The evolution of color vision in nocturnal mammals

Huabin Zhao[a,b], Stephen J. Rossiter[c,1], Emma C. Teeling[d,1], Chanjuan Li[b], James A. Cotton[c], and Shuyi Zhang[b,1]

[a]Institute of Zoology and Graduate University, Chinese Academy of Sciences, Beijing 100080, China; [b]School of Life Sciences, East China Normal University, Shanghai 200062, China; [c]School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, United Kingdom; and [d]UCD School of Biology and Environmental Science and UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved April 10, 2009 (received for review December 24, 2008)

Nonfunctional visual genes are usually associated with species that inhabit poor light environments (aquatic/subterranean/nocturnal), and these genes are believed to have lost function through relaxed selection acting on the visual system. Indeed, the visual system is so adaptive that the reconstruction of intact ancestral opsin genes has been used to reject nocturnality in ancestral primates. To test these assertions, we examined the functionality of the short and medium- to long-wavelength opsin genes in a group of mammals that are supremely adapted to a nocturnal niche: the bats. We sequenced the visual cone opsin genes in 33 species of bat with diverse sensory ecologies and reconstructed their evolutionary history spanning 65 million years. We found that, whereas the long-wave opsin gene was conserved in all species, the short-wave opsin gene has undergone dramatic divergence among lineages. The occurrence of gene defects in the short-wave opsin gene leading to loss of function was found to directly coincide with the origin of high-duty-cycle echolocation and changes in roosting ecology in some lineages. Our findings indicate that both opsin genes have been under purifying selection in the majority bats despite a long history of nocturnality. However, when spectacular losses do occur, these result from an evolutionary sensory modality tradeoff, most likely driven by subtle shifts in ecological specialization rather than a nocturnal lifestyle. Our results suggest that UV color vision plays a considerably more important role in nocturnal mammalian sensory ecology than previously appreciated and highlight the caveat of inferring light environments from visual opsins and vice versa.

bats | opsin gene | sensory tradeoff | echolocation | selection

Vision plays one of the most important roles in the survival of an individual, underpinning numerous key behaviors such as foraging, predator avoidance, and mate recognition. Color vision is conferred by the cone photopigments, each comprising an opsin transmembrane protein and a 11-cis-retinal chromophore (1, 2). Diversity in the properties and arrangement of photoreceptors in vertebrates reflects the evolutionary malleability of this system in response to specific visual challenges (3). Opsin proteins can be classified into medium/long wavelength sensitive (M/LWS) and short-wavelength-sensitive (SWS) based on the wavelength of their peak light sensitivity. Comparisons of visual pigments across taxa indicate that spectral tuning and, therefore, the wavelength of peak light sensitivity (λ_{max}) are modulated by 5 key critical amino acid sites in M/LWS opsins (4) and at least 11-aa sites in SWS opsins (5).

Most mammals possess both classes of opsin, with the M/LWS sensitive to green-red and the SWS1 sensitive to blue-violet (6), and a greater proportion of cones containing the former (85–95%) than the latter (only 5–15%) (3). Reported exceptions to this visual state include a number of monochromats, such as the blind mole rat (7), cetaceans (8), and the flying squirrel (9), all of which have acquired loss-of-function mutations in their SWS1 opsin. Such losses have been typically explained by relaxed selection from inhabiting poor photopic environments (3). Recently, the presence of a functional SWS1 opsin in several lineages of prosimians has been used to refute the longstanding hypothesis of nocturnality in ancestral primates (10). To understand further how opsin genes have adapted in mammals and to investigate whether functionality can indeed be used to infer activity patterns, we undertook an extensive survey of visual genes in bats, which are considered the sensory specialists (11) and arguably show the greatest adaptation for nocturnality of all of the vertebrates.

The unique ability of most bats to orient using laryngeal echolocation without the need for vision, coupled with their characteristically small eyes, has led to speculation that laryngeal echolocation and nocturnality coevolved, and that an evolutionary "tradeoff" occurred between vision and hearing in bats (12). Using this logic, it is hypothesized that bats have occupied a nocturnal niche for >52 million years, because this is the age of the oldest bat fossil that shows evidence of echolocation capabilities (13). To infer the impact of nocturnality on the evolution of vision in mammals, we sequenced the *SWS1* opsin gene (2.2 kb) in 32 species of bat and the *M/LWS* opsin gene (3.2 kb) in 14 species of bats. We included bats from both major lineages (Yangochiroptera and Yinpterochiroptera) and species that varied in their acoustic and roosting ecology [see supporting information (SI) Table S1 and Fig. 1]. We analyzed our sequences alongside the published opsin sequences of 35 additional and phylogenetically diverse mammal species, representing the largest single analyses of mammalian opsins to date (Table S1). We predicted that bats should show evidence for loss of function throughout their phylogenetic tree because of their long history of nocturnality and evolution of echolocation.

Results

Sequence Alignment and Analyses Based on the ORF. ***M/LWS opsin.*** For the *M/LWS* opsin sequences, phylogenetic reconstruction based on both Maximum Likelihood and Bayesian methods supported the published consensus species tree for mammals (see Fig. 2). *M/LWS* sequences were highly conserved across all species examined, including bats, and showed an intact ORF and consensus mammalian intronic splice sites, suggesting the gene is functional. Maximum likelihood estimates of the ratio of nonsynonymous to synonymous substitution rates (d_N/d_S or ω) on each branch of the tree were uniformly low among bat lineages (0.00–0.48), indicating the *M/LWS* opsin gene has been subject to purifying selection during the radiation of bats (Table S2). Estimated omega values were also below 1 under other models of variation in selection across branches that we examined, confirming a lack of positive selection (Table S3). Examination of 5 critical sites revealed 2 substitutions that are known

Author contributions: H.Z., S.J.R., E.C.T., and S.Z. designed research; H.Z. and C.L. performed research; S.Z. contributed new reagents/analytic tools; H.Z., S.J.R., E.C.T., and J.A.C. analyzed data; and H.Z., S.J.R., E.C.T., and S.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EU912338–EU912383).

[1]To whom correspondence may be addressed. E-mail: syzhang@bio.ecnu.edu.cn, s.j.rossiter@qmul.ac.uk, or emma.teeling@ucd.ie.

This article contains supporting information online at www.pnas.org/cgi/content/full/0813201106/DCSupplemental.

Fig. 1. *SWS1* species tree showing substitution rates, indels and stop codons based on the ORF. The tree topology and divergence dates follow consensus published phylogenies (*SI Text*). Numbers of insertions and deletions in the ORF are shown on the branches by downward (▼) and upward oriented (▲) triangles, respectively. Ancestral stop codons were inferred by probabilistic reconstruction and are denoted by squares (■), and critical sites implicated in spectral tuning were obtained using the same method for all functional ancestral nodes predating stops. Stops found in the tip sequences are shown in Fig. 3. Key changes in the critical spectral tuning amino acid sites are shown by sequence logos in which the height of amino acids at a given position are proportional to their posterior probability. Species with sequences containing stop codons are shown in red font, whereas the loss of functionality inferred from indels or stops is depicted by red branches. Branch length represents millions of years and numbers at the nodes represent divergence time in millions of years (*SI Text*). Numbers along the terminal branches are the d_N/d_S ratios after removing indels and stops, calculated by PAML (40) (see also *SI Text*).

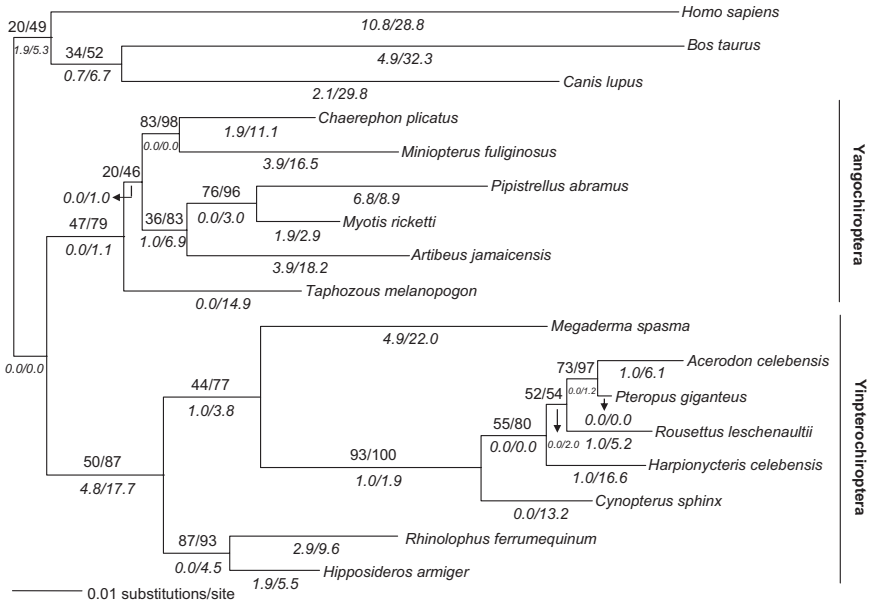

Fig. 2. The TBR maximum likelihood tree (-ln likelihood = 2,103.5) for the *M/LWS* dataset under the GTR+Γ+I model of sequence evolution. Numbers above the branches are the ML bootstrap values/Bayesian posterior probabilities as percentages, 100* = clades that received 100% ML bootstrap support and had posterior probabilities of 1; numbers below the branches are the d_N/d_S values estimated by PAML (40) (see also *SI Text*).

to influence spectral tuning (T285A and S180A) (1). The former occurred in most bats, was independent of phylogenetic relationships, and has been shown to reduce peak light absorption from 560 to 553 nm. The latter was recorded in one species only (*Myotis ricketti*) and has been shown to reduce absorption to 545 nm (1).

SWS1 opsin. Surprisingly, the *SWS1* opsin gene showed a sharply different evolutionary trajectory to the *M/LWS* opsin gene in bats. Phylogenetic analyses of the *SWS1* opsin gene differed from the recently published consensus tree (11) and instead showed an alternate topology in which bats with laryngeal echolocation formed a single monophyletic clade resembling the traditional grouping of Microchiroptera (Figs. S1 and S2). However, this could not be considered as a robust result, because statistical support for this clade was very weak, with a maximum-likelihood bootstrap value of 0.27 and a Bayesian posterior probability of 0.63 in Fig. S1 and similar values in Fig. S2. Moreover, a Shimodaira–Hasegawa test revealed no significant difference between the published species tree and our gene tree (Table S4). Branch lengths varied greatly among lineages, consistent with different selection pressures (Figs. S1 and S2, Table S2).

Among the echolocating bats studied, all 12 species examined from the major suborder of echolocating species (Yangochiroptera) were found to possess a functional *SWS1* opsin characterized by an intact ORF (Fig. 1) and conserved mammalian intronic splice sites (Fig. S3). In each case, the 11 critical amino acid sites implicated in spectral tuning matched those of the ancestral and consensus mammalian sequence, which is known to encode an UV sensitive pigment (Fig. 1). All of these bats use "low-duty-cycle" echolocation, in which the emitted calls are relatively short in duration and do not overlap in time with the returning echoes (11). Similarly, of the echolocating species studied that belong to the other major suborder (Yinpterochiroptera), the only representative to use low-duty-cycle echolocation (*Megaderma spasma*) (11) was also found to be the only 1 to possess an intact UV sensitive shortwave opsin (Fig. 1).

In all of the other surveyed members of the Yinpterochiroptera that use laryngeal echolocation (rhinolophids and hipposiderids), we found evidence that the *SWS1* opsin gene had accrued multiple insertions and deletions (indels) and premature stop mutations that disrupted the ORF and was thus nonfunctional (Figs. 1 and 3, Fig. S4). These bats have evolved a specialized form of echolocation in which calls are characterized by a constant frequency component, are relatively long in duration and are separated by shorter intervals ("high-duty-cycle echolocation").

