Rpl32_10005908	*RPL32*	ribosomal protein L32
Myotis_brandtii_Rpl32_10015803	*RPL32*	ribosomal protein L32
Myotis_brandtii_Rpl32_10032721	*RPL32*	ribosomal protein L32
Myotis_brandtii_Rpl32_10034974	*RPL32*	ribosomal protein L32
Myotis_brandtii_Rpl36_10005219	*RPL36*	Rpl36 60S ribosomal protein L36, probable (IC)
Myotis_brandtii_RPL36_10011147	*RPL36*	ribosomal protein L36, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_Rpl36a_10029682	*RPL36A*	ribosomal protein L36A
Myotis_brandtii_RPL37A_10007828	*RPL37A*	Rpl37ap
Myotis_brandtii_RPL4_10018988	*RPL4*	ribosomal protein L4, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL4_10018989	*RPL4*	ribosomal protein L4, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL4_10018990	*RPL4*	ribosomal protein L4, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL4_10035501	*RPL4*	ribosomal protein L4, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPL6_10033107	*RPL6*	Ribosomal protein L6, component of cytosolic 80S ribosome and 60S large subunit

Myotis_brandtii_RPL7_10026407	*RPL7*	Ribosomal protein L7, component of cytosolic 80S ribosome and 60S large subunit
Myotis_brandtii_RPLP1_10035388	*RPLP1*	ribosomal protein, large, P1
Myotis_brandtii_RPLP1_10035762	*RPLP1*	ribosomal protein, large, P1
Myotis_brandtii_RNF181_10031381	*RNF181*	ring finger protein 181
Myotis_brandtii_RLIM_10016583	*RLIM*	ring finger protein, LIM domain interacting
Myotis_brandtii_HNRNPA1_10025610	*HNRNPA1*	heterogeneous nuclear ribonucleoprotein A1
Myotis_brandtii_POLR2C_10024582	*POLR2C*	polymerase (RNA) II (DNA directed) polypeptide C, 33kDa
Myotis_brandtii_Rps11_10031571	*RPS11*	Rps11
Myotis_brandtii_Rps15_10033481	*RPS15*	ribosomal protein S15
Myotis_brandtii_Rps16_10019181	*RPS16*	ribosomal protein S16
Myotis_brandtii_RPS2_10017782	*RPS2*	Rps2p
Myotis_brandtii_rps2_10034595	*RPS2*	RPS2
Myotis_brandtii_RPS2_10035519	*RPS2*	Rps2p
Myotis_brandtii_rps24_10029631	*RPS24*	ribosomal protein S24
Myotis_brandtii_Rps26_10022142	*RPS26*	ribosomal protein S26
Myotis_brandtii_Rps26_10031098	*RPS26*	ribosomal protein S26
Myotis_brandtii_Rps6_10012767	*RPS6*	ribosomal protein S6
Myotis_brandtii_Rps6_10021894	*RPS6*	ribosomal protein S6
Myotis_brandtii_RPS6_10022417	*RPS6*	TIR-NBS-LRR class disease resistance protein
Myotis_brandtii_Rps8_10015562	*RPS8*	Rps8
Myotis_brandtii_RPSA_10028767	*RPSA*	ribosomal protein SA
Myotis_brandtii_MRPS33_10034391	*MRPS33*	mitochondrial ribosomal protein S33
Myotis_brandtii_SLC25A39_10027264	*SLC25A39*	solute carrier family 25, member 39
Myotis_brandtii_Slc25a39_10027266	*SLC25A39*	solute carrier family 25, member 39
Myotis_brandtii_SALL4_10017833	*SALL4*	sal-like 4 (Drosophila)
Myotis_brandtii_SALL4_10017834	*SALL4*	sal-like 4 (Drosophila)
Myotis_brandtii_NA_10024307	*NA*	NA
Myotis_brandtii_SF3B5_10011224	*SF3B5*	splicing factor 3b, subunit 5, 10kDa
Myotis_brandtii_SIAH1_10026007	*SIAH1*	siah E3 ubiquitin protein ligase 1
Myotis_brandtii_skp1_10027959	*SKP1*	S-phase kinase-associated protein 1
Myotis_brandtii_SLIT1_10033398	*SLIT1*	slit homolog 1 (Drosophila)
Myotis_brandtii_Son_10029366	*SON*	Son DNA binding protein
Myotis_brandtii_Spata5_10005562	*SPATA5*	spermatogenesis associated 5
Myotis_brandtii_SRSF1_10020959	*SRSF1*	serine/arginine-rich splicing factor 1
Myotis_brandtii_Srsf3_10012579	*SRSF3*	serine/arginine-rich splicing factor 3
Myotis_brandtii_Ssr3_10029717	*SSR3*	signal sequence receptor, gamma
Myotis_brandtii_Suclg1_10027229	*SUCLG1*	succinate-CoA ligase, GDP-forming, alpha subunit
Myotis_brandtii_GTF2F1_10033850	*GTF2F1*	general transcription factor IIF, polypeptide 1, 74kDa
Myotis_brandtii_TAF12_10028254	*TAF12*	transcription initiation factor TFIID subunit D10
Myotis_brandtii_TBC1D25_10017075	*TBC1D25*	TBC1 domain family, member 25
Myotis_brandtii_TCNA_10031674	*TCNA*	NA
Myotis_brandtii_Tead3_10027735	*TEAD3*	TEA domain family member 3
Myotis_brandtii_GTF2H1_10026470	*GTF2H1*	general transcription factor IIH, polypeptide 1, 62kDa
Myotis_brandtii_GTF2H1_10026471	*GTF2H1*	general transcription factor IIH, polypeptide 1, 62kDa
Myotis_brandtii_Tfpt_10021192	*TFPT*	TCF3 (E2A) fusion partner
Myotis_brandtii_TFPT_10021193	*TFPT*	TCF3 (E2A) fusion partner (in childhood Leukemia)
Myotis_brandtii_Timm22_10020667	*TIMM22*	translocase of inner mitochondrial membrane 22 homolog (yeast)
Myotis_brandtii_Tmem140_10034109	*TMEM140*	transmembrane protein 140
Myotis_brandtii_Trappc2l_10022522	*TRAPPC2L*	trafficking protein particle complex 2-like
Myotis_brandtii_UBE2R2_10009073	*UBE2R2*	ubiquitin-conjugating enzyme E2R 2

Myotis_brandtii_UBE2V1_10033069	*UBE2V1*	ubiquitin-conjugating enzyme E2 variant 1
Myotis_brandtii_FAU_10019804	*FAU*	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed
Myotis_brandtii_UBP1_10030291	*UBP1*	upstream binding protein 1 (LBP-1a)
Myotis_brandtii_USF2_10034396	*USF2*	upstream transcription factor 2, c-fos interacting
Myotis_brandtii_UTP18_10017844	*UTP18*	UTP18, small subunit (SSU) processome component, homolog (yeast)
Myotis_brandtii_UTP18_10017845	*UTP18*	UTP18, small subunit (SSU) processome component, homolog (yeast)
Myotis_brandtii_ATP6V1E1_10020062	*ATP6V1E1*	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1
Myotis_brandtii_YBX1_10032836	*YBX1*	Y box binding protein 1
Myotis_brandtii_Zmat2_10017843	*ZMAT2*	zinc finger, matrin type 2
Myotis_brandtii_ZWINT_10018315	*ZWINT*	ZW10 interactor

Supplementary Table S10 | *Myotis brandtii* lost genes compared to human.

GeneID	Gene Symbol	Gene Name
ENST00000416408	*AC004258.1*	NA
ENST00000436967	*AC008271.1*	NA
ENST00000535853	*AC016586.1*	NA
ENST00000409070	*AC016757.3*	Uncharacterized protein
ENST00000452718	*AC083862.1*	Uncharacterized protein
ENST00000356686	*AC092031.1*	NA
ENST00000418835	*AC092687.4*	Uncharacterized protein
ENST00000416839	*AC096644.1*	Uncharacterized protein
ENST00000409162	*AC112721.1*	Uncharacterized protein
ENST00000544589	*AC132186.1*	CDNA FLJ27243 fis, clone SYN08134
ENST00000535173	*AC135178.1*	HCG1985372; Uncharacterized protein; cDNA FLJ37541 fis, clone BRCAN2026340
ENST00000440990	*AL033532.1*	Uncharacterized protein
ENST00000370795	*AL078633.1*	NA
ENST00000398726	*AL162389.1*	Uncharacterized protein
ENST00000373137	*AL591845.1*	NA
ENST00000544076	*AP000974.1*	CDNA FLJ26432 fis, clone KDN01418; Uncharacterized protein
ENST00000398067	*AP001024.1*	Uncharacterized protein
ENST00000447610	*AP001024.2*	Uncharacterized protein
ENST00000369492	*ATP1A1OS*	ATP1A1 opposite strand
ENST00000239125	*C10ORF95*	chromosome 10 open reading frame 95
ENST00000378416	*C16ORF85*	chromosome 16 open reading frame 85
ENST00000382638	*C18ORF62*	chromosome 18 open reading frame 62
ENST00000301246	*C19ORF33*	chromosome 19 open reading frame 33
ENST00000420022	*C19ORF76*	chromosome 19 open reading frame 76
ENST00000379205	*C4ORF46*	chromosome 4 open reading frame 46
ENST00000503143	*C5ORF20*	chromosome 5 open reading frame 20
ENST00000436592	*C5ORF27*	chromosome 5 open reading frame 27
ENST00000509627	*C5ORF49*	chromosome 5 open reading frame 49
ENST00000519561	*C8ORF44*	chromosome 8 open reading frame 44
ENST00000457681	*C9ORF147*	chromosome 9 open reading frame 147
ENST00000374213	*CD52*	CD52 molecule
ENST00000378779	*CTD-2210P24.4*	Uncharacterized protein
ENST00000510604	*CTD-2228K2.5*	Uncharacterized protein
ENST00000357412	*CXORF24*	chromosome X open reading frame 24
ENST00000322246	*DEFB112*	defensin, beta 112
ENST00000398948	*DSCR4*	Down syndrome critical region gene 4
ENST00000357704	*DSCR8*	Down syndrome critical region gene 8
ENST00000327532	*DZIP1L*	DAZ interacting protein 1-like
ENST00000360827	*FAM177B*	family with sequence similarity 177, member B
ENST00000407599	*GAGE10*	G antigen 10
ENST00000361446	*GAGE12B*	G antigen 12B
ENST00000420398	*GAGE12C*	G antigen 12C
ENST00000405679	*GAGE12D*	G antigen 12D
ENST00000381698	*GAGE12E*	G antigen 12E
ENST00000440137	*GAGE12F*	G antigen 12F
ENST00000445148	*GAGE12G*	G antigen 12G
ENST00000381722	*GAGE12H*	G antigen 12H
ENST00000415752	*GAGE12I*	G antigen 12I

