As Blind as a Bat? Opsin Phylogeny Illuminates the Evolution of Color Vision in Bats.

Bruno F. Simões, Nicole M. Foley, Graham M. Hughes, Huabin Zhao, Shuyi Zhang, Stephen J. Rossiter and Emma C. Teeling

a UCD School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland;
b School of Earth Science, University of Bristol, Bristol, BS8 1TG, United Kingdom
%c School of Biological Science, The University of Adelaide, North Terrace, South Australia 5005, Australia.
d College of Life Sciences, Wuhan University, Wuhan, Hubei 430072, China
e College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, Shenyang 110866, China.
f School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom

*Corresponding Author: E-mail: emma.teeling@ucd.ie

Abstract

Through their unique use of sophisticated laryngeal echolocation bats are considered sensory specialists amongst mammals and represent an excellent model in which to explore sensory perception. While several studies have shown that the evolution of vision is linked to ecological niche adaptation in other mammalian lineages, this has not yet been fully explored in bats. Recent molecular analysis of the opsin genes, which encode the photosensitive pigments underpinning color vision, have implicated high-duty cycle (HDC) echolocation and the adoption of cave roosting habits in the degeneration of colour vision in bats. However, insufficient sampling of relevant taxa has hindered definitive testing of these hypotheses. To address this, novel sequence data was generated for the SWS1 and MWS/LWS opsin genes and combined with existing data to comprehensively sample species representing diverse echolocation types and niches (SWS1 n=115; MWS/LWS n=45). A combination of phylogenetic analysis, ancestral state reconstruction and selective pressure analyses were used to reconstruct the evolution of these visual pigments in bats and revealed that while both genes are evolving under purifying selection in bats, MWS/LWS is highly conserved but SWS1 is highly variable. Spectral tuning analyses revealed that MWS/LWS opsin is tuned to a long wavelength, 555 - 560 nm in the bat ancestor and the majority of extant taxa. The presence of UV vision in bats is supported by our spectral tuning analysis, but phylogenetic analyses demonstrated that the SWS1 opsin gene has undergone pseudogenization in several lineages. We do not find support for a link between the evolution of HDC echolocation and the pseudogenization of the SWS1 gene in bats, instead we show the SWS1 opsin is functional in the HDC echolocator, Pteronotus pannelli. Pseudogenization of the SWS1 is correlated with cave roosting habits in the majority of pteropodid species and divergence time estimates suggest this loss is coincident with climatic events. Together these results demonstrate that the loss of UV vision in bats is more widespread in bats than was previously considered and further elucidate the role of ecological niche specialization in the evolution of vision in bats.

Keywords: Photic Adaptation, Colour Vision, Sensory Trade-Off, Visual Pigments, Bats
Introduction

Bats possess some of the most unique and peculiar adaptations observed amongst extant mammals that render them as excellent models in which to study the evolution of sensory perception. They are successful, nocturnal animals (~1260 species), the only mammals that can truly fly (Simmons 2005) and they exist in diverse ecological niches throughout the globe, feeding on insects, small mammals, fish, blood, nectar, fruit and pollen (Teeling et al. 2018). Bats are the only mammals to use laryngeal echolocation (Teeling et al. 2005; Teeling 2009; Teeling et al. 2016) for hunting, avoiding obstacles and orienting in scotopic or low light conditions. This unique auditory capability shows great variation amongst the 21 families of echolocating bats (Teeling et al. 2016; Teeling et al. 2018) and appears to be related to lineage-specific selection pressures as well as shared ancestry (Teeling et al. 2016). It has been argued that bats have developed this acoustic sense at the expense of their other senses, such as vision, given the typically small size of an echolocating bat’s eyes (Eklöf et al. 2014) and has led to the popular use of the term “as blind as a bat”. One family of bats, the Pteropodidae (~186 species), does not use laryngeal echolocation but instead has large sensitive eyes, specialised for nocturnal vision.

In vertebrates, photosensitive molecules within the outer segments of rods and cones of the retina determine the spectral sensitivity of the eye. Each photosensitive molecule is encoded by a distinct opsin gene (Yokoyama and Radlwimmer 1998). Rods function in dim light and contain the rhodopsin pigment, which is most sensitive between 475 - 505nm (eg, Yokoyama 2008; Nathans et al. 1986). Cones are responsible for colour vision and mammals can possess up to two types of opsin molecules with different sensitivities ranging from ~360 - 440nm (short wavelength opsin; SWS1) and ~536 - 560nm (medium-to-long wavelength opsin; MWS/LWS) (Yokoyama and Yokoyama 1996). Comparisons of visual pigments across taxa indicate that spectral tuning and, therefore, the wavelength of peak light sensitivity ($\lambda_{\text{max}}$), is generally modulated by five key critical amino acid sites in MWS/LWS opsins (Yokoyama and Radlwimmer 1998) and at least 11 amino acid sites in SWS1 opsin (Yokoyama et al. 2006). Therefore, while the five-site rule may be subject to allelic variation such as that observed in guppies (Kawamura et al 2016) or may be reduced to a “three-site” rule in certain primates (Matsumoto et al. 2014), it is possible to infer the $\lambda_{\text{max}}$ from opsin sequences using sequence based analyses. In this way, the spectral type of the SWS1 (UV: ~360nm or Visible: >400nm) can be broadly determined, while $\lambda_{\text{max}}$ predictions for the MWS/LWS are typically more accurate (Hauser et al. 2014).