To ascertain when the functionality of the *SWS1* opsin gene was lost in these species we developed a probabilistic model of codon evolution to allow us to incorporate stop codons in our analyses and reconstruct the ancestral amino acid sequences based on the ORF alignment (Fig. 1, *SI Text*). Amino acid reconstructions suggest that stop codons arose independently at the ancestral nodes of both the families Hipposideridae and Rhinolophidae, and that both stops and indels appear at different positions in the 2 lineages.

We also found evidence of several cases of loss-of-function of *SWS1* in the sequences of 6 of the nonecholocating Old World fruit bats, although 9 closely related fruit bat genera possessed ORFs (Fig. 1, Fig. S4). Reconstructions of amino acid sequences revealed no ancestral stop codons among these taxa, suggesting independent losses (Fig. 1). Of those genera with nonfunctional SWS1 opsins, as inferred by stop or frameshift mutations, 3 (*Rousettus*, *Dobsonia*, and *Eonycteris*) roost in caves in poor photopic conditions, and one roosts in caves (*Eidolon*) occasionally (Fig. 1).

The strong contrast in the selective constraints acting on the *SWS1* opsin gene among different lineages of bats was also demonstrated by the results of branch model tests of selection. Of those *SWS1* genes considered to be functional (i.e., without stops or indels), we found no evidence that the dN/dS ratio (ω) calculated for ancestral branches of the fruit bats, Yinpterochiroptera or Yangochiroptera was significantly greater than that estimated using a model in which the ratio was fixed across the tree (Tables S2 and S3). Probabilistic reconstructions of the 11 critical sites implicated in spectral tuning (5) revealed that these sites are unlikely to have undergone significant evolutionary change in any branch of bats since their radiation, and the opsin has thus remained UV sensitive (Fig. 1). Thus, the independent losses of SWS1 in some bat lineages are probably not a consequence of changes in peak spectral sensitivity.

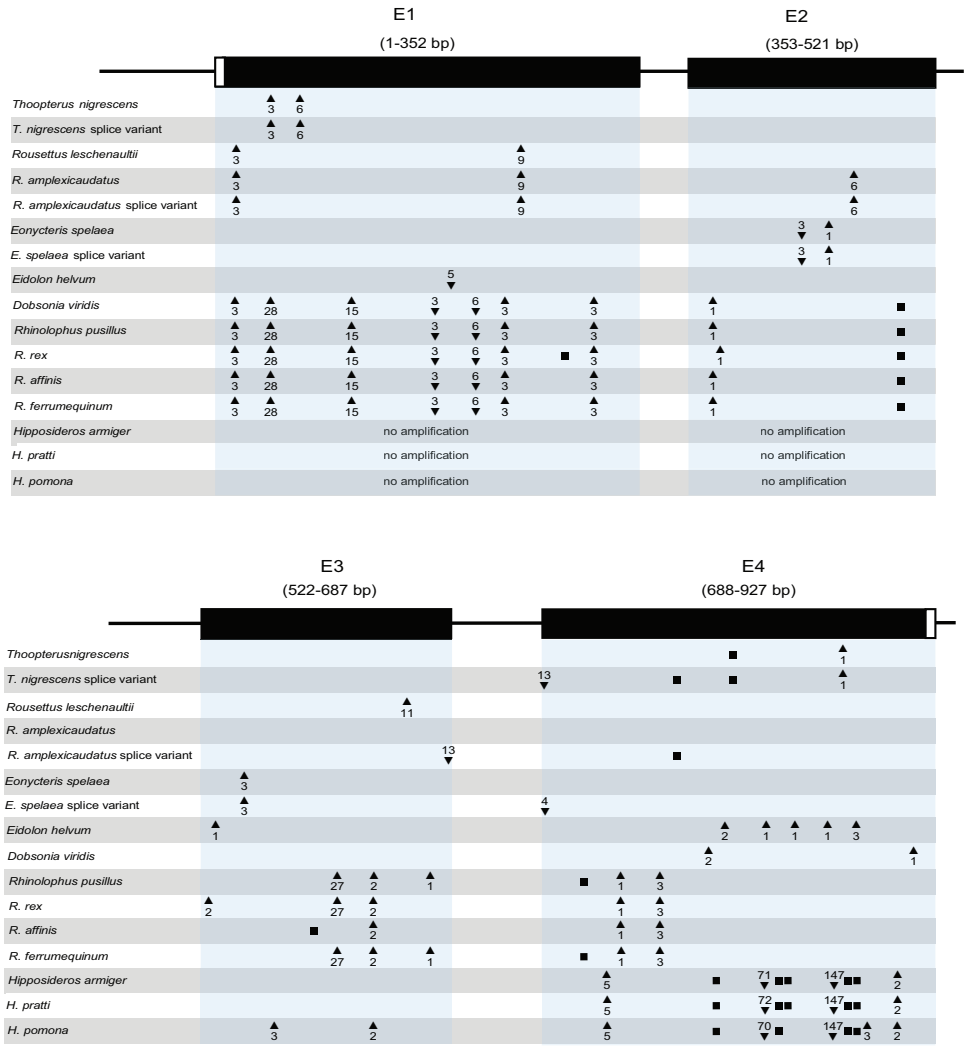

Fig. 3. Schematic to show *SWS1* bat sequences with insertions (▼), deletions (▲) and stops (■) in the ORF. Sizes of indels are given in base pairs.

Own Reading-Frame Analyses. To assess the extent of frameshift mutations resulting from indels, we removed gaps, translated each sequence in its own frame, and aligned the amino acids (Fig. S5, *SI Text*). Sequences with stop codons typically showed lower similarity to the consensus alignment than sequences with no stops (Fig. S5). Exceptions were the potential splice variant of *Eonycteris spelaea*, which had no stop yet low similarity score, and the sequence of *Thoopterus nigrescens*, which had a stop yet high similarity score (*SI Text*). The former case suggests that this potential splice form does not exist in nature, whereas the latter case supports our assertion that the loss of SWS1 in fruit bats has occurred recently, a conclusion that is also suggested by sliding window analyses (Fig. S6). Indeed, apparent elevated substitution rates were recorded mainly at or downstream of stop codons, indicating that both of these features in fact reflect shifts in frame because of indels rather than relaxed selection in fruit bats (Fig. S6). Sliding windows also revealed the impact of indels on genetic divergence, even where no stop codons are present (*Dobsonia* and *Rousettus*, Fig. S6).

Discussion

The ability of bats to orient and hunt using sound in complete darkness (echolocation) (14) represents one of the most striking examples of nocturnal sensory innovation in vertebrates. However, the extent to which bats use vision has long been debated. Early reports of all-rod-retinae suggested that bats lack color vision (3), whereas more recent visual studies of the visual capabilities in bats have yielded conflicting and inconsistent findings. The amplification of intact genes encoding the visual cone opsins (long/medium and short wavelength) in 3 species of bats suggested that bats possess dichromatic color vision (15). However, immunocytochemical examinations of cone type and density in 6 Old World fruit bats (Pteropodidae) indicated that all 6 species had long-wavelength cones, but that only 2 had short-wavelength cones (16), suggesting that some fruit bats are monochromats and so incapable of color vision. A recent behavioral study showed that one echolocating bat species can see UV visual stimuli; however, this was attributed to an adaptive change in the sensitivity of rods rather than true cone dichromatic color vision (17). Thus, the extent to which bats use vision to interpret their environment and therefore the evolutionary consequence of being a nocturnal specialist is unknown and, as these recent results (15–17) show, the answer is likely to be much more complicated than previously believed.

The study of functional genes within a phylogenetic framework can provide powerful insights into the relationship between genes, ecology, and behavior (18–21). Our phylogenetic and molecular evolutionary analyses of 2 classes of visual opsin gene reveal dramatic contrasting evolutionary histories in bats. All 14 species examined were found to possess a functional *M/LWS*

opsin gene tuned to red light, despite their long history of nocturnality and, in most cases, use of laryngeal echolocation. Indeed, a functional M/LWS opsin appears to have been retained for >80 million years, since bats diverged from other mammals (22). This supports other studies that have reported strong conservation of the M/LWS opsin in mammals (10, 15, 23). One possible explanation for such conservation in the face of nocturnality is that the M/LWS opsin might have additional roles to vision, such as in controlling the circadian rhythms of physiological processes (see ref. 23).

By comparison, the shortwave opsin (*SWS1*) shows much greater evolutionary divergence among mammals, and especially within bats. Differential conservation or losses via indels or stops appear to correspond closely to differences in species' sensory ecology. The retention of an intact UV sensitive shortwave opsin in taxa with 'low-duty-cycle' echolocation, both in all members of the Yangochiroptera, and independently in the Yinpterochiroptera taxon *Megaderma spasma* (11) suggests that these species are dependent on short wave vision for orientation and/or hunting, despite being nocturnal. At the same time, evidence of independent losses of shortwave opsin functionality early in the evolution of the Hipposideridae and Rhinolophidae lineages (extant members of which possess high-duty-echolocation) indicates that these losses are broadly coincident with the evolution a novel form of echolocation (Fig. 1). Here, bats exploit Döppler shifts to produce calls of lower frequencies than their returning echoes, allowing them to separate their calls from the returning echoes in terms of frequency rather than time and so receive a more continuous flow of acoustic information (24–26). This is considered perhaps the most advanced nocturnal sensory adaptation within mammals (27).

Such contrasting trajectories in the visual ecology of 2 main groups of echolocating bat within a nocturnal niche indicates that the neural "picture" obtained by the 2 divergent forms of echolocation is likely to differ markedly. We postulate that low-duty-cycle echolocators augment their acoustic "image" with shortwave vision, whereas the evolutionary innovation of high-duty-cycle echolocation has rendered dichromatic color vision redundant. This apparent tradeoff between vision and hearing is further supported by the recent discovery that a key cochlear protein implicated in high frequency audition has undergone a burst of adaptive selection in high-duty-cycle echolocators (21) that coincides with the loss of SWS1 function reported here. This highly specialized form of echolocation allows these bats to detect acoustic glints created by flapping insects and is considered especially well adapted for hunting in dense vegetation (narrow-space specialists) (11). Although tradeoffs between species' evolutionary adaptations have long been assumed to occur through ecological specialization, evidence for postulated evolutionary sensory tradeoffs is poor. Indeed, although most mammal species are assumed to be sensory specialists with 1 sensory modality enhanced above the others, most species also show multiple sensory modalities throughout their evolution and, where sensory losses occur, they are typically associated with a cessation of sensory input rather than via a tradeoff per se. The origin of a novel sensory modality (high-duty-cycle echolocation) within the evolutionary timeframe considered in this study means that we are able to directly relate the gain of one sense with the loss of another.

We also found evidence of the independent loss of a functional shortwave opsin in a number of fruit bat lineages, supporting the findings of emerging but taxonomically limited immunocytochemical studies (16). The retention of a functional *SWS1* gene in obligate tree roosters, yet loss in those species that roost in caves, indicates that cave roosting is associated with a relaxation in selective constraint. The close correspondence between the lack of *SWS1* functionality and roosting ecology, coupled with the lack of ancestral stops (Fig. 1), strongly suggests that cave roosting in fruit bats represents a recent behavioral innovation that has evolved independently several times (28). We also cannot rule out the possibility that the loss of SWS1 opsin in cave roosting fruit bats might also correspond to an undetected sensory tradeoff, perhaps because of an increase in olfactory capabilities, as is hypothesized for ancestral placental mammals (29), although this scenario cannot account for such close ties with roosting environment.