ENST00000442437	*GAGE12J*	G antigen 12J
ENST00000381751	*GAGE13*	G antigen 13
ENST00000381709	*GAGE1*	G antigen 1
ENST00000362097	*GAGE2A*	G antigen 2A
ENST00000381725	*GAGE2B*	G antigen 2B
ENST00000381708	*GAGE2C*	G antigen 2C
ENST00000404720	*GAGE2D*	G antigen 2D
ENST00000402590	*GAGE2E*	G antigen 2E
ENST00000381700	*GAGE4*	G antigen 4
ENST00000373005	*GTSF1L*	gametocyte specific factor 1-like
ENST00000293330	*HCRT*	hypocretin (orexin) neuropeptide precursor
ENST00000257572	*HRK*	harakiri, BCL2 interacting protein (contains only BH3 domain)
ENST00000377035	*IL31*	interleukin 31
ENST00000409539	*INMT*	indolethylamine N-methyltransferase
ENST00000535211	*KTN1-AS1*	KTN1 antisense RNA 1 (non-protein coding)
ENST00000460508	*LBX2*	ladybird homeobox 2
ENST00000538077	*LINC00346*	long intergenic non-protein coding RNA 346
ENST00000374555	*MDS2*	myelodysplastic syndrome 2 translocation associated
ENST00000261507	*MSMO1*	methylsterol monooxygenase 1
ENST00000536684	*MTRNR2L8*	MT-RNR2-like 8
ENST00000308946	*MYEOV*	myeloma overexpressed (in a subset of t(11;14) positive multiple myelomas)
ENST00000425699	*NANOS1*	nanos homolog 1 (Drosophila)
ENST00000356906	*NBR2*	neighbor of BRCA1 gene 2 (non-protein coding)
ENST00000552815	*ORPHAN*	Uncharacterized protein
ENST00000376150	*PAGE1*	P antigen family, member 1 (prostate associated)
ENST00000218068	*PAGE4*	P antigen family, member 4 (prostate associated)
ENST00000418336	*PKNOX1*	PBX/knotted 1 homeobox 1
ENST00000301698	*PRR25*	proline rich 25
ENST00000342055	*RHD*	Rh blood group, D antigen
ENST00000447795	*RP11-126K1.2*	Novel protein; Uncharacterized protein
ENST00000532511	*RP11-17M16.1*	Uncharacterized protein
ENST00000529278	*RP11-315O6.2*	HCG1815860; Uncharacterized protein
ENST00000523692	*RP11-325I22.2*	NA
ENST00000439236	*RP11-366L20.2*	Uncharacterized protein
ENST00000375670	*RP11-410N8.4*	NA
ENST00000512516	*RP11-487E13.1*	Uncharacterized protein
ENST00000551650	*RP11-650K20.3*	Uncharacterized protein
ENST00000508020	*RP11-701P16.2*	Uncharacterized protein
ENST00000354756	*RP11-89K21.1*	NA
ENST00000311067	*RP11-89N17.1*	HCG1643653; Uncharacterized protein
ENST00000415412	*RP5-837J1.2*	NA
ENST00000343542	*RXFP1*	relaxin/insulin-like family peptide receptor 1
ENST00000269205	*SLC25A52*	solute carrier family 25, member 52
ENST00000406890	*STEAP1B*	STEAP family member 1B
ENST00000303130	*TMEM133*	transmembrane protein 133
ENST00000296529	*TMEM144*	transmembrane protein 144
ENST00000434970	*TRDD2*	T cell receptor delta diversity 2
ENST00000250066	*USP6*	ubiquitin specific peptidase 6 (Tre-2 oncogene)
ENST00000506413	*WWC2-AS2*	WWC2 antisense RNA 2 (non-protein coding)
ENST00000255198	*ZBED3*	zinc finger, BED-type containing 3

Supplementary Table S11 | Overview of species examined. C = complete coverage. The UCSC genome name is shown in bold, while the full common name includes the text in brackets.

Common name	Binomial name	UCSC assembly	Coverage
Brandt's bat	*Myotis brandtii*	-	78X
little brown bat	*Myotis lucifugus*	myoLuc2	6.6X
megabat/fruit bat/flying fox	*Pteropus vampyrus*	pteVam1	2.9X
human	*Homo sapiens*	hg19	C
(common) **chimpanzee**	*Pan troglodytes*	panTro2	6.0X
(Western lowland) **gorilla**	*Gorilla gorilla gorilla*	gorGor1	2.1X
(Sumatran) **orangutan**	*Pongo pygmaeus abelii*	ponAbe2	6X
rhesus (monkey)	*Macaca mulatta*	rheMac2	5.1X
(hamadryas) **baboon**	*Papio hamadryas*	papHam1	5.3X
(common) **marmoset**	*Callithrix jacchus*	calJac1	6X
(Philippine) **tarsier**	*Tarsier syrichta*	tarSyr1	2.1X
(gray) **mouse lemur**	*Microcebus murinus*	micMur1	1.9X
bushbaby/small-eared greater galago	*Otolemur garnettii*	otoGar1	1.9X
(Northern) **tree shrew**	*Tupaia belangeri*	tupBel1	1.9X
(house) **mouse**	*Mus musculus*	mm9	C
(brown) **rat**	*Rattus norvegicus*	rn4	7.3X
(Ord's) **kangaroo rat**	*Dipodomys ordii*	dipOrd1	1.8X
guinea pig	*Cavia porcellus*	cavPor3	6.76X
(thirteen-lined ground) **squirrel**	*Spermophilus tridecemlineatus*	speTri1	1.9X
(European) **rabbit**	*Oryctolagus cuniculus*	oryCun2	7.5X
(American) **pika**	*Ochotona princeps*	ochPri2	1.9X
alpaca	*Vicugna pacos*	vicPac1	2.9X
(Atlantic bottlenose) **dolphin**	*Tursiops truncatus*	turTru1	2.8X
cow	*Bos taurus*	bosTau4	5.4X
horse	*Equus caballus*	equCab2	6.8X
cat	*Felis catus*	felCat3	1.9X
dog	*Canis lupus familiaris*	canFam2	7.6X
(European) **hedgehog**	*Erinaceus europaeus*	eriEur1	1.9X
(common) **shrew**	*Sorex araneus*	sorAra1	1.9X
(African) **elephant**	*Loxodonta africana*	loxAfr3	7X
rock hyrax	*Procavia capensis*	proCap1	2.4X
(lesser hedgehog) **tenrec**	*Echinops telfairi*	echTel1	1.9X
(nine-banded) **armadillo**	*Dasypus novemcinctus*	dasNov2	2.3X
(Hoffmann's two-toed) **sloth**	*Choloepus hoffmanni*	choHof1	2.2X
(tammar) **wallaby**	*(Macropus eugenii*	macEug1	2X
(gray short-tailed) **opossum**	*Monodelphis domestica*	monDom5	6.8X
platypus	*Ornithorhynchus anatinus*	ornAna1	6X
chicken	*Gallus gallus*	galGal3	6.6X
zebra finch	*Taeniopygia guttata*	taeGut1	6X
(green anole) **lizard**	*Anolis carolinensis*	anoCar1	6.8X
(Western clawed) **frog**	*Xenopus tropicalis*	xenTro2	7.65X
tetraodon/green spotted puffer	*Tetraodon nigroviridis*	tetNig2	7.9X
(tora)**fugu**	*Takifugu rubripes*	fr2	8.5X
(three-spined) **stickleback**	*Gasterosteus aculeatus*	gasAcu1	6X
medaka/Japanese rice fish	*Oryzias latipes*	oryLat2	10.6X
zebrafish	*Danio rerio*	danRer6	6.5X
(sea) **lamprey**	*Petromyzon marinus*	petMar1	5.9X

naked mole rat	*Heterocephalus glaber*	hetGla1	90X

Supplementary Table S12 | Positively selected genes in the *M. brandtii* genome. The genes in bold were manually inspected. '% sites under selection' refers to sites in the Gblocks alignment under positive selection (detected by Bayes Empirical Bayes (BEB) with posterior probabilities greater than 95%.