The evolution of the visual opsins is tightly correlated with ecological niche adaptation (Douglas and Jeffery 2014). The MWS/LWS is remarkably conserved across most mammals (Bowmaker and Hunt 2006), but recent studies suggest that the MWS/LWS opsin gene has undergone pseudogenization in some cetaceans, possibly coincident with the loss of the SWS1 opsin gene (Meredith et al. 2013; Emerling et al. 2015; Springer et al. 2016). This adaptation is thought to be driven by inhabiting a deep-water environment and feeding on bioluminescent prey (Meredith et al. 2013; Springer et al. 2016). Adaptation to a fossorial niche is hypothesized to have driven the loss of both cone opsins (SWS1 and MWS/LWS) in xenarhthans (Emerling and Springer 2015), the naked-mole rat (Kim et al. 2011), and the star-nosed and golden moles (Emerling and Springer 2014, 2015). Similar adaptations to burrowing habitats have been observed in fossorial snakes (Simões et al. 2016), caecilians (Mohun et al. 2010) and cave salamanders. Zhao et al. (2009a) showed that the MWS/LWS gene was conserved in all bat species but showed the SWS1 gene had undergone dramatic divergent selection among lineages. This and subsequent analyses have shown that the SWS1 opsin is functional in the majority of bat lineages and is sensitive to UV light (Zhao et al. 2009a; Emerling et al. 2015). These studies have also suggested that loss of function mutations in the SWS1 has potentially coincided with
the acquisition of high-duty cycle (HDC) echolocation, as well as with changes in roosting ecology in some lineages (e.g. cave roosting in Pteropodidae; Zhao et al. 2009a), however, incomplete sampling has hampered definitive testing of these hypotheses.

The loss of the SWS1 is notably widespread across mammals, especially in lineages living in low light environments (Hunt et al. 1998; Bowmaker and Hunt 2006; Jacobs 2013; Meredith et al. 2013; Davies et al. 2015; Emerling et al. 2015; Wu et al. 2017). Therefore, it has been suggested that the loss of the SWS1 opsin is a non-adaptive process that has been tolerated by most mammals given that the interaction between the rhodopsin and MWS/LWS opsin still allows some colour discrimination (Griebel and Schmid 1992). Increased interaction with other sensory systems (e.g. olfaction, hearing) may also reduce the impact of SWS1 opsin loss on species’ survival (Zhao et al. 2009a; Fujun et al. 2012; Jacobs 2013). Furthermore, spectral shifts in UV sensitivity to violet/blue vision has occurred at least 12 times during mammalian evolution (Emerling et al. 2015). Given the correlation between opsin evolution and ecological niche specialisation, the functionality and evolution of opsin genes has been used as a proxy to infer the ecology of ancestral lineages (Tan et al. 2005). Therefore, elucidating the evolution and functionality of cone visual pigments in ecologically divergent taxa can shed light on how vision has contributed towards mammalian adaptation to diverse ecological niches and also further our understanding of the evolutionary history of mammals.

Here, we have gathered, amplified and sequenced the most extensive dataset of opsin genes across the largest taxonomic representation of bat species (n = 115) to date. Our dataset, which combines existing and novel sequence data, encompasses the vast diversity of ecological niches and acquisition of unique senses (visual, olfaction, thermal and acoustic) across ~65 million years of bat evolutionary history (Miller-Butterworth et al. 2007; Meredith et al. 2011; Foley et al. 2014; Foley et al. 2016; Teeling et al. 2018). We used selection estimates (dN/dS ratios or ω) to elucidate the evolutionary forces acting on opsin genes (relaxed, neutral or positive selection) in bats and assessed their influence on the spectral tuning of these visual pigments. Using comprehensive taxonomic sampling we determined if the loss of SWS1 is always coincident with 1) the evolution of high-duty cycle echolocation in echolocating bats and 2) cave roosting in pteropodids. Furthermore, molecular clock dating, selection tests and analysis of substitution rates were used to elucidate the timing of the SWS1 pseudogenization event and aid our understanding of the selective pressures that have potentially driven these events in bats.

Results

The SWS1 dataset of 115 bats species resulted in a coding alignment of 1125-1325bp. The MWS/LWS dataset comprised 45 bat species and resulted in a coding alignment of 536bp. Where applicable, COX1 barcoding was used to confirm each species identity. In instances where the species was not present in the BOLD database a cut-off of <=90% sequence identity was used to identify samples to genus level.

MWS/LWS functionality, sensitivity and phylogenetics

The MWS/LWS opsin gene was found to be functional and highly conserved across all bats, with no indels, frameshifts or apparent alternative splice sites. Phylogenetic reconstructions conducted using Maximum Likelihood (ML) and Bayesian Analyses (BA) recovered the consensus bat species tree (fig. 1) (Teeling et al. 2000; Teeling et al. 2002; Teeling et al. 2005; Miller-Butterworth et al. 2007;
Meredith et al. 2011; Foley et al. 2014). ML bootstrap support (BS) was low across the tree (fig. 1) while Bayesian posterior probabilities (BPP) were higher throughout. Ancestral state reconstruction and analyses of the spectral tuning sites (λmax) exhibited high levels of conservation throughout bats (fig. 1). The majority of extant bat species (n = 31) and the putative crown-group ancestor have a substitution in the first spectral tuning site, S180A (S→A at site 180), present in eutherian and Laurasiatherian ancestors (Lui et al., 2018) which differs from the ancestral vertebrate condition of ‘SHYTA’ (~560 nm; Yokoyama et al. 2008) and indicates sensitivity to wavelengths of ~555 nm (fig. 1). The ancestral eutherian spectral pattern (‘SHYTA’) (Lui et al. 2018, Yokoyama et al. 2008) is observed in the Megadermatidae, *Rhinolophus sinicus*, the Vespertilionidae (except *Myotis ricketti*) and *Artibeus jamaicensis*. A medium-wavelength shift resulting from a mutation in the fourth spectral site, T285A is observed in *Myotis ricketti* and reduces the spectral sensitivity to 543 nm as independently reported by Zhao et al. (2009a). A combination of both aforementioned spectral site mutations are present in two cave dwelling pteropodid bats, *Dyacopterus rickarti* and *Thoopterus nigrescens* and reduces the spectral sensitivity to ~536 nm (fig. 1).