Reconstruction of ancestral amino acid sequences (30) and synthesis of their expressed pigments (31) suggest that the ancestral vertebrate short-wave opsin was UV (UV) sensitive (λ_{max} ≈360 nm). Derived critical site replacements in some lineages of amphibian, bird and mammal have usually been explained as adaptations to different light environments (5, 32), whereas cases of loss-of-function appear related to the cessation of sensory input (9). Conversely, the reconstruction of the ancestral intact shortwave opsin in primates has been used to infer diurnality in extinct species (10). Our finding, that the SWS1 opsin has been under purifying selection in the majority of echolocating bats and a large number of fruit bats despite a long history of nocturnality, shows that such logic is at best an over-simplification and strongly indicates that UV color vision is likely to play a considerably more important role in bat sensory ecology than previously appreciated. UV vision in mammals was until recently considered to be restricted to rodents and marsupials, with other orders showing either a loss or a switch to violet (3). The addition of most bats to this group advances our understanding of the extent to which vertebrates are able to perceive and use UV light (33, 34) and leads us to question the validity of earlier claims of color blindness in a nectarivorous bat (17) and diurnality in ancestral primates (10). More generally, the detected tradeoff reported here between vision and echolocation in bats supports the longstanding but weakly supported assumption that tradeoffs are indeed associated with ecological specializations and highlights the need to explore evolutionary hypotheses in phenotypically and phylogenetically divergent taxa.

Materials and Methods

Taxon Coverage. We amplified and sequenced ≈2.2 kb of the short wavelength opsin gene (*SWS1*) from exon 1 to exon 4 in 32 species of bats that vary in their acoustic and roosting ecology. Sequencing was based on genomic DNA and, where possible, from retinal mRNA (*SI Text*, Table S1). These data were supplemented with sequences from an additional 35 mammal species obtained from GenBank, comprising 4 marsupials, 1 proboscidean, 4 rodents, 19 primates, 1 canid, 1 suid, 1 bovid, 1 equid, and 3 other bats. In total, the *SWS1* dataset included 67 mammals. We also amplified and sequenced the *M/LWS* opsin gene in a subset of 14 bat species (Table S1) and added the sequences of 3 outgroups (1 primate, 1 bovid, and 1 canid) in our analyses. GenBank accession numbers and species names are given in Table S1.

Nucleotide Alignment, Phylogenetic and Molecular Evolution Analysis. Nucleotide sequences were aligned in the ORF using CLUSTAL X (35) and modified by eye with SE-AL (36). Intron-exon boundaries were identified with reference to published sequences, and, in the case of *SWS1*, c-DNA sequences obtained in this study. The best-fit model of nucleotide evolution was determined by Modeltest (37), and Maximum Likelihood and Bayesian phylogenetic reconstruction methods were performed with PAUP 4.0b10 (38) and MrBayes 3.1.1 (39), respectively (see also *SI Text*). We also applied a Maximum Likelihood approach to test for differences in selection pressure, using the CODEML program of PAML version 4 (40) (see also *SI Text*).

Ancestral Sequence Reconstruction. Ancestral sequences were reconstructed under an explicit phylogenetic model of coding sequence evolution in a Maximum Likelihood framework. Our model of coding sequence evolution is similar to the models of Muse and Gaut (41), extended to include substitutions to and from stop codons and so that substitution rates depend on the nucleotide composition. We modeled separate rates for changes between synonymous and nonsynonymous codons for each branch of the tree, and an additional parameter describing the rate of sense codons changing to stop

codons. Rates of change between codons also depended on the nucleotide composition at each codon position (see also *SI Text*).

Protein Sequence Alignment and Sliding Window Analysis. The indels of *SWS1* protein coding sequences among bats were removed and the sequences translated in their own reading frames using SE-AL (36) and realigned using CLUSTAL X (35) and T-Coffee (42). The sequences were then examined for stop codons, indicative of loss-of-function. We derived similarity scores of each sequence relative to the multiple sequence alignment using T-Coffee (see also *SI Text*). We also measured deviation from the functional state by estimating average rates of nonsynonymous (d_N) and synonymous (d_S) substitutions per site, and the d_N/d_S (ω) ratio, for a sliding window of 90 nt with a step size of 9 nt (Fig. S6) in the software SWAAP 1.0.2 (43).

ACKNOWLEDGMENTS. We thank A. Walsh, G. Li, and L. Wei for assistance with samples and L. Yuan, G. Li, Z. Wang, and J. Wang for technical support in the laboratory. We are grateful to A. Roca and W. Murphy for helpful comments on an earlier version of this manuscript. This work was supported by a grant under the Key Construction Program of the National ''985'' Project and ''211'' Project (S.Z.). Additionally, S.J.R. was supported by a Royal Society Research Fellowship and E.C.T. by a Science Foundation Ireland grant PIYRA (06/YI3/B932).

1. Yokoyama S, Radlwimmer FB (1999) The molecular genetics of red and green color vision in mammals. *Genetics* 153:919–932.
2. Yokoyama S, Yokoyama R (1996) Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annu Rev Ecol Syst* 27:543–567.
3. Peichl L (2005) Diversity of mammalian photoreceptor properties: adaptations to habitat and lifestyle? *Anat Rec* 287:1001–1012.
4. Yokoyama S, Radlwimmer FB (1998) The ''five-sites'' rule and the evolution of red and green color vision in mammals. *Mol Biol Evol* 15:560–567.
5. Yokoyama S, Starmer WT, Takahashi Y, Tada T (2006) Tertiary structure and spectral tuning of UV and violet pigments in vertebrates. *Gene* 365:95–103.
6. Yokoyama S (2000) Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* 19:385–419.
7. David-Gray ZK, et al. (2002) Adaptive loss of ultraviolet-sensitive/violet-sensitive (UVS/VS) cone opsin in the blind mole rat (*Spalax ehrenbergi*). *Eur J Neurosci* 16:1186–1194.
8. Levenson DH, Dizon A (2003) Genetic evidence for the ancestral loss of short-wavelength-sensitive cone pigments in mysticete and odontocete cetaceans. *Proc R Soc Lond B* 270:673–679.
9. Carvalho Ldos S, Cowing JA, Wilkie SE, Bowmaker JK, Hunt DM (2006) Shortwave visual sensitivity in tree and flying squirrels reflects changes in lifestyle. *Curr Biol* 16:R81–83.
10. Tan Y, Yoder AD, Yamashita N, Li WH (2005) Evidence from opsin genes rejects nocturnality in ancestral primates. *Proc Natl Acad Sci USA* 102:14712–14716.
11. Jones G, Teeling EC (2006) The evolution of echolocation in bats. *Trends Ecol Evol* 21:149–156.
12. Speakman J (2001) The evolution of flight and echolocation in bats: another leap in the dark. *Mammal Rev* 31:111–130.
13. Simmons NB, Seymour KL, Habersetzer J, Gunnell GF (2008) Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* 451:818–821.
14. Griffin DR ed. (1958) *Listening in the Dark* (Yale Univ Press, New Haven).
15. Wang D, et al. (2004) Molecular evolution of bat color vision genes. *Mol Biol Evol* 21:295–302.
16. Muller B, Goodman SM, Peichl L (2007) Cone photoreceptor diversity in the retinas of fruit bats (Megachiroptera). *Brain Behav Evol* 70:90–104.
17. Winter Y, Lopez J, Von Helversen O (2003) Ultraviolet vision in a bat. *Nature* 425:612–614.
18. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573.
19. Zhang J, Zhang YP, Rosenberg HF (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* 30:411–415.
20. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. *Nat Genet* 36:1326–1329.
21. Li G, et al. (2008) The hearing gene *Prestin* reunites echolocating bats. *Proc Natl Acad Sci USA* 105:13959–13964.
22. Murphy WJ, et al. (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.
23. Nei M, Zhang J, Yokoyama S (1997) Color vision of ancestral organisms of higher primates. *Mol Biol Evol* 14:611–618.
24. Trappe M, Schnitzler H-U (1982) Doppler-shift compensation in insect-catching horseshoe bats. *Naturwissenschaften* 69:193–194.
25. Schuller G, Beuter K, Schnitzler H-U (1973) Response to frequency shifted artificial echoes in the bat Rhinolophus ferrequinum. *J Comp Physiol* 89:275–286.
26. Hiryu S, Katsura K, Lin LK, Riquimaroux H, Watanabe Y (2005) Doppler-shift compensation in the Taiwanese leaf-nosed bat (*Hipposideros terasensis*) recorded with a telemetry microphone system during flight. *J Acoust Soc Am* 118:3927–3933.
27. Jones G (2005) *Echolocation Curr Biol* 15:R484–R488.
28. Giannini NP, Simmons NB (2003) A phylogeny of megachiropteran bats (Mammalia : Chiroptera : Pteropodidae) based on direct optimization analysis of one nuclear and four mitochondrial genes. *Cladistics* 19:496–511.
29. Kishida T (2008) Pattern of the divergence of olfactory receptor genes during tetrapod evolution. *PLoS ONE* 3:e2385.
30. Shi Y, Yokoyama S (2003) Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proc Natl Acad Sci USA* 100:8308–8313.
31. Shi Y, Radlwimmer FB, Yokoyama S (2001) Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc Natl Acad Sci USA* 98:11731–11736.
32. Fasick JI, Applebury ML, Oprian DD (2002) Spectral tuning in the mammalian short-wavelength sensitive cone pigments. *Biochemistry* 41:6860–6865.
33. Bennett ATD, Cuthill IC, Partridge JC, Maier EJ (1996) Ultraviolet vision and mate choice in zebra finches. *Nature* 380:433–435.
34. Fleishman LJ, Loew ER, Leal M (1993) Ultraviolet vision in lizards. *Nature* 365:397.
35. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
36. Rambaut A (1996) Se-Al. http://evolve.zoo.ox.ac.uk/Se-Al/Se-Al.html.
37. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
38. Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Sinauer Associates, Sunderland, Massachusetts).
39. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
40. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
41. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:15–724.
42. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217.
43. Pride DT (2004) SWAAP 1.0.2: a tool for analyzing substitutions and similarity in multiple alignments.

Supporting Information

Zhao et al. 10.1073/pnas.0813201106

SI Text

Gene and Taxon Coverage. We amplified and sequenced ≈2.2 kb of the short wavelength opsin gene (*SWS1*) from exon 1 to exon 4 in 32 species of bats (Table S1). We also obtained from GenBank the published sequences of 3 additional bat species (Table S1). Our dataset of bats contained 15 pteropodids, 1 megadermatid, 3 hipposiderids, 4 rhinolophids, 3 emballonurids, 3 phyllostomids, 1 miniopterid, 4 vespertilionids, and 1 molossid, *sensu Jones and Teeling* (1) (Table S1). Sequences from a number of outgroups were also obtained from GenBank and consisted of 4 marsupials, 1 proboscidean, 4 rodents, 19 primates, 1 canid, 1 suid, 1 bovid, and 1 equid (Table S1).

We also amplified and sequenced ≈3.2 kb of the *M/LWS* opsin gene, spanning exon 3 to exon 5, in a subset of 14 bat species (Table S1). These consisted of 5 pteropodids, 1 rhinolophid, 1 hipposiderid, 1 megadermatid, 1 emballonurid, 1 phyllostomid, 2 vespertilionids, 1 miniopterid, and 1 molossid. Outgroup sequences from GenBank consisted of 1 primate, 1 bovid, and 1 canid (Table S1).

Isolation, Amplification, and Sequencing of Genomic DNA. Genomic DNA was isolated using Qiagen DNeasy kits following the manufacturer's protocol. For the *SWS1* opsin and the *M/LWS* opsin genes, we designed degenerate primer pairs in conserved positions (Table S5) based on an alignment of published mammalian sequences of human, rhesus monkey, mouse, rat, gray squirrel, guinea pig, dog, and cow. PCR (10 μl) were carried out using high fidelity Taq (TransGen Biotech) on a DNA Engine Dyad Cycler (BioRad). Cycling reactions were as follows: 94°C for 30 s, annealing temperature (Table S5) for 40 s, 72°C for 1.5 min for 35 cycles, and 72°C for 10 min. The PCR mixtures contained 1 μl (50 ng/μl) of genomic DNA, 5 μl of 10 × buffer, 1.5 μl (50 mM) of $MgCl_2$, 1 μl (10 μM) of each primer, and 1 unit TaqDNA polymerase. PCR products were verified by agarose gel, ligated into a pEASY-Blunt vector (TransGen Biotech) and then transformed into TOP10 *Escherichia coli* competent cells (Tiangen). Positive clones were sequenced with the primer pair M13–47 (5′-CGC CAG GGT TTT CCC AGT CAC GAC-3′) and M13–48 (5′-GAG CGG ATA ACA ATT TCA CAC AGG-3′) using the Big Dye Terminator Cycle Sequencing kit (Applied Biosystems) and run on an ABI 3730 sequencer. Multiple clones (3–5) were sequenced for each PCR product to substantiate the results.