Symbol	Name	Ω	% sites tested under selection	P-value	FDR corrected P-value
AANAT	aralkylamine N-acetyltransferase	999	1.7%	5.70E-05	0.010253919
ACOT11	acyl-CoA thioesterase 11	32.89828	0.5%	0.0004647	0.043873529
ADAM19	ADAM metallopeptidase domain 19	9.73739	0.6%	4.20E-05	0.008650868
ADAM32	ADAM metallopeptidase domain 32	21.20338	1.6%	0.00019226	0.023175913
AFM	Afamin	29.99272	0.5%	7.76E-05	0.012939625
AIFM3	apoptosis-inducing factor, mitochondrion-associated, 3	74.24212	2.5%	7.50E-05	0.012813987
AKR1D1	aldo-keto reductase family 1, member D1 (delta 4-3-ketosteroid-5-beta-reductase)	8.10238	1.0%	0.00054663	0.048874514
ALS2CR11	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 11	8.64673	0.2%	4.84E-08	5.94E-05
ANO9	anoctamin 9	999	0.5%	3.55E-06	0.001776832
ARMC9	armadillo repeat containing 9	998.99916	0.3%	2.08E-05	0.00616651
ATP2A3	ATPase, Ca++ transporting, ubiquitous	118.81451	0.1%	0.00015224	0.020761538
C17orf78	chromosome 17 open reading frame 78	999	1.8%	4.35E-07	0.000345165
C3	complement component 3	4.34235	3.1%	0.00045188	0.043379383
C3orf38	chromosome 3 open reading frame 38	20.1861	1.1%	0.00012583	0.018465553
C4orf29	chromosome 4 open reading frame 29	67.68103	0.6%	4.29E-05	0.008652126
C8A	**complement component 8, alpha polypeptide**	**14.85375**	**0.6%**	**1.33E-05**	**0.005003996**
CD200R1	**CD200 receptor 1**	**6.23285**	**1.3%**	**8.68E-16**	**5.86E-12**
CD300LF	CD300 molecule-like family member f	46.49154	3.6%	9.63E-07	0.000650262
CD46	CD46 molecule, complement regulatory protein	15.78262	0.8%	6.09E-05	0.010818742
CD5	CD5 molecule	100.26543	0.6%	0.00016994	0.022060125
CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	7.53893	3.1%	0.00046374	0.043873529
CDCA5	cell division cycle associated 5	62.78854	0.8%	2.34E-05	0.006330079
CDH19	cadherin 19, type 2	2.72528	2.1%	0.00048135	0.044511687
CEPT1	choline/ethanolamine phosphotransferase 1	882.17093	0.7%	2.76E-05	0.006861945
CGNL1	cingulin-like 1	998.985	1.8%	0.00013527	0.019428514
CIDEC	**cell death-inducing DFFA-like effector c**	**999**	**0.4%**	**0.0001562**	**0.020808747**
CNPY4	canopy 4 homolog (zebrafish)	998.99983	0.9%	2.72E-05	0.006861945
CPN2	carboxypeptidase N, polypeptide 2	29.29815	0.9%	9.58E-06	0.003920772
CUL4B	cullin 4B	29.73298	0.4%	4.23E-05	0.008650868
DDX10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	75.82903	9.3%	0.0003824	0.037411467
DHRS3	dehydrogenase/reductase (SDR family) member 3	999	1.9%	0.00010761	0.016595611
DMBT1	deleted in malignant brain tumors 1	389.32792	1.1%	7.13E-09	1.20E-05
DPP4	dipeptidyl-peptidase 4	43.84608	0.3%	0.0002997	0.030961606
ENG	Endoglin	998.99864	0.2%	0.00010914	0.016595611
ENPP5	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)	998.9989	0.3%	0.00032098	0.032172383
EPX	eosinophil peroxidase	58.04964	0.2%	5.24E-05	0.010104148
EVI2B	ecotropic viral integration site 2B	15.05115	0.2%	7.73E-05	0.012939625

FAM82A2	family with sequence similarity 82, member A2	54.06927	0.3%	0.00047235	0.044286093
FER	fer (fps/fes related) tyrosine kinase	16.77718	0.9%	0.00014239	0.020025077
FGB	**fibrinogen beta chain**	**12.7338**	**0.4%**	**1.60E-05**	**0.005408838**
FOXH1	forkhead box H1	999	1.3%	0.0003217	0.032172383
FREM1	FRAS1 related extracellular matrix 1	207.21303	0.2%	5.37E-05	0.010194532
FRK	fyn-related kinase	6.58452	1.3%	0.00014713	0.020478372
FSTL4	follistatin-like 4	621.50634	9.1%	0.00019092	0.023175913
FTSJ1	FtsJ RNA methyltransferase homolog 1 (E. coli)	998.99974	0.5%	0.0001413	0.020025077
GIMAP7	GTPase, IMAP family member 7	31.11229	0.5%	3.16E-05	0.007198058
GLT25D1	glycosyltransferase 25 domain containing 1	43.1004	0.8%	0.00028236	0.030497139
GPR125	G protein-coupled receptor 125	998.99919	0.0%	3.75E-06	0.001807205
GPR146	G protein-coupled receptor 146	999	0.4%	0.00027345	0.030145647
GTPBP3	GTP binding protein 3 (mitochondrial)	83.18907	0.5%	0.00016234	0.021279149
HEMGN	Hemogen	18.70414	0.3%	0.0002682	0.029925357
HIST1H1T	histone cluster 1, H1t	757.18013	3.1%	9.09E-07	0.000646229
HNRPLL	heterogeneous nuclear ribonucleoprotein L-like	56.32263	0.4%	2.26E-05	0.00630662
HPS6	Hermansky-Pudlak syndrome 6	999	0.2%	0.00017911	0.022390408
HSDL2	hydroxysteroid dehydrogenase like 2	83.61153	0.2%	0.00011526	0.017201164
IFLTD1	intermediate filament tail domain containing 1	19.85907	0.4%	1.46E-05	0.005145266
IQCF5	**IQ motif containing F5**	**10.33215**	**3.9%**	**2.85E-05**	**0.006877795**
IRF9	interferon regulatory factor 9	14.19644	1.3%	0.00012763	0.018528308
IZUMO1	izumo sperm-egg fusion 1	8.39178	0.4%	0.00022051	0.025229708
KALRN	kalirin, RhoGEF kinase	65.76299	0.5%	0.00015202	0.020761538
KCNK4	**potassium channel, subfamily K, member 4**	**998.99924**	**0.8%**	**0.00015721**	**0.020808747**
KIAA1551	KIAA1551	130.67397	0.2%	9.81E-08	0.000110397
KNG1	**kininogen 1**	**7.93538**	**1.9%**	**7.35E-06**	**0.003201827**
LGALS3	lectin, galactoside-binding, soluble, 3	998.93159	0.6%	0.00015436	0.020808747
MAP2K6	**mitogen-activated protein kinase kinase 6**	**999**	**0.9%**	**2.80E-05**	**0.006861945**
MAP3K15	mitogen-activated protein kinase kinase kinase 15	41.92301	18.2%	3.06E-05	0.007198058
MEIS3	Meis homeobox 3	71.83448	0.7%	0.00010752	0.016595611
MYH1	myosin, heavy chain 1, skeletal muscle, adult	3.3455	4.5%	0.00021686	0.025024161
MYH15	myosin, heavy chain 15	51.92737	0.2%	3.28E-05	0.007258669
MYOT	Myotilin	409.11566	0.3%	0.00054497	0.048874514
NBN	Nibrin	52.78951	0.3%	0.0001006	0.016169055
NDUFB1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	138.29874	1.5%	0.00036896	0.036568986
NEBL	**Nebulette**	**173.55538**	**0.2%**	**2.78E-05**	**0.006861945**
NFYA	nuclear transcription factor Y, alpha	123.986	0.6%	4.14E-05	0.008650868
NID2	nidogen 2 (osteonidogen)	3.94666	0.2%	0.00018386	0.022604356
NMBR	neuromedin B receptor	998.99999	0.5%	0.00037108	0.036568986
NOX3	**NADPH oxidase 3**	**998.99859**	**0.2%**	**2.29E-05**	**0.00630662**
NPNT	Nephronectin	229.05817	0.6%	1.70E-06	0.001040682
NTNG2	netrin G2	999	3.4%	7.92E-05	0.013043777
P2RY11	purinergic receptor P2Y, G-protein coupled, 11	67.09373	0.6%	2.73E-05	0.006861945
P2RY12	purinergic receptor P2Y, G-protein coupled, 12	998.9993	0.3%	0.00019962	0.023640962