*Functionality and Spectral Tuning of the SWS1 opsin*

Unlike the MWS/LWS gene, the SWS1 opsin gene is highly variable across bats. Multiple indels leading to frame-shift mutations, premature stop codons and loss-of-function mutations were found in the major bat lineages of the Yinpterochiroptera (Pteropodidae, Rhinolophidae, Hipposideridae, Rhinonycteridae, and Megadermatidae), and the Yangochiroptera lineage Mormoopidae (fig. 2). Of the 115 bat species examined, the SWS1 opsin appears to be non-functional in 26 species (fig. 2). In the sub-order Yangochiroptera, loss-of-function mutations can be found exclusively in the New World family Mormoopidae. The genus *Mormoops* shares two ORF-disrupting deletions in the first and second exon. *Pteronotus gymnonotus* contains a single base pair deletion leading to a premature stop codon in exon 1. This deletion was confirmed through PCR and sequencing repeated in two different laboratories with two different specimens. No ORF disrupting mutations were observed in *Pteronotus pammelii*, suggesting the presence of a functional SWS1 protein in this species.

In the superfamily Rhinolophoidea, ORF disrupting indels can be found in the Rhinolophidae (genus *Rhinolophus*), Hipposideridae (genus *Hipposideros*) and Rhinonycteridae (genus *Paratriaenops*) which all use HDC echolocation (fig. 2). Shared indels are found within all three families (fig. 2) but not between Rhinolophidae and Hipposideridae + Rhinonycteridae, suggesting that the loss-of-function events in these families, for the most part, post-date the divergence of each lineage (Foley et al. 2014). Additionally, independent ORF-disrupting mutations can be found in 2 of the 5 species from the family Megadermatidae (*Megaderma lyra* and *Macroderma gigas*), however within *Macroderma gigas* the premature stop codon is only present in exon 5 (fig. 2). In the Pteropodidae, a shared ORF-disrupting mutation is present within the genus *Rousettus* and independent frameshift indels can be found in *Eonycteris spelaea*, *Dobsonia viridis*, *Thoopterus nigrescens* and *Eidolon helvum* (fig. 2). An 11bp deletion in exon 2 with premature stop codons was found in the tree roosting bat *Scotonycteris zenkeri* (fig. 2). All other pteropodid bats, including the cave roosting *Lissonycteris angolensis*, were shown to have a functional and intact SWS1 opsin gene. The association of the loss of the SWS1 opsin gene and acquisition of primarily cave-roosting ecologies are supported by the estimation of contrasts estimated by the Brunch algorithm estimated in cafer (supplementary information, table S5)

The intron-exon boundaries were found to be conserved across all echolocating bats and all tree roosting pteropodid bats, retaining the GT/AG pattern. Potential alternative splice sites were found in *Rousettus aegyptiacus* as well as *Rousettus amplexicaudatus*, *Eonycteris spelaea* and *Thoopterus nigrescens* as was reported in previous studies (Shi et al. 2001; Zhao et al. 2009a). An investigation of
the 11 amino acid sites responsible for the spectral tuning of the SWS1 visual pigment revealed that for all bat species analysed the SWS1 opsin is UV sensitive. The amino acid positions 86, 93 and 113 (bovine rhodopsin numbering) play a major role in the tuning of the vertebrate SWS1 visual pigments to UV (Shi et al. 2001; Yokoyama et al. 2008) with site 86 being the biggest contributor. All these spectral sites are highly conserved across bats. Independently of the method used the ancestral reconstruction of the SWS1 opsin for all major lineages spectral sites indicated that the bat ancestral SWS1 opsin was sensitive to UV with a phenylalanine present at site 86 in all extant and ancestral lineages. This suggests that no major changes occurred in bats’ spectral sensitivity prior to multiple losses of the SWS1 in select bat lineages.

SWS1 opsin phylogenetics

ML and BA of the SWS1 complete datasets support the consensus phylogenetic position of the majority of bat families and the monophyly of the two sub-orders Yangochiroptera and Yinpterochiroptera (fig. 3). However, the ML BS and BPP support across the topology of the tree were low (BS 30 and BPP 71, for the Yangochiroptera/Yinpterochiroptera split, fig. 3). The branch lengths in the SWS1 opsin tree vary greatly across lineages and are particularly long in non-functional branches due to the large numbers of indels/stop codons in Rhinolophidae, Hipposideridae, Rhinonycteridae and Mormoopidae (fig. 3).

Phylogenetic analyses of the functional SWS1 data unite the laryngeal echolocating bats (fig. S1). Alternative topology tests were used to compare phylogenetic support for the gene tree (fig. 3) topology versus species tree topology (Teeling et al. 2016) from our data. Alternative topology tests could not reject either topology (SH = 0.811 and SH = 0.189, respectively). Convergence between echolocating bats may be due to non-neutral convergent amino acid evolution between the Rhinolophoidea and Yangochiroptera. When this was tested, it was found that among 2 branch pairs examined, nine convergent amino-acid sites, which included one spectral site F47L (F46L in the bovine rhodopsin), drives the convergent signal uniting echolocating bats in a single clade.

dN/dS analyses in the MWS/LWS and SWS1 opsin genes

The ω (dN/dS ratio) estimates across all branches and all sites for the MWS/LWS opsin gene was 0.0827, indicating that the MWS/LWS opsin gene is evolving under strong purifying selection (Table 1). The ω estimates for each branch of the tree were consistently <1 (0.00 – 0.61) across all bat lineages. The two-ratio model for lineages with HDC echolocation had a higher ω value (ω = 0.1504) than the background (ω = 0.0736) (Table 1). Conversely, Yinpterochiroptera and non-echolocating bats have higher ω estimates (0.0943 and 0.0940, respectively) than the backgrounds branches to which they were compared (0.0759 and 0.0831, respectively). The ω values are lower for cave roosting bats and insectivore/carnivore bats (0.081 and 0.082, respectively) than the background branches (0.090 and 0.086, respectively). The site models M8beta and M2a detected 4 individual sites under positive selection under Bayes empirical Bayes analysis (supplementary information table S4). Both models detected positive selection in one of the 5 amino acid positions involved in the spectral tuning of the MWS/LWS opsin, S180A, located in the transmembrane domain 4. The FUBAR analysis indicated that this site is evolving under pervasive diversifying selection. The other sites under positive selection are located in transmembrane domains 4 and 5 and one in the extracellular loop 4. Episodic selection (selection affecting only a subset of lineages) was detected in only one lineage, the cave-roosting pteropodid Lissonycteris angolensis (ω mean = 0.64, p < 0.005).