Isolation, Amplification, and Sequencing of Retinal mRNA. To gain information on retinal expression and, where appropriate, verify coding regions or absence of coding regions, we undertook tissue-specific *SWS* opsin RT-PCR in representative species from both major subordinal bat clades. For the Yinpterochiroptera, we collected 1 pteropodid (*Rousettus leschenaultii*) from Sichuan in China, 1 megadermatid (*Megaderma spasma*) from Lang Son in Vietnam, 2 rhinolophids (*Rhinolophus pusillus* and *Rhinolophus ferrumequinum*), and 2 hipposiderids (*Hipposideros pratti* and *Hipposideros armiger*) from Guangxi in China. For the Yangochiropteran, we collected 1 emballonurid (*Taphozous melanopogon*), 1 molossid (*Chaerephon plicatus*), and 1 vespertilionid (*Myotis ricketti*), all from Guangxi, China.

Individuals were humanely killed, and eyes were excised rapidly, placed in liquid nitrogen, and stored at −80°C. Total RNA was isolated from the eye using the RNAiso kit (Takara). The first-strand cDNA was synthesized using 2 μg of total RNA treated with DNase I, 2.5 μM oligo(dT), 500 μM dNTPs, 40 units pf RNase inhibitor (Invitrogen), 1 × superScript RT buffer, and 200 units of superScript reverse transcriptase (Invitrogen) in a volume of 20 μl at 50°C for 1 h. The reaction was terminated and then inactivated at 70°C for 15 min.

We used the same primer sets as listed in Table S5 to amplify the *SWS1* gene from cDNAs. PCRs were performed in 10-μl reaction volumes that included 1 μg of the first-strand cDNA, 0.2 μM primers, and 1 unit of TaqDNA polymerase (Takara). After an initial step of 5 min at 95°C, PCR was carried out for 35 cycles at 94°C for 30 s, annealing temperature for 30 s and 72°C for 1 min with a final extension of 72°C for 5 min. Products were purified using the Takara agarose gel DNA purification kit and cloned into the pMD19-T vector (Takara) and sequenced commercially. Once again, to verify results and avoid artifacts, multiple clones of each PCR product were sequenced.

Of the 9 species tested, the cDNA of four species (*Megaderma spasma*, *Taphozous melanopogon*, *Chaerephon plicatus*, and *Myotis ricketti*) was sequenced successfully and showed no differences with the aligned corresponding coding sequence based on genomic DNA. However, in the 5 remaining species (*Rousettus leschenaultii*, *Rhinolophus pusillus*, *R. ferrumequinum*, *Hipposideros pratti*, and *H. armiger*), we were unable to amplify any *SWS1* cDNA. Given that these 5 species all possess stops or indels in their SWS1 opsins, as inferred from genomic DNA sequences, the cDNA results verify our assertion that the *SWS1* gene is nonfunctional in these bats.

Nucleotide Sequence Alignment and Species Phylogenetic Reconstruction. Nucleotide sequences were aligned using CLUSTAL X (2) and modified by eye with SE-AL (3). Intron-exon boundaries were identified with reference to published sequences, and, in the case of *SWS1*, cDNA sequences obtained in this study. In the *M/LWS* dataset, intronic-splice signals (GT/AG rule) were completely conserved across all species, and no alternative splice signals were found. In the *SWS1* dataset, splice sites were conserved in most mammals. However, alternative splice signals (GT/AG rule) (Fig. S3) were also found within 15 bp of the conserved sites in 3 species (*Rousettus amplexicaudatus, Eonycteris spelaea*, and *Thoopterus nigrescens*) and used to generate potential variant splice forms, which were included in phylogenetic analyses. Introns varied in length and were removed from all sequences where possible, and exons were concatenated in the ORF. Note that in some species of hipposiderid bats, the *SWS1* genomic sequences were so divergent that no splice signals were found for certain introns. The resulting alignments of *SWS1* spanned 1,107 bp, including gaps, and that of *M/LWS* spanned 536 bp with no gaps.

We undertook phylogenetic reconstruction based on both genes. For the *SWS1* opsin, separate analyses were undertaken for all sequences and for a reduced dataset comprising only sequences from members of the superordinal clade Boreoeutheria (Laurasiatheria and Euarchontoglires). In each case, potential alternative splice forms were included. Maximum likelihood (ML) analyses were performed with PAUP 4.0b10 (4) for all datasets using the GTR (general time reversible) model of sequence evolution with the parameter settings estimated by Modeltest (5), as follows: (*i*) *M/LWS*: R-Matrix = (1.0000 9.8884 1.0000 1.0000 6.8334); base frequencies = (0.1856 0.3079 0.2607); proportion of invariant sites = 0.5562; and shape parameter of gamma distribution 0.7259; (*ii*) *SWS1* (all mammals): R-Matrix = (1.2215 4.9648 0.6497 0.9324 4.9648); base frequencies = (0.1927 0.2921 0.2564); proportion of invariant

sites = 0.0927; and shape parameter of gamma distribution = 1.0129; (*iii*) *SWS1* (Boreoeutheria): R-Matrix = (1.2544 5.0643 0.6379 1.0154 5.0643); base frequencies = (0.1901 0.2883 0.2659); proportion of invariant sites = 0; and shape parameter of gamma distribution = 0.8052. For each dataset, we generated a single ML tree using TBR branch swapping. One hundred ML bootstrap replicate trees were generated for each dataset using NNI branch swapping for the *SWS1* dataset and TBR branch swapping for the *M/LWS* dataset. In all ML analyses, starting trees were obtained via neighbor-joining (NJ), and heuristic searches were used.

We also undertook phylogenetic reconstruction for all datasets using a Bayesian approach, implemented in MrBayes 3.1.1 (6, 7). MrBayes 3.1.1 simultaneously initiates 2 Metropolis coupled Markov Chain Monte Carlo (MCMCMC) runs. When the standard deviation of split frequencies was less than 0.01, convergence was reached. Four simultaneous chains were run, 3 hot and 1 cold. Analyses were run for as many generations as was required for the average standard deviation of split frequencies to be less than 0.01. Chains were sampled every 1,000 generations. Starting trees were random and the prior indicated that all trees were equally probable. Results are shown in Fig. 2 and Figs. S1–S2.

Functionality and Protein Sequence Alignment. The alignment of bat *SWS1* sequences and potential splice variants revealed numerous indels, some of which could potentially result in frameshift mutations and stop codons. By comparison, *M/LWS* sequences contained no indels or stops. To compare *SWS1* protein coding sequences among bats, indels were removed and the sequences translated in their own reading frames using SE-AL (3), and realigned using CLUSTAL X (2) and T-Coffee (8). The sequences were then examined for stop codons, indicative of loss-of-function. We derived similarity scores of each sequence relative to the multiple sequence alignment using T-Coffee (8), and found that scores ranged from 12 to 79 (mean). The lowest scores (< 70) corresponded to 13 sequences characterized by numerous stop codons and/or frameshifts, which included all *Rhinolophus* and *Hipposideros* species and some but not all of the fruit bat species (*Rhinolophus rex*, *R. affinis*, *R. pusillus*, *R. ferrumequinum*, *Hipposideros pratti*, *H. armiger*, *H. pomona*, *Eidolon helvum*, *Eonycteris spelaea*, *Thoopterus nigrescens*, *Dobsonia viridis*, *Rousettus leschenaultii*, and a splice variant of *R. amplexicaudatus*). A plot of similarity scores versus score rank showed a stepwise pattern, indicative of major jumps in sequence dissimilarity (Fig. S5*a*). To assess sequence similarity among the pteropodids (fruit bats) in more detail (see below), we removed *Rhinolophus* and *Hipposideros* species, realigned the sequences, and repeated the entire alignment process (Fig. S5*b*).

Both plots revealed that points fall into 3 broad groups, the first of which comprised all Yangochiroptera, the majority of fruit bats without stop codons, and (the least conserved of this group) one fruit bat sequence with a stop (*Thoopterus nigrescens*). The second group comprised three fruit bat sequences without stops (*Rousettus leschenaultii* and potential splice variants of *Thoopterus nigrescens* and *Rousettus amplexicaudatus*) and 1 fruit bat with a stop (*Eonycteris spelaea* potential splice variant). The third group contained all *Rhinolophus* and *Hipposideros* species (shown in plot *a* only), and 2 fruit bats (*Eidolon helvum* and *Eonycteris spelaea*), all of which possess stop codons.

Molecular Evolution of Opsin Genes in Bats. To compare the distribution of mutational change along the *SWS1* gene in fruit bats, especially with respect to the impact of stop codons, we applied a sliding window approach. Each fruit bat nucleotide sequence was aligned in turn with the functional sequence of *Pteropus giganteus* using CLUSTAL X (2). We used the software SWAAP 1.0.2 (9) to estimate average rates of nonsynonymous (d_N) and synonymous (d_S) substitutions per site, and the d_N/d_S (ω) ratio, for a sliding window of 90 nt with a step size of 9 nucleotides (Fig. S6).

We also used a maximum likelihood approach to test for positive selection acting on both opsin genes. For this, *SWS1* genes considered to be nonfunctional (with stop codons) were removed from the dataset. All tests on *SWS1* and *M/LWS* opsin dataset were conducted using the CODEML program of PAML version 4 (10). For each model, the ratio of nonsynonymous (d_N) to synonymous (d_S) substitution rates was estimated (Table S2). Likelihood ratio tests (LRTs) were performed to compare pairs of nested models (10) (Table S3). Twice the difference in log-likelihood values ($2\Delta l$) between the 2 models was calculated with a χ^2 distribution, and the degree of freedom equals the difference the numbers of parameters estimated in the nested models. Maximum likelihood methods were applied to identify putative heterogeneous selective pressures along different branches.

For both genes, we undertook the same suite of tests. First, we undertook a free-ratio model, which assumes an independent ω value for each branch. Second, we undertook branch models assuming 2 ratios, in which specific branches were constrained as the foreground. Foreground branches were specified as (*i*) fruit bats, (*ii*) all Yinpterochiroptera, and (*iii*) Yangochiroptera for the *M/LWS* dataset, whereas foregrounds in *SWS1* dataset were Yinpterochiroptera and Yangochiroptera lineages. The null hypotheses of the free-ratio and 2-ratio models are both the 1-ratio model (M0), which assumes the same ω value for all branches. Third, 4 site-specific models were analyzed to examine the possible heterogeneous selective forces among amino acid sites. The neutral model (M1a) estimates 2 ω values ($0<\omega_0<1$, $\omega_1 = 1$); the positive selection model (M2a) adds an additional ω value to M1a, which allows ω_2 greater than 1, if present; M7 (β model) constrains ω ratio smaller than 1 following the β distribution; M8 ($\beta\&\omega$ model) is an extension of M7 and takes into account possible positively selected (PS) sites. Bayes Empirical Bayes (BEB) analysis was used to calculate the Bayesian posterior probability of positively selected sites (11). LRT statistics was calculated between 2 nested models (i.e., M1 vs. M2, M7 vs. M8). The same methods applied in *SWS1* dataset were also used in *M/LWS* dataset.