PCDHGB7	protocadherin gamma subfamily B, 7	112.82458	0.3%	3.18E-06	0.001650861
PDE9A	phosphodiesterase 9A	999	0.2%	0.00053116	0.048187697
PGC	progastricsin (pepsinogen C)	289.02029	1.8%	1.08E-07	0.000111799
PHACTR4	phosphatase and actin regulator 4	177.71686	0.2%	1.81E-05	0.005812502
PITPNB	phosphatidylinositol transfer protein, beta	999	0.8%	3.83E-05	0.008348409
PLD4	phospholipase D family, member 4	54.57308	1.0%	0.00017816	0.022390408
PLSCR1	phospholipid scramblase 1	69.34939	0.5%	0.00020615	0.023993372
POLRMT	polymerase (RNA) mitochondrial (DNA directed)	26.38593	1.6%	5.52E-05	0.010194532
PPP1R12C	protein phosphatase 1, regulatory subunit 12C	24.51994	0.4%	4.46E-05	0.008859237
PRKRIR	protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)	28.225	1.7%	0.00027932	0.030412091
PRRG1	proline rich Gla (G-carboxyglutamic acid) 1	147.53862	1.1%	1.24E-05	0.004767396
PRSS53	protease, serine, 53	236.31104	5.1%	9.76E-22	1.32E-17
PWP1	PWP1 homolog (S. cerevisiae)	678.2464	0.7%	0.00030893	0.031597454
RAD51AP2	RAD51 associated protein 2	14.54511	0.1%	4.24E-10	1.15E-06
RAPGEF3	Rap guanine nucleotide exchange factor (GEF) 3	999	1.6%	3.20E-05	0.007198058
RHNO1	RAD9-HUS1-RAD1 interacting nuclear orphan 1	47.65033	1.3%	1.41E-06	0.000904503
RPL23	ribosomal protein L23	142.1821	2.8%	3.61E-07	0.000304464
RSPO1	**R-spondin 1**	**999**	**1.1%**	**2.05E-05**	**0.00616651**
RTL1	retrotransposon-like 1	2.65774	1.0%	0.00010189	0.016183728
SCEL	Sciellin	834.95257	0.5%	2.59E-07	0.000233252
SCGB1C1	secretoglobin, family 1C, member 1	22.72521	3.8%	0.00029517	0.030961606
SCN5A	sodium channel, voltage-gated, type V, alpha subunit	63.05693	0.4%	0.00029688	0.030961606
SGSM3	small G protein signaling modulator 3	999	0.5%	0.00045304	0.043379383
SIGLEC10	sapiens sialic acid binding Ig-like lectin 10	22.82778	0.2%	0.00019692	0.023527583
SIGLEC5	sialic acid binding Ig-like lectin 5	975.4514	0.3%	0.0000669	0.011731317
SLC13A3	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 3	6.25625	1.8%	0.0000138	0.005036603
SLC16A1	solute carrier family 16, member 1 (monocarboxylic acid transporter 1)	14.75259	0.3%	0.00048978	0.044983128
SLC22A14	solute carrier family 22, member 14	194.45004	0.6%	0.00011594	0.017201164
SLC22A6	solute carrier family 22 (organic anion transporter), member 6	508.68971	3.3%	0.00053181	0.048187697
SLC39A2	solute carrier family 39 (zinc transporter), member 2	7.0811	0.3%	7.44E-05	0.012813987
SLC44A2	solute carrier family 44, member 2	250.76395	0.3%	0.00027464	0.030145647
SLC9C2	solute carrier family 9, member C2 (putative)	31.37037	0.4%	3.16E-05	0.007198058
SLFN11	schlafen family member 11	12.36061	0.2%	0.00032043	0.032172383
SMAD4	**SMAD family member 4**	**16.0078**	**1.4%**	**4.96E-06**	**0.002231445**
SPAM1	**sperm adhesion molecule 1 (PH-20 hyaluronidase, zona pellucida binding)**	**19.19882**	**2.4%**	**5.78E-10**	**1.30E-06**
SPEG	SPEG complex locus	17.14402	0.0%	0.00024291	0.027329399
SPRYD7	SPRY domain containing 7	170.1284	1.6%	4.41E-06	0.00205299
SPTBN4	spectrin, beta, non-erythrocytic 4	999	0.1%	2.14E-05	0.00616651
STRADA	**STE20-related kinase adaptor alpha**	**999**	**0.5%**	**2.42E-06**	**0.001306681**
SUMO1	SMT3 suppressor of mif two 3 homolog 1 (S. cerevisiae)	998.99994	2.6%	1.00E-04	0.016169055
TCEA2	transcription elongation factor A (SII), 2	999	3.8%	2.10E-07	0.000202563
TCOF1	**Treacher Collins-Franceschetti syndrome 1**	**10.74103**	**0.4%**	**5.03E-05**	**0.009834207**
TLR9	toll-like receptor 9	15.49447	0.1%	0.00056325	0.049702211
TM4SF1	transmembrane 4 L six family member 1	66.95365	2.1%	1.71E-05	0.005645394

TMPRSS11F	transmembrane protease, serine 11F	946.23245	1.5%	3.64E-10	1.15E-06
TNXB	tenascin XB	22.93863	1.3%	2.28E-08	3.07E-05
TOX2	TOX high mobility group box family member 2	998.99999	0.6%	5.58E-05	0.010194532
TPM2	tropomyosin 2 (beta)	71.24693	0.7%	0.00029993	0.030961606
TRMT61A	tRNA methyltransferase 61 homolog A (S. cerevisiae)	703.13541	1.3%	0.00017311	0.022060125
TRPV5	**transient receptor potential cation channel, subfamily V, member 5**	**999**	**0.2%**	**0.00040994**	**0.039817266**
UBE2W	**ubiquitin-conjugating enzyme E2W (putative)**	**999**	**0.9%**	**2.15E-05**	**0.00616651**

Supplementary Table S13 | Significant GO categories of *M. brandtii* positively selected genes. Benjamini-Hochberg adjusted p-value < 0.05. BP = biological process; CC = cellular compartment.

GO categories	Description	Taxonomy	Number of Genes (all)	Number of Genes (in list)	P-value (hyper-geometric)	Corrected P-value (Benjamini-Hochberg)
GO:0006956	complement activation	BP	32	4	0.0004	0.0469
GO:0006959	humoral immune response	BP	62	5	0.0006	0.0469
GO:0002541	activation of plasma proteins involved in acute inflammatory response	BP	35	4	0.0005	0.0469
GO:0009566	Fertilization	BP	50	5	0.0002	0.0469
GO:0006958	complement activation, classical pathway	BP	25	4	0.0001	0.0469
GO:0002455	humoral immune response mediated by circulating immunoglobulin	BP	30	4	0.0003	0.0469
GO:0005576	extracellular region	CC	1182	5	0.0003	0.0146
GO:0032444	activin responsive factor complex	CC	3	2	0.0003	0.0146
GO:0015629	actin cytoskeleton	CC	202	8	0.0011	0.0267
GO:0005886	plasma membrane	CC	2441	40	0.0009	0.0267
GO:0016459	myosin complex	CC	53	4	0.0021	0.0404
GO:0031226	intrinsic to plasma membrane	CC	919	19	0.0025	0.0404
GO:0044459	plasma membrane part	CC	1468	26	0.0035	0.0485

Supplementary Table S14 | Proteins that uniquely changed amino acids in *Myotis*. AA changes correspond to residue locations in the translated human RefSeq mRNA entry (typically the longest RefSeq entry of each gene). *M. lucifugus* proteins correspond to ENSEMBL IDs. Known and putative functions are listed. All proteins were manually verified.

Gene name	Symbol	*M.brandtii* protein	M.lucifugus protein	human RefSeq	AA change(s)	Function(s)
adenomatous polyposis coli	*APC*	10033156	ENSMLUG00000014785	NM_001127510	D2684N RH2721C	tumor suppressor
cyclic nucleotide gated channel beta 3	*CNGB3*	-	ENSMLUG00000000019	NM_019098	Y359C	channel function in cone photoreceptors
DEAD (Asp-Glu-Ala-Asp) box helicase 1	*DDX1*	10020169	ENSMLUG00000003027	NM_004939	K502R Q713H	RNA helicase. Template-guided repair of transcriptionally active regions of the genome
DENN/MADD domain containing 1B	*DENND1B*	10008239	ENSMLUG00000002301	NM_144977	S340C	activates Rab GTPases. Immune function
Dystonin	*DST*	10026669	ENSMLUG00000005610	NM_015548	T210K/R L/M2806R	cytoskeleton organisation
enhanced RNAI (RNA interference) family member 3	*ERI3*	10008306	ENSMLUG00000003870	NM_024066	H161Q	regulation of RNA interference (RNAi)
follicle stimulating hormone, beta polypeptide	*FSHB*	10028764	ENSMLUG00000015565	NM_000510	ST107R	development of ovarian follicle and spermatogenesis in the reproductive organs. Residue implicated in binding with FSHR
helicase (DNA) B	*HELB*	10022899	ENSMLUG00000017176	NM_033647	C280F	facilitates cellular recovery from replication stress
mediator complex subunit 13-like	*MED13L*	10031057	ENSMLUG00000000740	NM_015335	R2775W	Mediator complex subunit, coregulator or RNA PolII transcription
mediator complex subunit 23	*MED23*	10035672	ENSMLUG00000016810	NM_004830	D775N P787del	Mediator complex subunit, coregulator or RNA PolII transcription
non-SMC condensin II complex, subunit D3	*NCAPD3*	10003684	ENSMLUG00000009710	NM_015261	HR1020Q	mitotic chromosome assembly
SET and MYND domain containing 4	*SMYD4*	10015717	ENSMLUG00000006108	NM_052928	FL360C	development of muscle
slingshot homolog 2	*SSH2*	10015941	ENSMLUG00000003080	NM_033389	SG1398H	regulates actin filament dynamics
uridine monophosphate synthetase	*UMPS*	10032668	ENSG00000114491	NM_000373	P360S	catalyses the final two steps of the de novo pyrimidine biosynthetic pathway
bombesin-like receptor 3	*BRS3*	10013462	ENSG00000102239	NM_001727	T272I	GPCR with physiological effects including regulation of reproduction and insulin secretion
vacuolar protein sorting 4 homolog B (S. cerevisiae)	*VPS4B*	10020306	ENSG00000119541	NM_004869	V353I	mediates endosomal membrane protein sorting. Host factor for viral budding: hijacked by viruses such as HIV
putative homeodomain transcription factor 1	*PHTF1*	10029989	ENSG00000116793	NM_006608	N694T Y732F	differentiation of male germ cell and spermatogenesis
protease, serine, 23	*PRSS23*	10014366	ENSG00000150687	NM_007173	P287L	protease involved in ovulation
histidyl-tRNA synthetase 2, mitochondrial (putative)	*HARS2*	10027842	ENSG00000112855	NM_012208	L140M	mitochondrial translation. Mutations of adjacent residue 141, implicated in binding with histidyl-adenosine monophosphate (HAM), causes ovarian dysfunction and hearing loss
DEAD (Asp-Glu-Ala-Asp) box polypeptide 4	*DDX4*	10010767	ENSG00000152670	NM_024415	T543A	RNA helicase. Expressed in reproductive tissues of both sexes. May function in sperm motility
chromosome 6 open reading frame 62	*C6orf62*	10006523	ENSG00000112308	NM_030939	M227L	unknown

Supplementary Table S15 | Data statistics of *M. brandtii* transcriptomes.
H2 refers to bats hibernating for 2 months, H6 to bats hibernating for 6 months, and A to summer active bats. Replicate samples represent different animals.