The estimates of ω values for the SWS1 opsin on all branches and all sites was 0.1947, which is significantly smaller than 1, suggesting that the SWS1 opsin gene is also evolving under purifying selection in bats (Table 1). The free ratio model for all bats fits the data significantly better than the simple one-ratio model. This finding, together with results from multiple two-ratio models, in which
the foreground branch is allowed to have a different ratio from the rest of the tree (background), suggests that different selective pressures may be acting along one or more lineages. Results from the two-ratio models confirm that \( \omega \) values in lineages where \( SWS1 \) is expected to be non-functional (\( \omega = 0.511 \)) are higher than branches where \( SWS1 \) is functional (\( \omega = 0.112 \)). In addition, the two-ratio analysis showed that lineages with HDC echolocation had a higher \( \omega \) value (\( \omega = 0.4646 \)) compared to background branches (\( \omega = 0.1638 \)). Furthermore, \( \omega \) estimates were higher in the Yinpterochiroptera (\( \omega = 0.2448 \)) than the Yangochiroptera (\( \omega = 0.1334 \)). Similar differences in the \( SWS1 \) between Yinpterochiroptera and Yangochiroptera can be found if we remove the pseudogenes (0.116 vs 0.092, \( P=9.39 \)). This analysis also showed that non-echolocating bats (Pteropodidae) had a higher \( \omega \) value (0.2047; \( n=46 \)) than echolocating bats (0.1836; \( n=69 \)). The \( \omega \) values are also higher in primarily cave roosting bat species (\( \omega = 0.2957, n=53 \)) than bats that roost in trees or other roosting sites (\( \omega = 0.1219, n=62 \)).

The \( \omega \) values are similar between species that feed on insects and other animals (\( \omega = 0.1845, n=56 \)) and animal that feed on fruits and nectar (\( \omega = 0.2014, n=59 \)). The site models M8beta and M2a (BEB) detected 5 and 1 individual sites under positive selection, respectively across the entire data set (supplementary information, Table S4). Further examination showed that none of these sites are involved in the spectral sensitivity of the visual pigment and all are located in extra- and intra cellular loops of the \( SWS1 \) opsin. Episodic directional selection was detected in the \( SWS1 \) opsin gene in \textit{Pteronotus gymnonotus} (\( \omega_{mean} = 0.56, p < 0.005 \)), in the Rhinolophidae (\( \omega_{mean} = 0.69, p < 0.005 \)) and within this family in \textit{Rhinolophus rex} (\( \omega_{mean} = 0.64, p < 0.005 \)).

Since the Pteropodidae rely mainly on vision rather than echolocation as their primary mode of sensory perception, we performed targeted selective pressure variation analyses within this group. The \( \omega \) estimate across the Pteropodidae was 0.2247, confirming that the \( SWS1 \) opsin gene is evolving under purifying selection within this bat family. Since the majority of cave roosting pteropodid bats have a non-functional \( SWS1 \) opsin, the \( \omega \) estimate for foreground branches ‘non-functional’ and ‘cave roosting’ were similar (\( \omega_{non-functional SWS1} = 0.5173; \omega_{cave-roosting} = 0.5384 \)) and higher than their respective background comparative branches (\( \omega_{functional SWS1} = 0.1154; \omega_{tree-roosting} = 0.1132 \)) (Table 1). The site models M8beta and M2a detected 7 and 4 sites under positive selection, respectively (Table 1). These sites are located in transmembrane domains and extra and intra cellular loops.

\textit{Dating the Loss of the SWS1 opsin}

The sequencing and phylogenetic analysis of \textit{Cyt b} completed gaps in the published phylogenies for Mormoopidae, Megadermatidae and Pteropodidae (fig. S2-4). ML and BA estimated similar tree topologies, the majority of which were in agreement with published phylogenies, although some BS and BPP values were low (fig. S2-4). Within Mormoopidae, \textit{Pteronotus personatus} was placed as basal to \textit{Pteronotus quadridens}, \textit{Pteronotus macleayii}, \textit{Pteronotus gymnonotus} and \textit{Pteronotus davyi} (fig. S2). In Megadermatidae, the \textit{Cyt b} phylogeny did not support the monophyly of the genus \textit{Megaderma} within the Megadermatidae (fig. S3). Dating the divergence of the 3 families mentioned suggested that \textit{Megaderma lyra} diverged from the other megadermatids around 26.6MYA and \textit{Macroderma gigas} diverged from \textit{Lavia frons} 7.9MYA (fig. S3). Within Mormoopidae, \textit{P. gymnonotus} diverged from \textit{P. davyi} approximately 3.8MYA. The genus \textit{Mormoops} diverged from the other Mormoopidae 33.8MYA and both Mormoops species diverged from each other approximately 17MYA (fig. S2). In the Pteropodidae, \textit{Dyacopterus} diverged from \textit{Saphaerias} 14.9MYA, \textit{Thoopterus} diverged 16.3MYA from the other Pteropodids, \textit{Scotonycteris} diverged from \textit{Casinycteris} 8.4MYA, \textit{Eonycteris} diverged approximately 17.8MYA and the crown-group \textit{Rousettus} radiated 10.4MYA (fig. S4).