Consensus Species Tree and Timeline. Overwhelming evidence has led to the rejection of the old bat suborders of Microchiroptera (echolocating bats) and Megachiroptera (nonecholocating bats) in favor of 2 new clades: the Yinpterochiroptera, which contains all nonlaryngeal echolocating Old World fruit bats and 5 families of echolocating insectivorous bats; and Yangochiroptera, which contains all other families of echolocating bats (1, 12, 13) (Fig. 1). Molecular data in the form of large nuclear and mitochondrial concatenations provide strong support a basal division between Yinpterochiroptera (rhinolophoid microbats and pteropodids) and Yangochiroptera (all other bats) and for the association of 4 major groups of echolocating microbat lineages (Fig. 1; refs. 1, 12, 13 and refs. therein): (*i*) Rhinolophoidea (rhinolophids, hipposiderids, rhinopomatids, craseonycterids, and megadermatids); (*ii*) Emballonuroidea (nycterids and emballonurids); (*iii*)Vespertilionoidea (vespertilionids, molossids, natalids, and miniopterids); (*iv*) Noctilionoidea (noctilionids, phyllostomids, mormoopids furipterids, thyropterids, mystacinids, and myzopopdids). Bat family divergence dates (Fig. 1) were taken from the studies detailed in refs. 1, 12–13., Phylogenetic relationships within the Old World fruit bats (Pteropodidae) are based on a molecular analysis of 4 mitochondrial loci and 1 nuclear gene (14) (Fig. 1). This topology implies that laryngeal echolocation either had a single origin in the ancestor of bats but was lost in the lineages leading to the Old World fruit bats, or that laryngeal echolocation was gained multiple times in bats (1,

12). Paleontological evidence has not resolved these scenarios: the 2 oldest fossil bats both appear to have been insectivorous and capable of flight, yet only 1 had laryngeal echolocation capabilities, and it is unclear whether they were nocturnal (15, 16). Such conflicts mean that the sensory ecology of ancestral bats still remains poorly understood (1).

Gene Tree vs. Species Tree. We used the Kishino–Hasegawa and Shimodaria–Hasgawa tests as implemented in PAUP4.0b10 (4, 17) with RELL optimization and 1,000 bootstrap replicates to identify whether the gene tree topology (Fig. 1) was significantly better supported by the *SWS1* dataset than the published bat species tree (1, 12, 13). This test was carried out on a subset of taxa that included only bat species, with dog and cow as outgroups.

Ancestral Sequence Reconstruction. Ancestral sequences were reconstructed under an explicit phylogenetic model of coding sequence evolution in a maximum-likelihood framework. Our model of coding sequence evolution is similar to the models of Muse and Gaut (18), extended to include substitutions to and from stop codons and so that substitution rates depend on the nucleotide composition. We specify the substitution rate for codons $i = i_1i_2i_3$ and $j = j_1j_2j_3$, which differ at a single position k, as:

$$q_{i,j} = \begin{cases} 0, & \text{if the codons differ at more than one position} \\ \rho_{i_k,j_k}\pi^{(k)}_{j_k}, & \text{for synonymous substitutions} \\ \omega\rho_{i_k,j_k}\pi^{(k)}_{j_k}, & \text{for non-synonymous substitutions} \\ \sigma\omega\rho_{i_k,j_k}\pi^{(k)}_{j_k}, & \text{for substitutions to and from stop codons} \end{cases}$$

where $\pi^{(k)}_n$ is the observed frequency of nucleotide n at codon position k in our data, $\rho_{m,n}$ is the rate of substitutions between nucleotide m and n for $m \neq n$, which was modeled as being independent of codon position. To correctly model the dependence of substitution rate on the nucleotide sequence, we performed a hierarchical series of likelihood ratio tests to select the best-fitting reversible model of substitution (so $\rho_{m,n} = \rho_{n,m}$ for all m,n for all models tested), including testing different models of rate heterogeneity across codons and across branches. The results of these tests are summarized in Table S6. The best-fitting model, used for reconstruction of ancestral sequences, has (nucleotide constraints for ρ), included a gamma distribution of rates across sites (19), and a separate omega value for each branch of the tree. For all analyses, a fixed tree topology was used as shown in Fig. 1. All model fitting, likelihood-ratio tests and ancestral state reconstruction were performed using the Hyphy software package (20). In Hyphy, we constructed ancestral sequences as 1,000 samples for each internal node, where codons were sampled in proportion to their likelihood at the estimated maximum-likelihood model parameter values, using the *sampleancestors* command of Hyphy. The proportion of codons in these reconstructed samples is an estimate of the empirical Bayes posterior probability with a prior that places equal probability on each codon for every site.

Like any model, this model approximates only the evolution of our sequences. In particular, our alignment includes both coding and noncoding sequences, so the codon structure imposed by this model is not relevant for pseudogenes or parts of the sequence downstream of any stop codons. In the best-fitting model, we capture some of this complexity by allowing independent ω values for each branch of the tree: we expect omega to be low for branches on which genes are functional, but close to 1 for these parts of the tree on which the gene has become psuedogenized. Our model thus becomes similar to a standard nucleotide model for these branches. Our results show that most of the stop codons have evolved close to the tips of the tree, so the gene is functional (and our codon model should fit well) over most of the tree. We are not aware of any phylogenetic work that properly models sequences that switch between coding and non-coding across a tree, which would require a significant advance in the state-of-the-art of in models of sequence evolution.

1. Jones G, Teeling EC (2006) The evolution of echolocation in bats. *Trends Ecol Evol* 21:149–156.
2. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
3. Rambaut A (1996) Se-Al. http://evolve.zoo.ox.ac.uk/Se-Al/Se-Al.html.
4. Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 (Sinauer Associates, Sunderland, MA).
5. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
6. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
7. Huelsenbeck JP, Ronquist F, eds (2001) *MrBayes 2.01, Bayesian Inference of Phylogeny* (University of Rochester, Rochester, NY).
8. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217.
9. Pride DT (2004) SWAAP 1.0.2: a tool for analyzing substitutions and similarity in multiple alignments [http://www.bacteriamuseum.org/SWAAP/SwaapPage.htm].
10. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
11. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.
12. Teeling EC, et al. (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307:580–584.
13. Miller-Butterworth CM, et al. (2007) A family matter: conclusive resolution of the taxonomic position of the long-fingered bats, Miniopterus. *Mol Biol Evol* 24:1553–1561.
14. Giannini NP, Simmons NB (2003) A phylogeny of megachiropteran bats (Mammalia : Chiroptera : Pteropodidae) based on direct optimization analysis of one nuclear and four mitochondrial genes. *Cladistics* 19:496–511.
15. Speakman J (2001) The evolution of flight and echolocation in bats: another leap in the dark. *Mammal Rev* 31:111–130.
16. Simmons NB, Seymour KL, Habersetzer J, Gunnell GF (2008) Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* 451:818–821.
17. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 29:170–179.
18. Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:15–724.
19. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* 39:306–314.
20. Pond SL, Frost SD, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
21. Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W (2007) Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* 17:413–421.
22. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190.

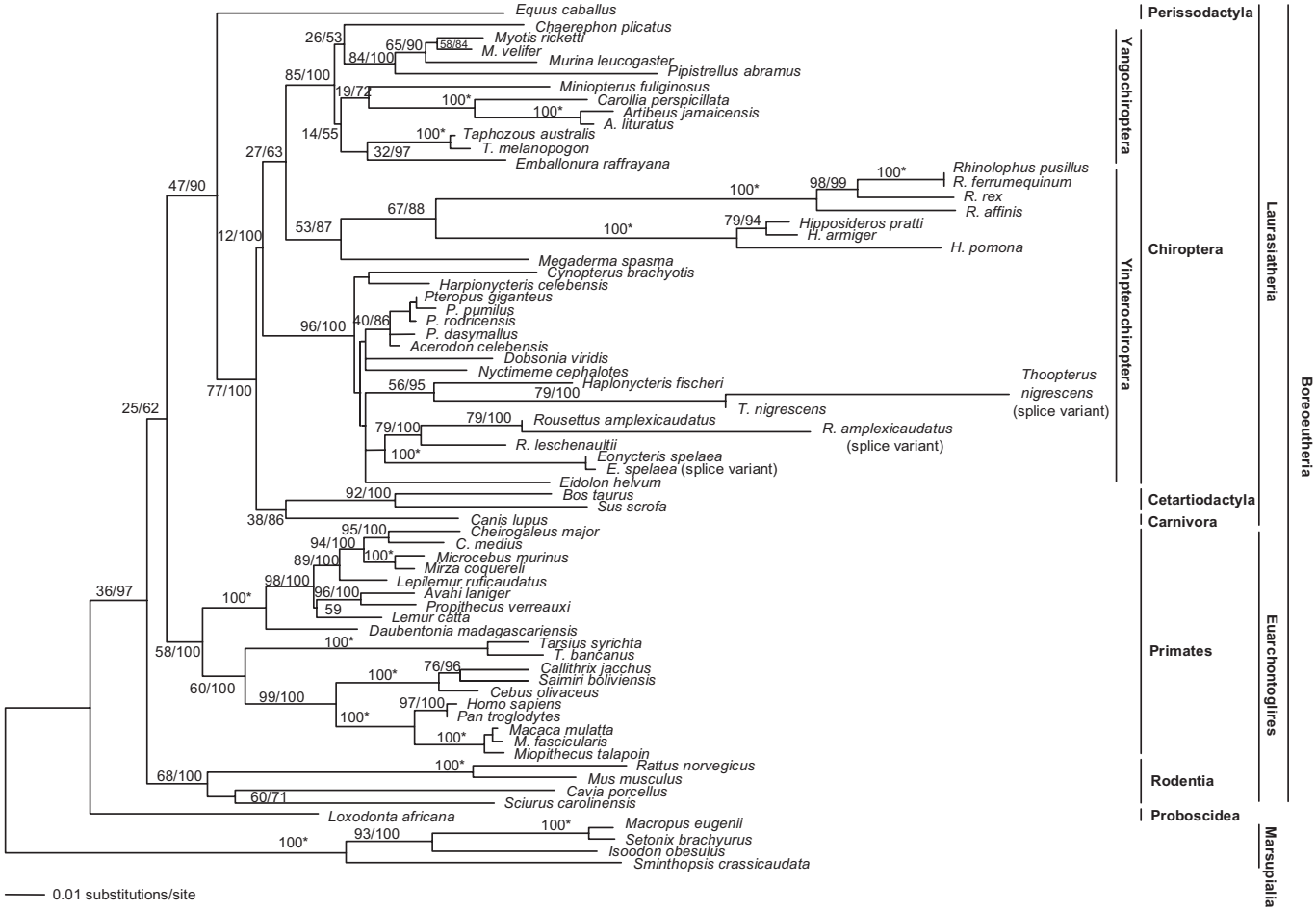

Fig. S1. *SWS1* phylogenetic tree of aligned nucleotide sequences in the ORF. The TBR maximum likelihood tree (-ln likelihood = -ln = 11574.95318) for the *SWS1* dataset under the GTR+Γ+I model of sequence evolution. Numbers at the nodes are the ML bootstrap values/Bayesian posterior probabilities as percentages, 100* = clades that received 100% ML bootstrap support and had posterior probabilities of 1. Shimodaira–Hasegawa tests (Table S4) indicated there was no significant difference in support for the published species tree (Fig. 1) and the gene tree depicted below. Our Bayesian and ML phylogenetic analyses of all bats based on the *SWS1* gene yielded little congruent phylogenetic signal compared to the *M/LWS* opsin gene and other published molecular datasets for both bats and other orders of mammals (Fig. 2 and Figs. S1-S2). This might be expected from highly contrasting selection pressures acting on this gene within bats and across the tree. Indeed a relaxation in selection pressure effected by the loss of gene function in several lineages is well reflected by the long-branch lengths and elevated rates of substitution.