Sample	TotalReads(M)	MappedReads(M)	Reads(%)	TotalBases(G)	MappedBases(G)	Bases(%)
H6 liver 1 RNA	47.18	31.43	66.61	4.25	2.83	66.61
H6 liver 2 RNA	37.51	26.63	70.99	3.38	2.40	70.99
H2 liver 1 RNA	65.08	44.05	67.69	5.86	3.96	67.69
H2 kidney 1 RNA	68.62	41.06	59.84	6.18	3.70	59.84
H2 brain 1 RNA	64.92	45.45	70.01	5.84	4.09	70.01
H2 liver 2 RNA	65.04	42.61	65.52	5.85	3.84	65.52
H2 kidney 2 RNA	64.52	39.35	60.99	5.81	3.54	60.99
H2 brain 2 RNA	67.79	49.27	72.68	6.10	4.43	72.68
A liver 1 RNA	58.90	37.85	64.26	5.30	3.41	64.26
A kidney 1 RNA	73.70	43.20	58.62	6.63	3.89	58.62
A liver 2 RNA	73.54	44.49	60.50	6.62	4.00	60.50
A kidney 2 RNA	77.01	44.71	58.06	6.93	4.02	58.06
A brain 1 RNA	74.25	55.73	75.05	6.68	5.02	75.05
H6 kidney 2 RNA	69.46	41.36	59.54	6.25	3.72	59.54
H6 kidney 1 RNA	63.86	41.56	65.08	5.75	3.74	65.08

Supplementary Table S16 | Insulin/IGF1 gene expression of *M. brandtii* and *GHR*$^{-/-}$ mice. The log2 fold change in gene expression in the liver of the *M. brandtii* (Brandt's bat; both active and hibernating) versus mice was calculated, with values of at least ±1.5 highlighted in red (increase) and green (decrease). Adjusted p-value < 0.05 (by Benjamini-Hochberg procedure) were highlighted in yellow. Values were calculated using R-package edgeR. *GHR*$^{-/-}$ mice: the previously reported direction of changes (up- or down-regulation with respect to wild-type mice); Log2(FC): normalized log 2 fold change.

Mouse Gene Symbol	Description	GHR-/- *vs* WT mice	Bat (active) *vs* Mouse		Bat (hibernating) *vs* Mouse	
			Log2 (FC)	Adjusted p-value	Log2 (FC)	Adjusted p-value
Akt1	thymoma viral proto-oncogene 1	Up	-2.4	1.18E-07	-2.6	1.10E-10
Akt2	thymoma viral proto-oncogene 2	Up	2.1	2.38E-09	2.2	1.28E-09
Foxo1	forkhead box O1	Up	2.5	3.35E-10	2.1	5.63E-06
G6pc (G6Pase)	glucose-6-phosphatase, catalytic	Up	3.7	3.98E-13	2.7	6.73E-07
Insr	insulin receptor	Up	1.6	2.55E-07	1.0	0.001632
Irs1	insulin receptor substrate 1	Up	1.5	0.020234	1.1	0.095843
Irs2	insulin receptor substrate 2	Up	2.7	6.26E-08	2.4	0.000124
Pck2 (PEPCK)	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	Up	7.2	6.26E-64	7.0	7.46E-47
Ppargc1a (PGC-1α)	peroxisome proliferative activated receptor, gamma, 1 alpha	Up	2.7	7.71E-06	1.5	0.110424
Prkaa1 (AMPK)	protein kinase, AMP-activated, alpha 1 catalytic subunit	Up	1.0	0.174343	0.8	0.279847
Slc2a2 (GLUT2)	solute carrier family 2 (facilitated glucose transporter), member 2	Up	-0.6	0.14345	-1.9	0.000143

Supplementary Table S17 | Comparison of wildtype mouse and long-lived mice and bat. The log2 fold change in gene expression in the liver tissues of the Brandt's bat *M. brandtii* (both active and hibernating) versus wild type (WT) mice was calculated, with values of at least ±1.5 highlighted in red (increase) and green (decrease). Adjusted p-value < 0.05 (by Benjamini-Hochberg procedure) were highlighted in yellow. Values were calculated using the R-package edgeR. Long-lived Mouse *vs* WT mice: the previously reported direction of change in long-lived mice (up- or down-regulation) with respect to wild-type mice; Log2(FC): normalized log 2 fold change.

Mouse Gene Symbol	Description	Long-lived Mouse *vs* WT Mouse	Bat (active) *vs* Mouse		Bat (hibernating) *vs* Mouse	
			Log2 (FC)	Adjusted p-value	Log2 (FC)	Adjusted p-value
Alas2	aminolevulinic acid synthase 2	Down	-8.7	1.74E-07	-8.1	5.88E-13
Apoa4	apolipoprotein A-IV	Down	0.3	1	1.2	0.283652
Cmas	cytidine monophospho-N-acetylneuraminic acid synthetase	Down	-8.3	6.69E-14	-8.8	3.11E-21
Dclre1a	DNA cross-link repair 1A, PSO2 homolog	Up	1.5	0.018106	1.3	0.058231
Dpp7	dipeptidylpeptidase 7	Down	2.8	2.12E-18	2.6	9.15E-14
Egfr	epidermal growth factor receptor	Down	-2.4	9.50E-05	-2.7	1.41E-07
Fabp2	fatty acid binding protein 2	Down	-8.8	2.26E-22	-9.3	8.27E-34
Ero1lb	ERO1-like beta	Down	2.9	1.99E-12	3.5	1.28E-11
Hsd17b12	hydroxysteroid (17-beta) dehydrogenase 12	Down	-0.9	0.003325	-1.9	1.47E-09
Hsd17b2	hydroxysteroid (17-beta) dehydrogenase 2	Down	-8.4	8.11E-38	-6.3	1.17E-46
Igf1	insulin-like growth factor 1	Down	0.1	1	0.4	0.478455
Igfals	insulin-like growth factor binding protein, acid labile subunit	Down	-6.2	4.52E-16	-5.9	8.97E-25
Lifr	leukemia inhibitory factor receptor	Down	-2.6	4.37E-11	-3.0	2.97E-18
Serpina12	serine peptidase inhibitor, clade A member 12	Down	-1.9	0.007352	-3.0	1.62E-05
Tfpi2	tissue factor pathway inhibitor 2	Down	-1.9	0.001159	-2.1	4.82E-05

SUPPLEMENTARY NOTE 1: SPECIES RANGE MAP

The species range map was obtained from the IUCN Red List of Threatened Species[61].

SUPPLEMENTARY NOTE 2: VISION

The retinas of mammals contain rod and cone photoreceptors for night vision and for daylight vision, respectively[62]. It is less established what kind of color vision adaptations individual species have accumulated. Bats are among the most numerous group of mammals (with regard to species) and have adapted to occupy a wide range of habitats since their common ancestor. The current evidence suggests that the echolocating bats, including *Myotis*, possess dichromatic color vision, dim-light sensitive rods and UV-sensitive cones[63-66].

RHO – rhodopsin

The ability to see under dim-light conditions has been largely attributed to the rod visual pigment rhodopsin, whose gene (*RHO*) is highly conserved in vertebrates[67]. A recent study reported that several bat branches share amino acid changes that map to the transmembrane domains and the intradiscal space of rhodopsin[66]. Our analysis of *M. brandtii* and *M. lucifugus* sequences confirms all findings in the previous study, except that residue 286 previously found to be unique to the *Myotis* and Hipposideridae families, is not changed in *M. brandtii* and *M. lucifugus* (Supplementary Fig. S12).

OPN1SW – Short-wavelength sensitive opsin/S opsin

UV is more abundant at dawn and dusk and the mouse is able to tune its short wavelength sensitive opsin/S opsin to UV by altering a set of amino acids in the protein[68,69]. It has recently been shown that the megabat *Pteropus dasymallus* and several echolocating species harbor the same amino acids at these sites as the mouse. In contrast, loss of the S opsin gene (*OPN1SW*) has been associated with the development of high-duty cycle echolocation in bats, while low-duty echolocating bat species, such as *Myotis*, retain the gene[65]. High-duty cycle echolocation is considered an energy expensive, more advanced adaptation of low-duty echolocation that allowed bats to hunt more efficiently in dense forests and other cluttered habitats[70]. We found that the echolocating *M. brandtii* and *M. lucifugus* bat harbor the same S opsin-tuning residues as the non-echolocating megabat *Pteropus vampyrus* and mouse (Fig. 3). This suggests that *M. brandtii* has UV vision.

SUPPLEMENTARY NOTE 3: Comparison with insulin/IGF1 signaling gene expression in the *GHR*$^{-/-}$ mouse

Several studies have reported the changes in mRNA and protein levels in the liver of *GHR*$^{-/-}$ mice as compared to wild-type controls[44,71-73]. Out of 13 genes previously reported to be differentially expressed in the liver of *GHR*$^{-/-}$ mice, we analyzed the genes that had unique one-to-one orthologs in *M. brandtii*. We observed up-regulation of genes including *AKT2*, *FOXO1*, G6Pase, insulin receptor substrate 2, *PEPCK* and *PGC-1α* in the Brandt's bat compared to mouse (Supplementary Table S17), which was largely in-line with the effect of *GHR* knockout in mice.

SUPPLEMENTARY NOTE 4: PANCREAS HISTOLOGY

The pancreas was obtained from a bat that had hibernated for six months. The α-cell/β-cell ratio is known to be unaltered in hibernating bats[74,75]. Diminished size of pancreatic islets of Langerhans in *M. brandtii* is shown in Supplementary Fig. S16. *M. brandtii* has tiny islets and the average number of cells of each islet is 5.65±5 with median value of 3.5 (n=40).