Divergence time estimates were consistent with finding from previous studies (Table S1) and indicated the oldest pseudogenization event in the \textit{SWS1} opsin gene occurred in \textit{Mormoops},
Rhinolophidae and Hipposideridae approximately 29.8, 29.1 and 28.4 MYA, respectively. In the megadermatids, *Megaderma lyra* and *Macroderma gigas* lost the SWS1 opsin approximately 4.06 and 2.86 MYA, respectively (fig. S2-4). Pseudogenization of the SWS1 in the Pteropodidae occurred multiple times in the majority of the cave roosting bats ranging from 6.79 MYA in *Dobsonia* and 12.01 MYA in *Rousettus* (fig. S4).

**Discussion**

The functionality of the bat visual system has been debated for many years, with common misconceptions compounded in the popular simile “as blind as a bat”. However, genetic (Zhao et al. 2009a, 2009b; Wu et al. 2017), behavioural (Fujun et al. 2012) and immunocytochemical studies of the visual pigments (Müller and Peichl 2005; Müller et al. 2007) have revealed that earlier expectations of blindness or pure rod retinas in bats (Peichl 2005) were incorrect.

**MWS/LWS evolution in bats**

Wang et al., (2004) proposed the MWS/LWS opsin gene had potentially undergone duplication in the pteropodid *Haplonycteris fisheri* and further suggested that the presence of two copies may increase the expression and sensitivity of this visual pigment in this species. Our sequence comparison of clones derived from each pteropodid species, examined as part of this study, did not recover any indels or mutations, indicative of additional copies of the MWS/LWS opsin gene (Yokoyama and Radlwimmer 1998; Wang et al. 2004). This result provides strong support for the presence of a single copy of the MWS/LWS opsin gene in the pteropodid species we examined, suggesting that any duplication event in the Pteropodidae is likely to be limited to the monotypic genus *Haplonycteris*. Together, these results suggest that the majority of bats are potentially dichromatic given the additional presence of a SWS1 functional opsin. Our spectral tuning analysis of the 5 amino-acid sites responsible for the \( \lambda_{\text{max}} \) revealed that the majority of bat MWS/LWS visual pigments are tuned to a long-wavelength (~555 – 560nm) (fig. 1). Ancestral state reconstructions demonstrated that the ancestor of all bats and those of the four major bat lineages (Rhinolophoidea, Emballonuroidea, Noctilionoidea and Vespertilionoidea) were tuned to 555nm (fig. 1). An increase of \( \lambda_{\text{max}} \) to ~ 560nm was observed in the majority of species sampled from the families Vespertilionidae and Megadermatidae and also in isolated species within the Rhinolophidae, Hipposideridae and Phyllostomidae (fig. 1). Given that the laurasiatherian ancestor had a red-tuned MWS/LWS opsin gene tuned to ~ 555nm (Yokoyama and Radlwimmer 1998), we show that ancestral laurasiatherian pattern is retained in the majority of bat species (fig. 1).

Assays conducted in vitro have demonstrated that the light reflected by the tapetum lucidum in *Pteropus giganteus* is red tuned (Blackwood et al. 2010). This suggests that 555 - 560nm wavelength sensitive vision is complemented by other visual abilities which may be particularly advantageous in nocturnal conditions. However, the MWS/LWS appears to have undergone a green-shift in two pteropodid species, *Dyacopterus rickarti* \( (\lambda_{\text{max}} \sim 536\text{nm}) \) and *Thoopterus nigrescens* \( (\lambda_{\text{max}} \sim 536\text{nm}) \), and one vespertilionid species, *Myotis ricketii* \( (\lambda_{\text{max}} \sim 543\text{nm}) \). *Myotis ricketii* exhibits the lowest frequency rhodopsin \( \lambda_{\text{max}} \) (497nm) observed in Chiroptera (Levenson and Dizon 2003; Zhao et al. 2009b), suggesting that parallel \( \lambda_{\text{max}} \) changes in the MWS/LWS opsin gene may be adaptive. The medium-wavelength MWS/LWS shift observed in the two pteropodid species, *Dyacopterus rickarti* and *Thoopterus nigrescens*, may be linked with the loss of the SWS1 opsin. Changes in spectral sensitivity in functional opsins as a consequence of the pseudogenization of the SWS1 opsin has been suggested in
cetaceans (Levenson and Dizon 2003; Meredith et al. 2013), prior to the discovery that the MWS/LWS opsin gene was non-functional in this lineage (Meredith et al. 2013). The ω estimates from our selection test analyses showed strong purifying selection acting on the MWS/LWS opsin gene in bats (Table 1). This suggests that despite the acquisition of laryngeal echolocation and a long history of nocturnality, the MWS/LWS opsin gene has evolved under very strong functional constraint in bats since they diverged from other laurasiatherians (Cavallari et al. 2011; Meredith et al. 2011). The MWS/LWS opsin gene is however under significant lower functional constrain in HDC lineages (table 1), suggesting that this type of echolocation may be correlated with relaxed long-wavelength vision. This is also supported by codon-based likelihood clade models (Gutierrez et al., 2018). With the exception of cetaceans and fossorial rodents, the MWS/LWS opsin is highly conserved among mammals (Yokoyama and Radlwimmer 1998; Zhao et al. 2009a), which highlights the importance of medium to long wavelength vision for mammals. Therefore, MWS/LWS vision may be fundamental for bat survival and may play an important role in the regulation of circadian rhythms which calibrate physiological processes and behaviour (Nowak 1994; Cavallari et al. 2011). Furthermore, the ability to process longer wavelengths may be more important for image formation than shorter wavelengths such as UV light (Douglas and Jeffery 2014). Longer wavelengths are used preferentially in luminance vision and acromatic intensity detection providing more signal power and thus visual information (Osorio and Vorobyev, 2005). The importance of long-wavelength vision is particularly evident since the LWS opsin in vertebrates originated from a gene duplication predating all other cone opsins (Yokoyama, 2000).