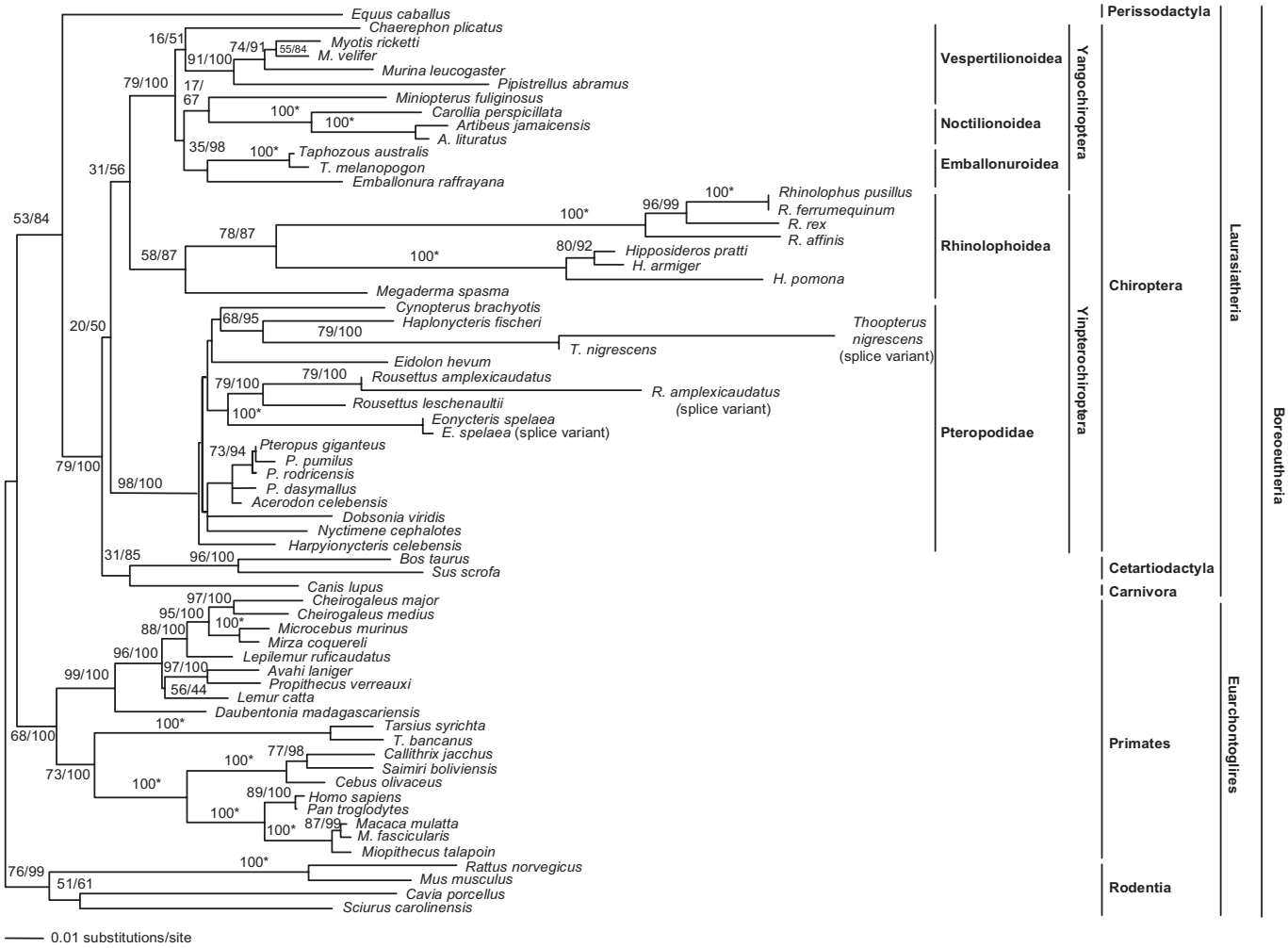

Fig. S2. The TBR maximum likelihood tree (-ln likelihood = 10421.2) for the *SWS1* opsin dataset under the GTR+Γ+I model of sequence evolution for boreoeutherian taxa. Numbers at the nodes are the ML bootstrap values/Bayesian posterior probabilities as percentages, 100* = clades that received 100% ML bootstrap support and had posterior probabilities of 1.

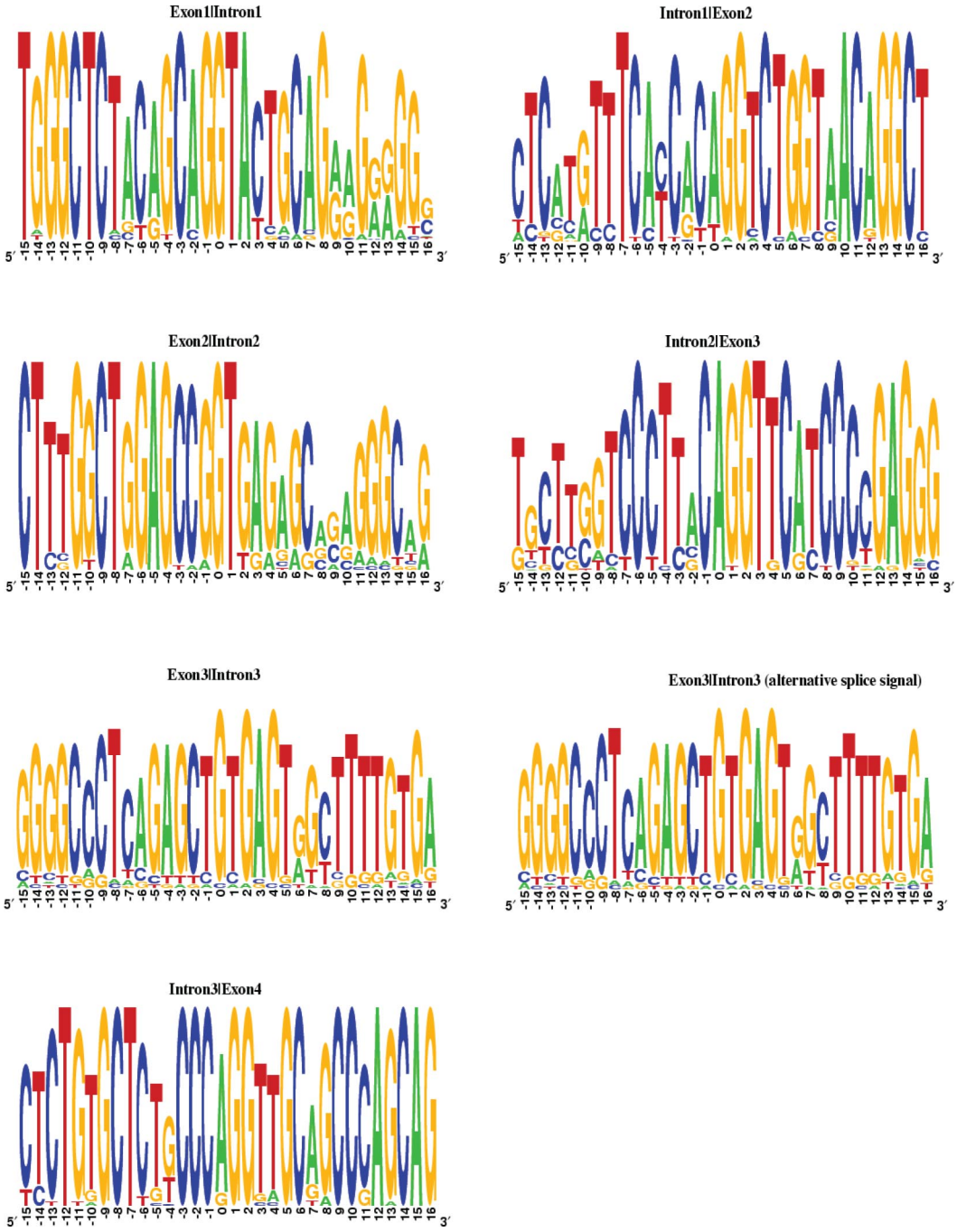

Fig. S3. Logo plots showing nucleotide variation at intron/exon boundaries among the *SWS1* sequences, constructed in the software WebLogo (22). The total height of the stack of nucleotides reflects the sequence conservation at that position, while the height of each nucleotide within the stack reflects its relative frequency.

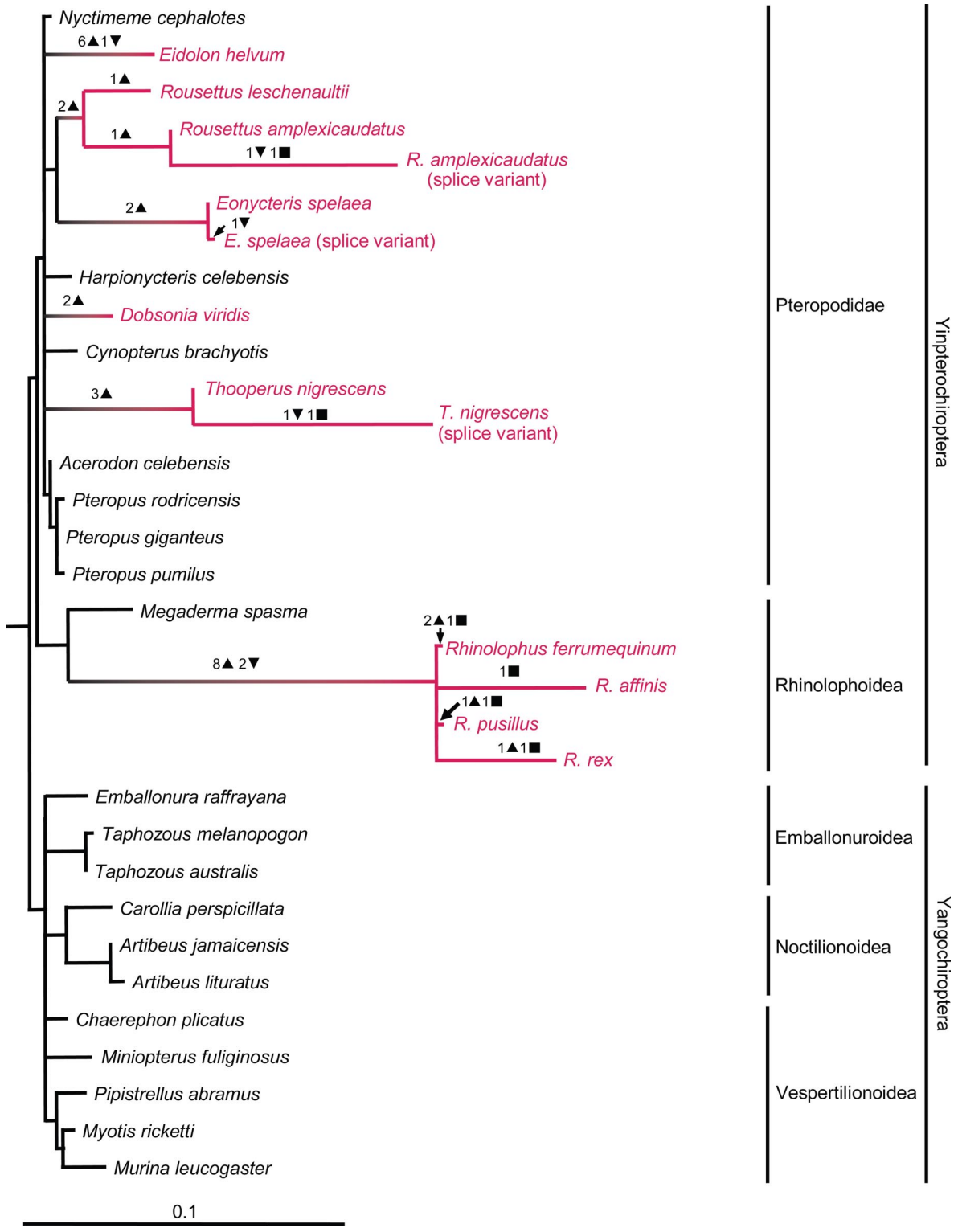

Fig. S4. *SWS1* species tree of aligned nucleotide sequences in the ORF. Branch lengths are scaled by the rate of nonsynonymous substitution and the tree topology follows published consensus studies. Species with sequences containing stop codons are shown in red font, while the loss of functionality inferred from indels or stops is depicted by red branches. Inferred ancestral stops are depicted in Fig. 1. In cases where it has been retained, *SWS1* appears to be under purifying selection and was even found to be significantly lower in the ancestral fruit bat branch than the fixed ratio ($\omega = 0.000$ vs. 0.145, respectively, $P < 0.001$). On the other hand, site models did identify 2 residues that appeared to be under positive selection (33V, 78H), although neither corresponded to sites of known spectral importance.