SUPPLEMENTARY NOTE 5: Comparison of differences in gene expression of the Brandt's bat, wild type mice and long-lived mice

Long-lived strains of dwarf mice are typified by impaired signaling of the GH/IGF1 axis. These changes can occur at many steps along the signaling pathway, from the inhibition of anterior pituitary development, to the impaired sensitivity to growth hormone releasing hormone, and to the disruption of the growth hormone receptor. Compared with wild-type mice, these dwarf mice live longer, are less susceptible to age-related declines in bodily functions, and have lower rates of cancer. Profiling of gene expression in the liver tissues of long-lived dwarf mice suggested a repertoire of genes that were differentially expressed in the dwarf mice compared with wild-type mice[46]. Many of these genes were also differentially expressed in mice undergoing caloric restriction, suggesting there might be a common mechanism underlying the observed extension in lifespan.

Gene expression in *M. brandtii* liver showed a similar pattern to that in long-lived mice. Out of the 43 genes previously reported to be differentially expressed in long-lived mice[46], we identified 15 genes that had unique one-to-one orthologs in *M. brandtii*. We compared the expression levels of these genes in the livers of *M. brandtii* and mice and found that 11 of the 15 genes followed the same directions of change as reported in long-lived mice (Supplementary Table S18).

SUPPLEMENTARY METHODS

Genome sequencing and assembly

Genome sequencing

To sequence the genome of *M. brandtii*, we applied a whole genome shotgun strategy and next-generation sequencing technologies using Illumina HiSeq 2000. In order to reduce the risk of non-randomness, we constructed 12 paired-end libraries by sequencing 12 lanes with insert sizes of about 170 base pairs (bp), 500 bp, 800 bp, 2 kbp, 5 kbp, 10 kbp, and 20 kbp. In total, we generated around 258 Gb of sequence, and 156 Gb (78 x coverage) was retained for assembly after filtering out low-quality and duplicated reads.

Estimation of genome size using k-mer

A k-mer refers to an artificial sequence division of K nucleotides. A raw sequence read with L bp contains (L-K+1) k-mers if the length of each k-mer is K bp. The frequency of each k-mer can be calculated from the genome sequence reads. Typically, k-mer frequencies plotted against the sequence depth gradient follow a Poisson distribution in any given dataset, whereas sequencing errors may lead to a higher representation of low frequencies. The genome size, G, can be calculated from the formula G=K_num/K_depth, where the K_num is the total number of k-mers, and K_depth denotes the frequency occurring more frequently than the other frequencies.

In this work, K was 17, K_num was 83,436,283,306 and K_depth was 44. We, therefore, estimated the *M. brandtii* genome size to be 1.9 Gb.

Nuclear genome assembly

The genome was *de novo* assembled by SOAPdenovo[56] (http://soap.genomics.org.cn). First, low-quality reads and those with potential sequencing errors were removed or corrected using the k-mer frequency. After these quality control and filtering steps, a total of 156 Gb (or 78 x) of data were retained for assembly. We assembled the genome using a previously described procedure[76]. About 101 Gb (or 50 x) of data were used to build contigs, while all high-quality data were used to build scaffolds. Some intra-scaffold gaps were filled by local assembly using the reads in a read-pair, where one end was uniquely aligned to a contig while the other end was located within the gap. The final total contig size and N50 were 2.05 Gbp and 21 Kbp, respectively. The total scaffold size and N50 were 2.18 Gb and 3.1 Mbp, respectively.

To access assembly quality, high quality short insert size reads that satisfied our filtering criteria were aligned onto the assembly using BWA[77] with parameters of -I. A total of 72 G (or 36 x) of high quality short insert size reads were used of which 96.28% reads could be mapped, and they covered 98.97% of the assembly, excluding gaps. Moreover, 91.64% of the assembly could be covered by more than 10 reads, and about 94.39% of all mapped reads were proper paired with the reasonable insert size

associated with the library built, suggesting a high quality assembly of high completeness and high accuracy.

Mitochondrial genome assembly

The mitochondrial genome was assembled using SOAPdenovo[56] with default settings, and verified by multiple alignment with 14 bat mitochondrial genomes available at NCBI (*Artibeus jamaicensis, Pteropus dasymallus, Pteropus scapulatus, Chalinolobus tuberculatus, Pipistrellus abramus, Rhinolophus pumilus, Rhinolophus monoceros, Mystacina tuberculata, Rousettus aegyptiacus, Rhinolophus formosae, Artibeus lituratus, Plecotus rafinesquii, Lasiurus borealis, Myotis formosus*). The size of the Brandt's bat mitochondrial genome is 16,868 nucleotides, which is consistent with the data for other species.

Cysteine-content of mitochondrial protein-coding genes

The data for Figure 4c and Supplementary Fig. S18 were extracted from the AnAge database (http://genomics.senescence.info, October 2012) and analyzed using a custom Perl script. A total of 1308 mammalian species were used in the analysis of body mass and longevity correlation, for which the data were known.

In addition to *M. brandtii*, the following 95 species were used for the analysis of cysteine content of mitochondrial protein-coding genes: *Balaenoptera physalus*, *Balaenoptera musculus*, *Homo sapiens*, *Megaptera novaeangliae*, *Elephas maximus*,

Loxodonta africana, *Eschrichtius robustus*, *Physeter catodon*, *Balaenoptera borealis*, *Berardius bairdii*, *Dugong dugon*, *Equus caballus, Pan troglodytes*, *Pongo pygmaeus*, *Hippopotamus amphibius*, *Gorilla gorilla*, *Balaenoptera acutorostrata*, *Ceratotherium simum*, *Pan paniscus*, *Tachyglossus aculeatus*, *Ursus arctos*, *Rhinoceros unicornis*, *Equus asinus*, *Halichoerus grypus*, *Cebus albifrons*, *Balaena mysticetus*, *Hylobates lar*, *Monodon monoceros*, *Odobenus rosmarus*, *Phoca vitulina*, *Ursus maritimus*, *Hyperoodon ampullatus*, *Papio hamadryas*, *Macaca mulatta*, *Tapirus terrestris*, *Felis catus*, *Ursus americanus*, *Cercopithecus aethiops*, *Zaglossus bruijni*, *Bos taurus*, *Canis familiaris*, *Eumetopias jubatus*, *Inia geoffrensis*, *Choloepus didactylus*, *Platanista minor*, *Lemur catta*, *Nycticebus coucang*, *Vombatus ursinus, Bos grunniens*, *Trachypithecus obscurus*, *Colobus guereza*, *Ovis aries*, *Orycteropus afer*, *Macaca sylvanus*, *Capra hircus*, *Sus scrofa*, *Macropus robustus*, *Acinonyx jubatus*, *Muntiacus muntjak*, *Oryctolagus cuniculus*, *Kogia breviceps*, *Ornithorhynchus anatinus*, *Pontoporia blainvillei*, *Cavia porcellus*, *Dasypus novemcinctus*, *Trichosurus vulpecula*, *Erinaceus europaeus*, *Procavia capensis*, *Echinops telfairi*, *Phocoena phocoena*, *Lepus europaeus*, *Sciurus vulgaris*, *Artibeus jamaicensis*, *Tupaia belangeri*, *Tamandua tetradactyla*, *Herpestes javanicus*, *Pseudocheirus peregrinus*, *Tarsius bancanus*, *Didelphis virginiana*, *Hemiechinus auritus*, *Macrotis lagotis*, *Ochotona princeps*, *Monodelphis domestica*, *Mus musculus*, *Ochotona collaris*, *Echinosorex gymnura*, *Metachirus nudicaudatus*, *Rattus norvegicus*, *Talpa europaea*, *Thryonomys swinderianus*, *Crocidura russula*, *Isoodon macrourus*, *Perameles gunnii*, *Tarsipes rostratus*, and *Notoryctes typhlops*.

Genome features

Heterozygosity analysis

We aligned all high quality short insert size reads to the assembly using BWA[77] with parameters -I. Since the alignment results were stored in BAM/SAM format, SAMtools, which is based on the Bayesian model, was selected for variation analysis[60]. After sorting alignments by leftmost coordinates and removing potential PCR duplicates, we used SAMtools mpileup to call SNPs and short InDels. We rejected SNPs and InDels within reads with depth either much lower or much higher than expected, since large copy number variation might lead to miscalling of SNPs. The sequencing depth ranged from 4 to 100, and the upper limit was about triple the sequencing depth. The SNP-miscalling caused by alignment around short InDels and low-quality sequences were removed. We applied samtools.pl varFilter as the filter tool, which can be found in the SAMtools package, with parameters -Q 20 -q 20 -d 4 -D 100 -S 20 -i 20 -N 5 -l 5 -W 5 -N 1.

Repeat annotation

We employed RepeatMasker[78] to identify and classify transposable elements (TEs) by aligning the *M. brandtii* genome sequences against a library of known repeats, Repbase (http://www.girinst.org/repbase/), with default parameters. To better compare the *M. brandtii* genome with those of other mammals, we used the same pipeline and

parameters to re-run the repeat annotations on the human, macaque, mouse, rat and dog genomes.

Gene annotation and evaluation of gene quality

To predict genes encoded by the *M. brandtii* genome, we used both homology-based and *de novo* methods. For the homology-based prediction, human, mouse, dog, cow and horse proteins were downloaded from Ensembl (release 64) and mapped onto the genome using BLAT[79]. Homologous genome sequences were then aligned against the matching proteins using GeneWise[80] to define gene models. For *de novo* prediction, Augustus[81], Genescan[82] and SNAP[83] were employed to predict coding genes, using appropriate parameters. RNA-seq data were mapped to genome using TopHat[58], and transcriptome-based gene structures were obtained by cufflinks (http://cufflinks.cbcb.umd.edu). Finally, homology-based, *de novo* derived gene sets and transcript gene sets were merged to form a comprehensive and non-redundant reference gene set using GLEAN (http://sourceforge.net/projects/glean-gene/), removing all genes that only had weak *de novo* support. We obtained a reference *M. brandtii* gene set containing 22,256 genes. Of these, 12,986 genes (58.35%) had three types of evidences and 20,910 genes (93.95%) had at least two types of evidence. Gene structures generated by GLEAN that differed from those generated by orthology analysis were subject to manual inspection and correction; a total 2,051 genes were refined.