SWS1 evolution in bats: UV vision, HDC echolocation and cave roosting

Previous studies have suggested that bats possess UV short-wavelength vision, based on analyses of the 11 amino acids responsible for the tuning of this visual pigment (Jacobs 2009; Zhao et al. 2009a). Although caution is recommended when predicting $\lambda_{\text{max}}$ from SWS1 amino acid sequences (Hauser et al. 2014), UV vision in bats is further supported by electoretinographic recording in two Phyllostomidae species, Carollia perspicillata and Glossophaga soricina (Wang et al. 2004; Müller et al. 2009) and behavioural studies (Fujun et al. 2012). Our spectral tuning analysis of the 11 sites responsible for light sensitivity in the SWS1 opsin gene in both ancestral and extant bat species, provide further support for the presence of UV vision in bats. Furthermore, our results demonstrate that this visual pigment has been UV sensitive in all bats since they diverged from other laurasiatherians ~78MYA (Foley et al. 2016). UV sensitive SWS1 opsins have a slower retinal release, a smaller binding pocket, increased dark stability, and a narrower absorption curve (Hauser et al. 2014), which together suggest that this visual pigment is adapted to mesopic and scotopic (mid - dim light) conditions. Amongst mammals, a UV sensitive SWS1 opsin is associated with a nocturnal lifestyle (Veilleux et al. 2012) and is thought to be particularly advantageous at dawn and dusk (Peichl 2005; Müller et al. 2009; Zhao et al. 2009a; Zhao et al. 2009b; Jacobs 2013). This may be particularly important for nectar-feeding lineages such as the Phyllostomidae which use UV vision to detect UV-reflecting flowers (Winter and von Helversen 2003).

In mammals, the evolutionary history of the SWS1 opsin gene is remarkably different compared to the highly conserved MWS/LWS or the RH1 opsin (Zhao et al. 2009a; Zhao et al. 2009b). Our data show that the SWS1 opsin gene has undergone multiple pseudogenization events in bats, evidenced by the presence of frame-shift indels and premature stop codons in key locations in many pteropodid and echolocating bat lineages (fig. 2). Previous studies have suggested a sensory trade-off between SWS1 opsin functionality and HDC echolocation (Zhao et al. 2009a), however, for the first time, our analyses show that the SWS1 opsin has undergone pseudogenization in more echolocating bat lineages and is not limited to HDC echolocators (fig. 2). All Old World HDC echolocators (families Rhinolophidae
Hipposideridae, Rhinonycteridae) showed loss of function mutations, however this was not observed in the only other bat species reported capable of HDC echolocation, the New World Pteronotus parnellii. This suggests that acquisition of HDC echolocation does not always coincide with loss of function of the SWS1 opsin in bats, as previously suggested (Zhao et al., 2009a). Loss of functionality of the SWS1 was also observed in low-duty cycle (LDC) echolocating species from the families Megadermatidae and Mormoopidae (fig. 2), suggesting a more complex interplay of echolocation and vision in bats. The presence of partial SWS1 fragments in the Mormoops blainvillii retinal transcriptome suggest the 3 deletions may result in a translated pseudogene (Gutierrez et al. 2018).

The loss of the SWS1 opsin gene is observed in several pteropodid lineages due to presence of premature frame-shift indels and stop codons (fig. 2). These lineages include species which roost in caves (Roussettus, Stenonycteris and Eonycteris), those which roost in caves and trees (Eidolon) and a subset with as yet uncharacterised roosting ecology (Thoopterus and Dyacopterus) (Nowak 1994). Preliminary observations suggest the genus Dyacopterus may have mixed roosting habits since it has been found roosting in trees but has also been caught near caves (Nowak 1994; Giannini and Simmons 2003). Previous hypotheses which suggest the pseudogenization of the SWS1 opsin may be related to the adoption of cave roosting habits (Zhao et al. 2009a) is supported by the majority of the species sequenced as part of this study. However, two exceptions appear to contradict this pattern, specifically, the pseudogenization of the SWS1 opsin in Scotonycteris zenkeri and the presence of a functional SWS1 opsin in Lissonycteris angolensis, a cave roosting species. Potentially, these observed differences can be attributed to a very recent change in the roosting ecology of these species, which may have occurred so recently that this switch has not yet exerted an observable effect at the genomic level. Generally, it is accepted that the acquisition of cave roosting habits in many extant pteropodid genera may have occurred independently from an ancestral tree roosting habit (Giannini and Simmons 2003; Li et al. 2008). This fits well with our results and lends further support to the hypothesis that cave roosting ecology drives SWS1 opsin loss in the Pteropodidae. The measurement of spectral reflectance of fruits and flowers under different twilight and full moon conditions for monochromat primates suggest that the loss of the SWS1 may not be disadvantageous when the luminance contrasts of critical stimuli are high (Moritz, 2015).

**SWS1 loss and ecological niche adaptation**

Results from this paper and previous studies show that the evolution of the SWS1 opsin in mammals is radically different from the other cone visual pigments. Both the SWS2 and Rh2 cone opsins were lost in ancestor of eutherian mammals (Yokoyama and Shi 2000), most likely in response to inhabiting a nocturnal environment (Heesy and Hall 2010). In extant taxa, the occupation of extreme photic environments has led to the degeneration of the SWS1 and the LWS opsin genes in some cetaceans (Mizutani et al. 2013; Springer et al. 2016) and in fossorial and subterranean rodents (Emerling and Springer 2014). Furthermore, in fossorial rodents the loss of visual pigments is associated with further inactivation of genes in the cone visual pathway (Emerling and Springer 2014). However, losses of function in the SWS1 opsin gene have puzzled biologists for decades (reviewed in Jacobs et al 2013) ever since they were first discovered in some nocturnal primates, nocturnal, aquatic carnivores and in several rodent lineages (David-Gray et al. 2002; Levenson and Dizon 2003; Carvalho et al. 2006; Jacobs 2013). Interestingly, the SWS1 appears to lose function over relatively short evolutionary time frames, as observed in the rodent family Sciuridae (Carvalho et al. 2006), in the bat families Pteropodidae, Mormoopidae and Megadermatidae (fig. 2) and Lemuriforme primates (Tan et al. 2005; Jacobs 2013). Together these results suggest that the pseudogenization of the SWS1 opsin may be a more widespread
evolutionary event among mammals than was previously thought, the full extent of this event will be revealed through increased taxonomic sampling and sequencing.