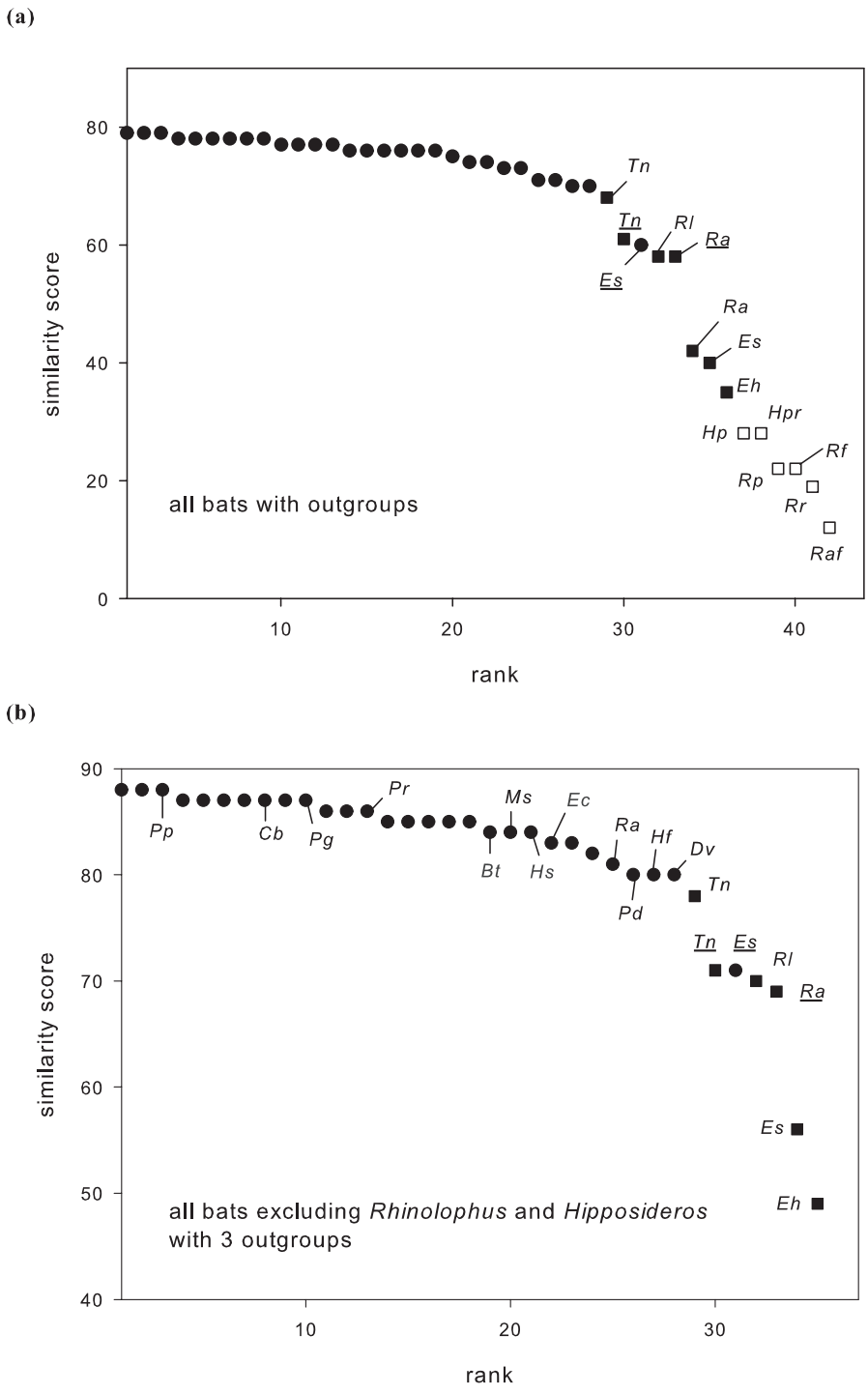

Fig. S5. Plots of T-Coffee similarity scores vs. score rank for *SWS1* amino acid sequences of bats and 3 outgroups (human, *Homo sapiens*; horse, *Equus caballus*; and cow, *Bos taurus*). Sequences have been aligned following translation in their own frames. Plot (*a*) shows ranked scores for all bat sequences and outgroups and (*b*) for the same sequences but without *Rhinolophus* and *Hipposideros*. Circles denote sequences without stop codons and squares sequences with stop codons. In *a*, unfilled squares correspond to *Rhinolophus* and *Hipposideros* species. An underlined sequence denotes a potential splice variant, and blue font denotes an outgroup. Individual sequences of interest are identified using the first letters of the genus and species (listed in Table S1).

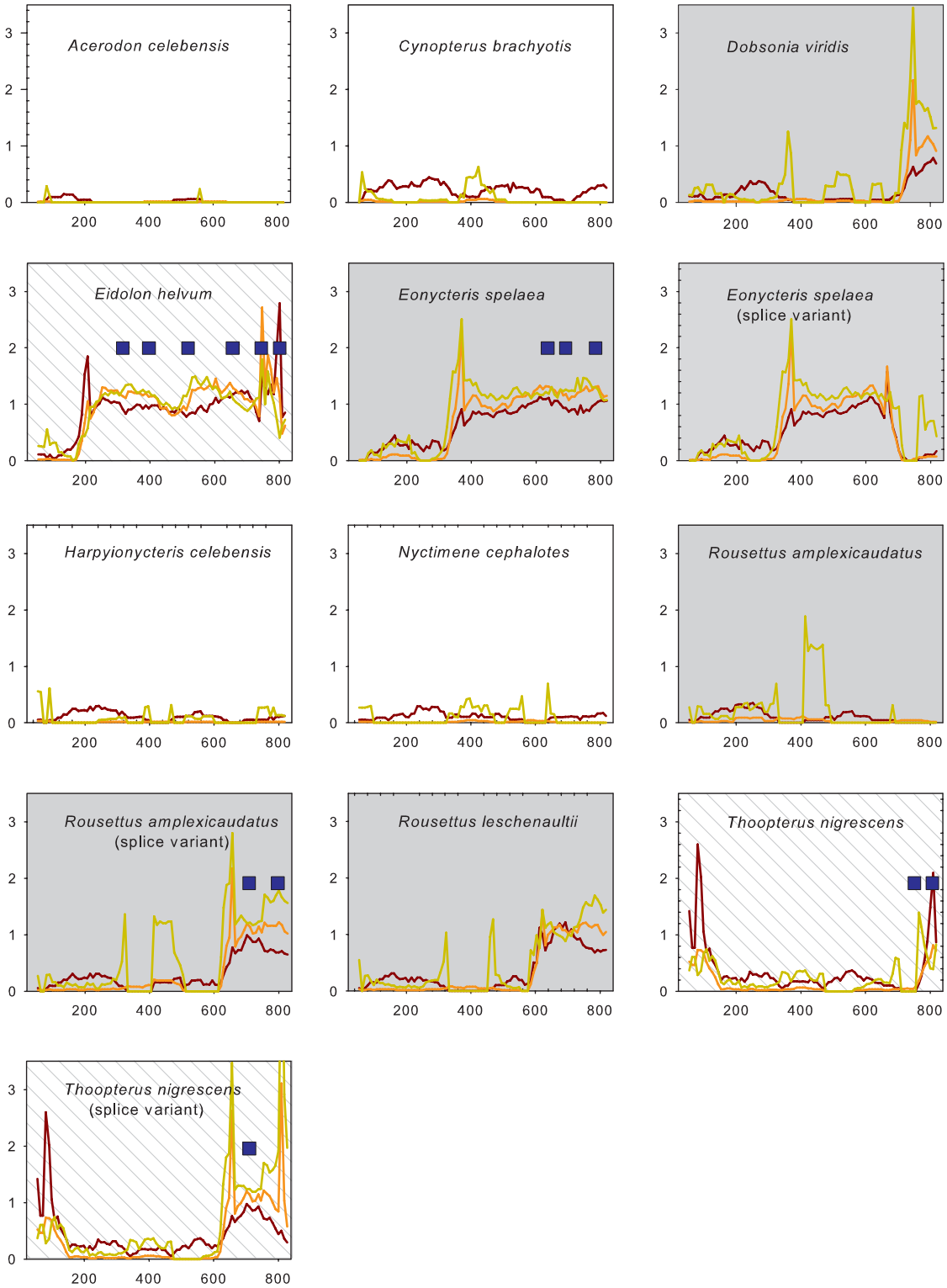

Fig. S6. Sliding window analysis of evolutionary change along the *SWS1* gene in Old World fruit bats (Pteropodidae). Each plot shows the synonymous substitution rate (d_S = green), non-synonymous substitution rate (d_N = orange) and omega (d_N/d_S = brown). Sequences have been aligned in their own frames, and the positions of stops are shown as dark blue squares. Gray shading indicates species that always roost in caves, white indicates always in trees, and hatched shading indicates species that either roost in both caves and trees (*Eidolon*) or for which roosting ecology is not known (*Thoopterus*). Species are ordered alphabetically.

Table S1. Taxa studied with GenBank accession nos.