Functional annotation of *M. brandtii* genes

Gene functions were assigned according to the best match of the alignments using BLASTp against SWISS-PROT databases[84]. Gene Ontology IDs for each gene were obtained from the corresponding InterPro entry[85,86]. All genes were aligned against KEGG proteins, and the pathway in which the gene might be involved was derived from the matching genes in KEGG[87].

Orthology relationship of *M. brandtii* and related mammals

To determine orthology relationships between *M. brandtii* and other mammalian proteins, nucleotide and protein data for six mammals (human/*Homo sapiens*, mouse/*Mus musculus*, rat/*Rattus norvegicus*, cow/*Bos taurus*, dog/*Canis lupus familiaris*, and horse/*Equus caballus*) were downloaded from the Ensembl database (release 64). For genes with alternative splicing variants, the longest transcripts were selected to represent the genes. We then subjected all proteins to BLASTp analysis with the similarity cutoff of e=1e-5. With the *M. brandtii* protein set used as a reference, we found the best hit for each *M. brandtii* protein in other species, with the criteria that more than 30% of the aligned sequence showed an identity above 30%. Reciprocal best-match pairs were defined as orthologs.

Gene evolution

Gene family cluster

A gene family is a group of similar genes descended from a single gene in the last common ancestor of the targeted species. In this study, we used TreeFam (http://www.treefam.org)[88] to define gene families among 9 mammalian genomes (human, rhesus macaque, mouse, rat, cow, dog, cat, horse and *M. brandtii*). We employed the same pipeline and parameters as in our previously published studies[7,76] and obtained 11,527 gene families and 2,654 single-copy orthologs.

Phylogenetic analysis

We constructed a phylogenetic tree of *M. brandtii* and the 8 other mammalian genomes (human, rhesus macaque, mouse, rat, cow, dog, cat and horse). A total of 2,654 single-copy gene families obtained as described above were used to reconstruct the phylogenetic tree. Coding sequences (CDS) from each single-copy family were aligned by MUSCLE[89] with the guidance of aligned protein sequences and concatenated to one super-gene for each species. Then, PhyML[90] was applied for these sequence sets to build a phylogenetic tree under GTR+gamma. We employed 1,000 rapid bootstrap replicated to assess the branch reliability (Supplementary Fig. S5). We also used codon 1,2 and 1+2 sequences extracted from the CDS alignment and used as input for building trees, along with protein sequences, and all results

indicated that *M. brandtii* and horse cluster together with high bootstrap. To estimate the divergence time, 4-fold degenerate sites were extracted from each single copy gene family and concatenated to one supergene for one species. We used PAML's MCMCtree[91,92] to determine split times with approximate likelihood calculation.

Gene family expansion and contraction

Gene family expansion analysis was performed by using CAFE 2.1[93]. A random birth and death model was proposed to study gene gain and loss in gene families across a user-specified phylogenetic tree. A global parameter λ, which described both the gene birth (λ) and death (μ = -λ) rates across all branches of the tree for all gene families, was estimated using maximum likelihood. Then, a conditional p-value was calculated for each gene family, and families with conditional p-values less than the threshold (0.05) were considered to have an accelerated rate of expansion and contraction. Our analysis revealed that a total of 67 gene families underwent expansion and 44 gene families underwent contraction in the *M. brandtii* lineage.

***M. brandtii* gained and lost genes**

To determine the orthologous relationship between *M. brandtii* and human proteins, sequences of the human protein dataset were downloaded from Ensembl (release 64). The longest transcript was chosen to represent each gene with alternative splicing

variants. We then subjected human and *M. brandtii* proteins to BLASTp analysis with the similarity cutoff threshold of e-value=$1e^{-5}$. With the human protein set as a reference, we found the best hit for each protein of *M. brandtii*, with the criteria that more than 30% of the aligned sequence showed an identity above 30%. Reciprocal best-match pairs were defined as orthologs. Then, gene order information was used to filter out false positive orthologs caused by draft genome assembly and annotation. Orthologs not in gene synteny blocks were removed from further analysis. For example, in the case of 3 continuous genes in the human genome A, B and C, even if all three orthologs could be identified in both humans and *M. brandtii* based on the cutoff threshold described above, but the B gene in the *M. brandtii* genome was not located between A and C genes, it was removed filtered out, as it could be located in another scaffold or another place within the same scaffold. Using this method, we identified gene synteny relationships for human and *M. brandtii* lineages.

Orthology information was obtained as described above. Since it showed synteny information at the protein level, it could be used to analyze gene-gain and -loss between human and *M. brandtii* lineages. In the protein synteny blocks, if a human protein had no *M. brandtii* ortholog, and excluding false positive predictions that could be caused by annotation or genome assembly (gap > 5%), this protein could be defined as either being lost in the *M. brandtii* lineage or gained in the human lineage. Using *M. brandtii* as a reference to generate the orthology relationship, we applied this procedure to identify the genes gained in the *M. brandtii* lineage compared to the

human lineage.

M. brandtii pseudogenes

To detect homozygous pseudogenes in the *M. brandtii* genome *in silico*, we first aligned all human genes (protein sequences downloaded from Ensembl (release 64) onto the *M. brandtii* genome using BLASTp with parameters (-F F -e 1e-5). Solar[94] was used to conjoin the fragmental alignments for each gene. Best hit regions of each gene with 5 kb flanking sequence were cut down and re-aligned with their corresponding human orthologous protein sequences using GeneWise with parameters (-genesf -tfor -quiet), which helped define the detailed exon-intron structure of each gene. Genes with frameshifts and premature stop codons as reported by GeneWise were considered candidate pseudogenes. We further carried out a series of filtering processes: (i) To avoid the reported frameshifts and premature stop codons which were due to flaws in the GeneWise algorithm, we also aligned all human proteins to their corresponding loci in the human genome using GeneWise as a control, and genes with frameshifts and premature stop codons in the human-to-human alignment reported by GeneWise were filtered out; (ii) Using the results of the human-to-human alignment from GeneWise, candidate pseudogenes with obvious splicing errors near their frameshifts and premature stop codons were filtered out; (iii) Candidate pseudogenes with a low number of reads covering their frameshift and premature stop codon sites were considered assembly errors. In addition, cases with a considerable number of reads resulting from different genotypes at these sites were treated as

heterozygous. The cases of assembly error and heterozygosis were filtered out. In total, we identified 194 pseudogenes in the *M. brandtii* genome. By comparison with all alternative splice forms in humans, we found that 153 pseudogenes with the pseudo-mutations could be compensated by alternative splicing.

Identification of proteins with unique amino acid changes

In addition to the Brandt's bat (*M. brandtii*), the little brown bat (*M. lucifugus*)[95] and the naked mole rat (*Heterocephalus glaber*)[7], the following organisms obtained via the UCSC multiway track[96] were examined (the full common name includes the text in brackets): the megabat/fruit bat/flying fox *Pteropus vampyrus*, human (*Homo sapiens)*, (common) chimpanzee (*Pan troglodytes*), (Sumatran) orangutan (*Pongo pygmaeus abelii*), rhesus monkey (*Macaca mulatta*), (hamadryas) baboon (*Papio hamadryas*), (Western lowland) gorilla (*Gorilla gorilla gorilla*), (common) marmoset (*Callithrix jacchus*), (Philippine) tarsier (*Tarsier syrichta*), (gray) mouse lemur (*Microcebus murinus*), bushbaby/small-eared greater galago (*Otolemur garnettii*), (Northern) tree shrew (*Tupaia belangeri*), (house) mouse (*Mus musculus*), (brown) rat (*Rattus norvegicus*), (Ord's) kangaroo rat (*Dipodomys ordii*), guinea pig (*Cavia porcellus*), (thirteen-lined ground) squirrel (*Spermophilus tridecemlineatus*), (European) rabbit (*Oryctolagus cuniculus*), (American) pika (*Ochotona princeps*), alpaca (*Vicugna pacos*), cow (*Bos taurus*), (Atlantic bottlenose) dolphin (*Tursiops truncatus*), horse (*Equus caballus*), cat (*Felis catus*), dog (*Canis lupus familiaris*), (European) hedgehog (*Erinaceus europaeus*), (common) shrew (*Sorex araneus*), rock

hyrax (*Procavia capensis*), (African) elephant (*Loxodonta africana*), (Hoffmann's two-toed) sloth (*Choloepus hoffmanni*), (lesser hedgehog) tenrec (*Echinops telfairi*), (nine-banded) armadillo (*Dasypus novemcinctus*), (tammar) wallaby (*Macropus eugenii*), (gray short-tailed) opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), (green anole) lizard (Anolis carolinensis), chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and (Western clawed) frog (*Xenopus tropicalis)*. We acknowledge the contribution of numerous research centers in generating genome sequences for these organisms. Of particular relevance to our current work are the genomes of bats *M. lucifugus* and *P. vampyrus* generated by the Broad Institute and Baylor College of Medicine, respectively[81]. To convert predicted *Myotis* sequences into human RefSeq IDs, which are employed in the UCSC multiway alignments, NCBI tBLASTn in v2.2.26+ of the BLAST+ suite[97] (protein sequences queried against a translated human RefSeq nucleotide database) with an e-value cut-off set at 10e-5 was employed. The best match, with at least 50% overall amino acid identity (along the entire sequence) and spanning ≥75% of the length of the query sequence, was used to rename the sequences. *Myotis* proteins were aligned to their orthologs using Clustal W v2.1[98]. In-house Perl scripts were used to parse the Clustal output and identify unique amino acids.