The loss of functional SWS1 opsins across mammals has been suggested to be non-adaptive, given that many of these pseudogenization events are not ecologically advantageous (Jacobs 2013). Previous studies have shown that among nocturnal mammals, the spectral tuning of the SWS1 opsin appears to be strongly associated with ways of foraging, while the MWS/LWS opsin is tuned to maximise the amount of light that can be absorbed from a given environment (Veilleux and Cummings 2012). Significant shifts in selective pressures can be found in the bat SWS1 in cave roosting species (table 1) and among vegetation foragers (Gutierrez et al., 2018) and lemurs that inhabit open canopy forests generally experience higher purifying selection in the SWS1 when compared with closed canopy rainforests, which have lower short-wavelength light levels (Veilleux et al., 2013). These cases highlight the potential role of different ecological factors in SWS1 evolution. A loss of function in SWS1 has also been hypothesised as having an adaptive role in species where luminance contrasts for critical stimuli, specifically gums/saps and flowers are high (Moritz, 2015), and may explain its rapid loss of function in the frugivorous Pteropodidae. The short time frame in which Pteronotus parrnellii has developed HDC echolocation poses important questions: How fast can HDC echolocation evolve or have other mormoopids regressed to a LDC echolocation during their evolutionary histories? Is it possible that HDC echolocation was more widespread over the course of bat evolutionary history? To date, it is not known how and why HDC echolocation arose in bats or if any of the extant LDC lineages used HDC echolocation at some point in their evolutionary history (Fenton et al. 2012). Although Doppler shift compensation is mainly used by HDC bats, it also used by LDC bats, such as Pteronotus personatus, which demonstrates that LDC bats have the potential to develop their call emission and information processing capabilities (Smotherman and Guillén-Servent 2008). The observed convergent phylogenetic signal observed in the SWS1 opsin gene (fig. S1) suggests that echolocation may be linked to the evolution of photoreception in bats, possibly as consequence of changes in habitat and their visual system.

So why is the loss of SWS1 vision so widespread across mammals? Clues may lie in the lower functional constraints acting on the SWS1 opsin in bats (Table 1) and other mammals (Invergo et al. 2013; Invergo et al. 2015) when compared with the evolutionary pressures acting on the MWS/LWS (0.0827) and Rhodopsin gene (Zhao et al. 2009b). Evolutionary pressures acting on the SWS1 opsin in mammals differ to those acting on the SWS1 opsin in other vertebrate lineages such as squamate reptiles (Simões et al. 2015), snakes (Simões et al. 2016) and birds (Zhang et al. 2014). Has the nocturnal sensory bottleneck in the ancestor of all mammals (Heesy and Hall 2010) impacted the evolutionary dynamics of visual pigments in all mammals? The complete loss of coloured or oil droplets (Walls 1942; Bowmaker 2008), coupled with multiple losses of SWS1 vision seems to suggest that this is the case. Although numbers vary across mammalian lineages, the number of S-cones (expressing SWS1) is relatively low when compared with the number of L-cones (expressing a MWS/LWS) and rods (Peichl 2005). Therefore, loss of SWS1 vision may not be particularly disadvantageous since higher numbers of rods and L-cones, may still allow some colour discrimination (Griebel and Schmid 1992).

Conclusion

Our data suggest the pseudogenization of the SWS1 opsin gene is more widespread in bats than was previously thought. Our data confirm the loss of short-wavelength vision in the Old World HDC echolocators (Rhinolophidae, Hipposideridae, Rhinonycteridae); conversely our data show that the New World HDC echolocator Pteronotus parrnellii has a functional SWS1. The evolution of HDC echolocation is a relatively recent event in Pteronotus parrnellii compared to the Old World HDC
echolocators. This suggests that the short period of time between the divergence of *Pteronotus parnellii* and other *Pteronotus* species may not have been sufficient to relax the functionality of the *SWS1* opsin gene. The adoption of cave roosting habits is highly correlated with pseudogenization of the *SWS1* opsin gene in the Pteropodidae. While, typically the loss of *SWS1* in mammals appears to be non-adaptive, potentially bats may have evolved mechanisms to compensate for the loss of UV vision. This study highlights the utility of studying sensory ecology and adaptation across divergent taxa to elucidate the diversity of selective pressures driving the evolution of vision in mammals.

**Materials and Methods**

**Taxon and Genic Coverage**

The wide sensory and ecological diversity represented in extant bat taxa is reflected in the taxonomic representation of all datasets included in this study. The *SWS1* opsin gene (exons 1 - 4) was amplified, cloned and sequenced (~2.2 – 2.8kb) in echolocating bat species across the suborders Yangochiroptera and Yinpterochiroptera and included representatives from all but one (Cistugidae) currently recognised bat families (Teeling et al. 2016) (Table S2 and S3). Exons 1-5 were amplified in a subset of species, including pteropodid species, to confirm the presence of a full-length open reading frame (ORF) in these taxa (Table S2). In addition to the *SWS1* gene, the *MWS/LWS* opsins gene (exons 3 - 5) was amplified, cloned and sequenced (~3.5kb) in 18 echolocating and 7 non-echolocating (Pteropodidae) bat species (Table S2 and S3). To ensure our results were robust we sequenced multiple (1-5) specimens per species. Both novel datasets were supplemented with data from previous analyses (Wang et al. 2004; Zhao et al. 2009a; Shen et al. 2010) and with data mined from the publicly available *Myotis lucifugus* and *Pteropus vampyrus* genomes (Ensembl v63) and consisted of 115 bat species for the *SWS1* opsin gene and 45 species for *MWS/LWS* opsin gene. GenBank accession numbers and detailed sample information are provided in Table S2. A total of 11 mammals from the laurasiatherian orders Carnivora, Cetartiodactyla, Perissodactyla and Eulipotyphla were chosen as outgroups for downstream phylogenetic analyses (Table S2).