Classification			Common name	Species name	*SWS1* gene	*M/LWS* gene
Bo	La	ORDER CHIROPTERA (BATS)				
		SUBORDER YINPTEROCHIROPTERA				
		Family Pteropodidae	African straw-colored fruit bat	*Eidolon helvum*	**EU912373**	
			lesser dawn bat	*Eonycteris spelaea*	**EU912375**	
			Geoffroy's rousette	*Rousettus amplexicaudatus*	**EU912352**	
			Leschenault's rousette	*Rousettus leschenaultii*	**EU912381**	**EU912350**
			swift fruit bat	*Thoopterus nigrescens*	**EU912355**	
			Philippine pygmy fruit bat	*Haplonycteris fischeri*	AM263196	
			Pallas's tube-nosed fruit bat	*Nyctimene cephalotes*	**EU912357**	
			greenish naked-backed fruit bat	*Dobsonia viridis*	**EU912371**	
			Sulawesi fruit bat	*Acerodon celebensis*	**EU912354**	**EU912338**
			Ryukyu flying fox	*Pteropus dasymallus*	AM263192	
			Rodrigues flying fox	*Pteropus rodricensis*	**EU912363**	
			little golden-mantled flying fox	*Pteropus pumilus*	**EU912362**	
			Indian flying fox	*Pteropus giganteus*	**EU912361**	**EU912348**
			Sulawesi harpy fruit bat	*Harpyionycteris celebensis*	**EU912356**	**EU912339**
			lesser short-nosed fruit bat	*Cynopterus brachyotis*	**EU912353**	
			greater short-nosed fruit bat	*Cynopterus sphinx*		**EU912342**
		Family Megadermatidae	lesser false vampire bat	*Megaderma spasma*	**EU912378**	**EU912346**
		Family Rhinolophidae	king horseshoe bat	*Rhinolophus rex*	**EU912365**	
			intermediate horseshoe bat	*Rhinolophus affinis*	**EU912364**	
			least horseshoe bat	*Rhinolophus pusillus*	**EU912382**	
			greater horseshoe bat	*Rhinolophus ferrumequinum*	**EU912380**	**EU912349**
		Family Hipposideridae	Pratt's leaf-nosed bat	*Hipposideros pratti*	**EU912369**	
			great leaf-nosed bat	*Hipposideros armiger*	**EU912368**	**EU912343**
			pomona leaf-nosed bat	*Hipposideros pomona*	**EU912369**	
		SUBORDER YANGOCHIROPTERA				
		Family Emballonuridae	Raffray's sheath-tailed bat	*Emballonura raffrayana*	**EU912374**	
			coastal tomb bat	*Taphozous australis*	**EU912383**	
			black-bearded tomb bat	*Taphozous melanopogon*	**EU912366**	**EU912351**
		Family Phyllostomidae	Jamaican fruit-eating bat	*Artibeus jamaicensis*	**EU912367**	**EU912340**
			great fruit-eating bat	*Artibeus lituratus*	**EU912372**	
			Seba's short-tailed bat	*Carollia perspicillata*	**EU912358**	
		Family Miniopteridae	western long-fingered bat	*Miniopterus fuliginosus*	**EU912376**	**EU912344**
		Family Molossidae	wrinkle-lipped free-tailed bat	*Chaerephon plicatus*	**EU912359**	**EU912341**
		Family Vespertilionidae	rickett's big-footed myotis	*Myotis ricketti*	**EU912377**	**EU912345**
			cave myotis	*Myotis velifer*	AM263197	
			tube-nosed bat	*Murina leucogaster*	**EU912379**	
			Japanese pipistrelle	*Pipistrellus abramus*	**EU912360**	**EU912347**
		ORDER CARNIVORA				
		Family Canidae	domestic dog	*Canis lupus*	XM_539386	XM_538203
		ORDER CETARTIODACTYLA				
		Family Bovidae	domestic cow	*Bos taurus*	NM_174567	
			wild boar	*Sus scrofa*	AY091587	NM_174566
		ORDER PERISSODACTYLA				
		Family Equidae	horse	*Equus caballus*	XM_001502735	
	Eu	ORDER RODENTS				
		Family Sciuridae	eastern gray squirrel	*Sciurus carolinensis*	DQ302163	
		Family Caviidae	domestic guinea pig	*Cavia porcellus*	AY552608	
		Family Muridae	house mouse	*Mus musculus*	NM 007538	
			brown rat	*Rattus norvegicus*	NM 031015	
		ORDER PRIMATES				
		Family Cercopithecidae	crab-eating macaque	*Macaca fascicularis*	AF158977	
			rhesus monkey	*Macaca mulatta*	XM_001091869	
			Angolan talapoin	*Miopithecus talapoin*	L76226	
		Family Hominidae	human	*Homo sapiens*	DQ822478	NM_020061
			common chimpanzee	Pan troglodytes	NM_001009127	
		Family Cebidae	weeper capuchin	*Cebus olivaceus*	AF039424	
			black-capped squirrel monkey	*Saimiri boliviensis*	U53875	
			common marmoset	*Callithrix jacchus*	L76201	
		Family Tarsiidae	Horsfield's tarsier	*Tarsius bancanus*	DQ191949–53	
			Philippine tarsier	*Tarsius syrichta*	DQ191954–58	
		Family Daubentoniidae	aye-aye	*Daubentonia madagascariensis*	DQ191898–02	
		Family Lemuridae	ring-tailed lemur	*Lemur catta*	DQ191940–44	

		Classification	Common name	Species name	*SWS1* gene	*M/LWS* gene
		Family Indridae	Coquerel's sifaka	*Propithecus verreauxi*	DQ191935–39	
			eastern woolly lemur	*Avahi laniger*	DQ191893–97	
		Family Lepilemuridae	red-tailed sportive lemur	*Lepilemur ruficaudatus*	DQ191945–48	
		Family Cheirogaleidae	giant mouse lemur	*Mirza coquereli*	DQ191903–07	
			gray mouse lemur	*Microcebus murinus*	DQ191922–26	
			lesser dwarf lemur	*Cheirogaleus medius*	DQ191908–12	
			greater dwarf lemurs	*Cheirogaleus major*	DQ191917–21	
At	Af	ORDER PROBOSCIDEA				
		Family Elephantidae	African bush elephant	*Loxodonta africana*	AY686753	
Au		ORDER DIPROTDONTIA				
		Family Macropodidae	tammar wallaby	*Macropus eugenii*	AY286017	
			quokka	*Setonix brachyurus*	AY726545	
		ORDER PERAMELINA				
		Family Peramelidae	southern brown bandicoot	*Isoodon obesulus*	AY726544	
		ORDER DASYUROMORPHIA				
		Family Dasyuridae	fat-tailed dunnart	*Sminthopsis crassicaudata*	AY442173	

*Newly generated sequences are indicated by bold font. Superordinal groupings of species are abbreviated as follows: Boreoeutheria (Bo), Laurasiatheria (La), Euarchontoglires (Eu), Atlantogenata (At), Afrotheria (Af), and the Australidelphia (Au) *sensu* (ref. 21 and refs. therein).

Table S2. Likelihood values and parameter estimates for bat opsin genes

Model	ℓ	P	Parameter estimate	Positive selected (PS) sites
SWS1				
One-ratio	−8120.97	98	ω=0.143	None
Free-ratio	−8011.43	193	variable ω by branch	n/a
Two-ratio (fruit bats)	−8177.00	99	ω_0=0.145, ω_1=0.000	n/a
Two-ratio (Yinpterochiroptera)	−8120.89	99	ω_0=0.143, ω_1=0.294	n/a
Two-ratio (Yangochiroptera)	−8120.46	99	ω_0=0.144, ω_1=0.067	n/a
M1a	−7919.60	99	ρ_0=0.853 (ρ_1=0.147), ω_0=0.070 (ω_1=1.000)	n/a
M2a	−7905.07	101	ρ_0=0.851, ρ_1=0.140 (ρ_2=0.008), ω_0=0.072, ω_1=1.000 (ω_2=4.372)	33V (P = 1.000), 78H (P = 0.999)
M7	−7877.90	99	P = 0.248, q = 1.112	n/a
M8	−7855.55	101	ρ_0=0.989, P = 0.289, q = 1.576, ρ_1=0.011, ω=3.701	33V (P = 1.000), 78H (P = 0.999)
M/LWS				
One-ratio	−2191.62	38	ω=0.092	None
Free-ratio	−2171.58	73	variable ω by branch	n/a
Two-ratio (fruit bats)	−2176.42	39	ω_0=0.067, ω_1=0.098	n/a
Two-ratio (Yinpterochiroptera)	−2175.71	39	ω_0=0.070, ω_1=0.000	n/a
Two-ratio (Yangochiroptera)	−2176.37	39	ω_0=0.067, ω_1=0.142	n/a
M1a	−2135.19	39	ρ_0=0.901(ρ1 = 0.099), ω0 = 0.015 (ω1 = 1.000)	n/a
M2a	−2135.19	41	ρ0 = 0.901, ρ1 = 0.099 (ρ2 = 0.000), ω0 = 0.015, ω1 = 1.000 (ω2 = 17.761)	None
M7	−2134.60	39	P = 0.047, q = 0.404	n/a
M8	−2133.91	41	ρ0 = 0.980, P = 0.059, q = 0.732, ρ1 = 0.020, ω=1.609	93V (P = 0.980)

Table S3. Likelihood ratio tests (LRTs) for bat opsin genes

Comparisons	$2\Delta\ell$	df	*P*-value
SWS1			
One-ratio vs. Free-ratio	219.08	95	<0.001
One-ratio vs. Two-ratio (fruit bats)	112.06	1	<0.001
One-ratio vs. Two-ratio (Yinpterochiroptera)	0.16	1	0.689
One-ratio vs. Two-ratio (Yangochiroptera)	1.02	1	0.313
M1a vs. M2a	29.08	2	<0.001
M7 vs. M8	44.70	2	<0.001
M/LWS			
One-ratio vs. Free-ratio	40.08	35	0.255
One-ratio vs. Two-ratio (fruit bats)	30.40	1	<0.001
One-ratio vs. Two-ratio (Yinpterochiroptera)	31.82	1	<0.001
One-ratio vs. Two-ratio (Yangochiroptera)	30.50	1	<0.001
M1a vs. M2a	0	2	1
M7 vs. M8	1.39	2	0.500

Table S4. Gene vs. species tree for the bat *SWS1* opsin dataset

tests	Log likelihood scores	Δ in–ln likelihood	*P* Values for KH tests	*P* Values for SH tests
Gene tree	6,827.29			
Species tree	6,843.05	15.75	0.095	0.054

Table S5. Details of primers used

Primer name	Primer sequence	Ta (°C)	Product
SWS1			
SFe	5′ ATGTCAG[G/A]GGA[G/A]GAGTTTTATCTGTTCAAG 3′	55–60°C	Exon1-Exon3 ~1200bp
SDC1R	5′ TATAG[G/T]ACTCGCTGCG[G/A]TA[T/C]TTGGTGCC 3′		
SDC2F	5′ GCAGTGTTCCTGTGG[C/T]CC[T/C]GACTGGTAC 3′	55–60°C	Exon3-Exon4 ~800bp
SINRb	5′ GGAACTGCTTATTCATGAAGCAGTAGATGATGGG 3′		
M/LWS			
LDF	5′ AGGGATCAC[G/A]GG[T/C]CTCTGGTC 3′	55–65°C	Exon3-Exon4 ~1200bp
LDaR	5′ GCTCGGATGGCCAGCCACACTTG 3′		
LDbF	5′ ACTGGCCCCA[T/C]GGCCTGAAGAC 3′	55–65°C	Exon4-Exon5 ~2000bp
LDR	5′ ACCTGCCGGTTCATAAAGACA 3′		

Table S6. Likelihood ratio tests for selecting the best-fitting model of codon evolution for the *SWS* in frame dataset

Nucleotide model name	Constraints	#Params	Likelihood	LRT p-value
JC	All ρ equal, $\omega=1$, $\sigma=1$	1	−13615.0	-
JC	All ρ equal, ω estimated, $\sigma=1$	2	−13084.7	0
JC	All ρ equal, ω, σ estimated	3	−13075.7	$2.21\text{x}10^{-5}$
K80	$\rho_{AC}=\rho_{GT}=\rho_{AT}=\rho_{CG}$ $\rho_{AG}=\rho_{CT}$	4	−12742.7	0
TrN	$\rho_{AC}=\rho_{GT}=\rho_{AT}=\rho_{CG}$	5	−12711.8	$3.78\text{x}10^{-15}$
TIM	$\rho_{AC}=\rho_{GT}$ $\rho_{AT}=\rho_{CG}$	6	−12707.9	0 (vs TrN)
GTR	6 independent ρ parameters	8	−12705.2	0.067
TIM+γ	as TIM, plus gamma distribution of rates across sites	9	−12576.6	0 (vs TIM)
TIM + I+γ	As above, plus invariant sites	10	−12576.6	1
TIM+γ+local ω	As TIM + G, with different ω value for each branch	139	−12025.3	0 (vs TIM+ γ)
TIM+γ+local ω + local σ	As above, with different σ value for each branch	270	−11982.8	0.99

Models are ordered from the simplest (lowest numbers of parameters) at the top of the table, to most complex at the bottom. Likelihood ratio tests are used to compare each model to the model directly above it, unless otherwise specified; tests not carried out in this order are separated by a blank line. If $P < 0.05$ (marked with an asterisk), then the simpler model is rejected in favor of the more complex model; otherwise, the simpler model is compared to an alternative, more complex model below the next blank line until the final comparison with the most parameter-rich model we investigated. The order of tests for the first 7 rows are as implemented in Modeltest version 3.7 (6). The best-fitting model is TIM+γ+local ω.