Because of extensive exon shuffling and gene duplications in teleost fish genomes[99-101], proteins that changed uniquely in *Myotis* compared to other tetrapods were independently aligned to their candidate fish medaka/Japanese rice fish (*Oryzias*

latipes), (three-spined) stickleback (*Gasterosteus aculeatus*), (tora)fugu (*Takifugu rubripes*), tetraodon/green spotted puffer (*Tetraodon nigroviridis*), and zebrafish (*Danio rerio*), as well as sea lamprey (*Petromyzon marinus*) orthologs.

The false positive rate of the detection of unique amino acids was estimated as follows. The estimated error rate for SOAPdenovo assembly was assumed to be 1 nucleotide per 81,025 base pairs[102]. Since only coding regions were used for amino acid conservation analysis, and total predicted size of coding sequences (CDS) of the Brandt's bat is approximately 33 Mb, then 418 nucleotides could be considered as candidates for false positives. Other factors (such as potential BLAST misalignment) are of minor importance, as parameters used in analysis were quite strict (i.e., e-value 1e-5). The procedure utilized in unique amino acid detection calls for a particular amino acid to be conserved among the tested vertebrate genomes, but differ in genome-in-question (e.g., bats in the genus *Myotis*, or *Myotis* and the Atlantic bottlenose dolphin). On a test dataset, 3,287 amino acids were conserved among 36 genomes, out of 1,018,988 tested cases; therefore, only 0.32% of all sequences fit our search criteria. Taken together, with 418 candidates, one would expect a false positive rate of 1.33 per analysis.

Positively selected genes

1:1 orthologous coding sequence (mRNA) alignments using the UCSC Table Browser

data retrieval tool[103] were used to query UCSC multiway alignments[96] of megabat (*Pteropus vampyrus*), dog (*Canis lupus familaris*), horse (*Equus caballus*), cow (*Bos Taurus*), the Atlantic bottlenose dolphin (*Tursiops truncatus*), and human (*Homo sapiens*). BLAST and custom Perl scripts were used to convert *M. brandtii* FASTA headers to matching human RefSeq identifiers compatible with the UCSC multiway alignment data set. The phylogenetic relationships between the 7 tested taxa were obtained from[95]. High-quality alignments were obtained from PRANK v.111130 (settings: -shortnames +F -termgap -codon -f=fasta) alignments using Gblocks v0.91b (settings: -t=c -e=.fa -d=y -b5=N -b0=4). Consistent with previous studies[104-106], we employed PRANK (codon)[107] and Gblocks[108,109] to minimise the impact of multiple sequence alignment errors and divergent regions. The guide tree was estimated by the PRANK program since this method has been shown to be the most accurate for downstream PAML branch-site analysis of sequences of insertions and deletions[104]. For manipulations of sequences, we employed in-house Perl scripts and public BioPerl modules[110].

Because manual inspection of positive selection data is currently recommended in the literature[105,106], we manually examined the alignments of a subset of PAML results significant at a 1% false discovery rate. Large, divergent protein families, such as collagens, zinc finger proteins and olfactory receptors were omitted from the analysis. Results were discarded if the candidate PSG harbored coiled-coil domains as predicted by the program COILS[111].

We analysed the list of genes under positive selection for GO term enrichment over the background using the WebGestalt online analysis tool[112]. Analysis was performed on the human orthologs of the bat gene lists. P-values were computed with the hypergeometric test and multiple test correction performed by the Benjamini-Hochberg method, using a significance cut-off of $P < 0.05$ after multiple test correction.

PRANK, Gblocks and CodeML computations were run on the Orchestra supercomputing cluster supported by the Harvard Medical School Research Information Technology Group.

Analysis of individual genes

The recently released genome sequence of the bat *Eptesicus fuscus* (the big brown bat, GenBank accession code GCA_000308155), as well as transcriptome data from the big brown bat (liver), the Mexican free-tailed bat (*Tadarida brasiliensis*) (liver) and the Jamaican fruit bat *Artibeus jamaicenis* (liver, kidney, lung, primary kidney cells) (NCBI SRA accession code SRR539297)[9], which are all bats in the Yangochiroptera suborder of echolocating bats (the second and most numerous suborder of bats), were used to manually validate the results pertaining to specific genes if sequence data could be obtained. We acknowledge the efforts of numerous research centers in

generating genome and transcriptome sequences for these organisms.

Comparison of gene expression of *M. brandtii, GHR*$^{-/-}$ mice and other long-lived mice to that of wild type mice

We compared the RNA-seq data in liver tissues of *M. brandtii* with previously reported RNA-seq data in liver tissues of wild-type mice[113,114] (NCBI Short Read Archive SRP007412 and SRA001030). RPKM values were calculated for all genes that had unique one-to-one orthologs between bat and mouse (gene homology based on Ensembl Release 69) and normalized log 2 fold changes were computed using the R package edgeR[115]. We also examined the list of genes previously reported as differentially expressed in *GHR*$^{-/-}$ mice and long-lived dwarf mice. As the previously reported changes were qualitative[44,46,71-73], in our case we considered only log 2 fold changes of at least ±1.5 and determined whether they agreed with previously reported directions of change.

SUPPLEMENTARY REFERENCES

61. Hutson, A. M. *et al.* Myotis brandtii. In *IUCN Red List of Threatened Species*, Version 2012.1. http://www.iucnredlist.org. Downloaded on 27 September 2012. (2008).
62. Jacobs, G. H. Evolution of colour vision in mammals. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **364**, 2957-67 (2009).
63. Muller, B. *et al.* Bat eyes have ultraviolet-sensitive cone photoreceptors. *PloS ONE* **4**, e6390 (2009).
64. Wang, D. *et al.* Molecular evolution of bat color vision genes. *Mol. Biol. Evol.* **21**, 295-302 (2004).
65. Zhao, H. *et al.* The evolution of color vision in nocturnal mammals. *Proc. Natl Acad. Sci. USA* **106**, 8980-5 (2009).
66. Shen, Y. Y., Liu, J., Irwin, D. M. & Zhang, Y. P. Parallel and convergent evolution of the dim-light vision gene RH1 in bats (Order: Chiroptera). *PloS ONE* **5**, e8838 (2010).
67. Palczewski, K. Chemistry and biology of vision. *J. Biol. Chem.* **287**, 1612-9 (2012).
68. Shi, Y., Radlwimmer, F.B. & Yokoyama, S. Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc. Natl Acad. Sci. USA* **98**, 11731-6 (2001).
69. Yokoyama, S., Starmer, W.T., Takahashi, Y. & Tada, T. Tertiary structure and spectral tuning of UV and violet pigments in vertebrates. *Gene* **365**, 95-103 (2006).
70. Fenton, M. B., Faure, P. A. & Ratcliffe, J. M. Evolution of high duty cycle echolocation in bats. *J. Exp. Biol.* **215**, 2935-44 (2012).
71. Al-Regaiey, K. A., Masternak, M. M., Bonkowski, M., Sun, L. & Bartke, A. Long-lived growth hormone receptor knockout mice: interaction of reduced insulin-like growth factor i/insulin signaling and caloric restriction. *Endocrinology* **146**, 851-60 (2005).
72. Masternak, M. M. *et al.* Effects of caloric restriction on insulin pathway gene expression in the skeletal muscle and liver of normal and long-lived GHR-KO mice. *Exp. Gerontol.* **40**, 679-84 (2005).
73. Panici, J. A. *et al.* Is altered expression of hepatic insulin-related genes in growth hormone receptor knockout mice due to GH resistance or a difference in biological life spans? *J. Gerontol. A Biol. Sci. Med. Sci.* **64**, 1126-33 (2009).
74. Hinkley, R.E. & Burton, P.R. Fine structure of the pancreatic islet cells of normal and alloxan treated bats (Eptesicus fuscus). *Anat Rec.* **166**, 67-85 (1970).
75. Mosca, L. Changes in the islets of Langerhans associated with age and hibernation. *Exp. Physiol.* **41**, 433-441 (1956).
76. Li, R. *et al.* The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311-7 (2010).
77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
78. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. In *Current protocols in Bioinformatics*, vol. 25, chap. 4 (Wiley, 2009).
79. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-64 (2002).
80. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988-95 (2004).
81. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003)
82. Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in Drosophila genomic DNA. *Genome Res.* **10**, 516-22 (2000).
83. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

84. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45-8 (2000).
85. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-9 (2000).
86. Mulder, N. & Apweiler, R. InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol.* **396**, 59-70 (2007).
87. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
88. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572-80 (2006).
89. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-7 (2004).
90. Huelsenbeck, J.P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-5 (2001).
91. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
92. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212-26 (2006).
93. Hahn, M.W., De Bie, T., Stajich, J.E., Nguyen, C. & Cristianini, N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* **15**, 1153-60 (2005).
94. Yu, X. J., Zheng, H. K., Wang, J., Wang, W. & Su, B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics* **88**, 745-51 (2006).
95. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-82 (2011).
96. Miller, W. *et al.* 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res.* **17**, 1797-808 (2007).
97. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **40**, D13-25 (2012).
98. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).
99. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* **297**, 1301-10 (2002).
100. Jaillon, O. *et al.* Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946-57 (2004).
101. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714-9 (2007).
102. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* **22**, 549-56 (2012).
103. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493-6 (2004).
104. Fletcher, W. & Yang, Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol. Biol. Evol.* **27**, 2257-67 (2010).
105. Markova-Raina, P. & Petrov, D. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. *Genome Res.* **21**, 863-74 (2011).
106. Vamathevan, J. J. *et al.* The role of positive selection in determining the molecular cause of species

differences in disease. *BMC Evol. Biol.* **8**, 273 (2008).

107. Loytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-5 (2008).
108. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-52 (2000).
109. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564-77 (2007).
110. Stajich, J.E. *et al.* The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**, 1611-8 (2002).
111. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-4 (1991).
112. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741-8 (2005).
113. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-8 (2011).
114. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621-8 (2008).
115. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).