For a subset of samples, species identity was confirmed through either CO1 or Cyt b sequencing. 44 samples were barcoded through the amplification and sequencing of the CO1 gene. Given that certain analyses, such as selection analyses in PAML, require a resolved bifurcating phylogeny, Cyt b was amplified and sequenced for 6 samples to ascertain their phylogenetic positions within Mormoopidae and Megadermatidae (see below) (Table S2).

**DNA Alignments**

DNA sequences were aligned with a combination of MAFFT v6 (Katoh et al. 2002), ClustalX (Larkin et al. 2007) and Muscle (Edgar 2004) and manually refined in Geneious (Kearse et al. 2012). Intron/Exon boundaries were identified by the GT/AG boundary rule and by aligning them with published mRNA sequences, following (Zhao et al. 2009a). Exons with variable length compared to reference sequences were considered splice variants. Introns were identified through comparison to reference sequences and subsequently removed to create alignments of opsin coding/exonic regions. Sequences which contained premature stop codons or frameshift indels were designated as non-functional. A total of 4 alignments, (3 *SWS1* and 1 *MWS/LWS*) were generated. As exon 5 was sequenced in only a subset of taxa, the first alignment consisted of *SWS1* exons 1 – 5, with the fifth exon encoded as missing data in relevant taxa. The second alignment consisted of the full *SWS1* opsin exons 1 - 4 only, for all taxa. The final *SWS1* alignment consisted of functional opsin genes only (exon 1 - 5). The *MWS/LWS* alignment
contained opsin genes for all taxa sequenced.

**Phylogenetic Analyses**

jModelTest v3.7 (Posada 2008) was used to find the best fitting model of sequence evolution for each gene fragment. GTR+I+G was the best fitting model for all SWS1 opsin datasets. HKY+G was the best fit model for MWS/LWS opsin gene. All model fits were assessed using the AIC.

RAxML v7.2 (Stamatakis 2006) and Mr.Bayes v3.1.2 (Ronquist and Huelsenbeck 2003) were used generate Maximum Likelihood (ML) and Bayesian Analysis (BA) gene trees for each gene. ML analyses were executed in RAxML using majority rule (MRE) based bootstrapping criteria; randomized MP starting trees, a fast hill-climbing algorithm and the best fit model of sequence evolution. The BA included 1,100,000 generations with chains sampled every 1000 generations, random starting trees, 4 chains (3 hot and 1 cold). Convergence was assessed through monitoring of the standard deviation of split frequencies and stationarity was reached when the standard deviation of split frequencies fell below 0.01.

**dN/dS Estimates**

The CodeML package, implemented in PAML 4.4 (Yang 2007), was used to estimate the ratio of the rates of non-synonymous substitutions (dN) to synonymous substitutions (dS; dN/dS or ω) for the SWS1 (exon 1 - 4) dataset for all bats, SWS1 in Pteropodidae (exon 1 - 4) and the MWS/LWS opsin gene. Given that dN/dS analyses do not allow the presence of indels, codons containing indels were removed from all sequences prior to analysis. The composite species tree used in the CodeML analysis was based on chiropteran intraordinal relationships from Teeling et al., (2005), the Mormoopidae topology from Davalos et al., (2006) the Phyllostomidae topology from Rojas et al., (2011), the Rhinolophidae topology from Stoffberg et al., (2010), the Hipposideridae topology from Sun et al., (2009) and the Pteropodidae phylogeny from Almeida et al., (2011) (fig. S5).

Branch models allow the ω ratio to vary in branches across a given topology and therefore can detect positive selection, where ω > 1, in lineages of interest. The simplest branch model (one-ratio) only allows one ω ratio across the tree, while the more complex free-ratio assumes independent ω ratios for each branch. Branch models for the SWS1 in all bats and MWS/LWS opsin genes were used to estimate the ω ratio for 4 foreground lineages: high-duty cycle echolocating bats, non-echolocating bats (Pteropodidae), primarily cave roosting bats, insectivore/carnivore bats and lineages with a non-functional SWS1 opsin gene. In the Pteropodidae SWS1 dataset, branch models were performed on the foreground lineages of cave roosting pteropodids, SWS1 non-functional pteropodids and the tongue-click echolocators. All branch models were compared to the simplest models (one-ratio) using a Likelihood Ratio Test (LRT) and accepted if p > 0.05. Site models (M1a nearly-neutral, M2a positive selection and M7β, M8 β&ω) allow ω ratios to vary among sites (amino-acid or codons). Site models M2a and M8 were compared by LRT with site-models M1a and M7β, respectively. Bayes empirical Bayes (BEB) (Yang et al. 2005) implemented in models M2a and M8 β&ω was used to identify sites evolving under positive selection across the opsin genes.

HyPhy (Pond et al. 2005) was used to identify directional episodic selection in the SWS1 and MWS/LWS opsin genes across branches in bats (BranchSiteREL) (Kosakovsky Pond et al. 2011). A Mixed Effects Model of Episodic Selection (MEME) (Murrell et al. 2012) and Fast Unconstrained Bayesian Approximation (FUBAR) (Murrell et al. 2013) was used to detect episodic selection on amino acid sites and was implemented in Datamonkey (Delport et al. 2010).
Ancestral State Reconstruction

Ancestral sequences were inferred using a combination of Bayesian and Parsimony methods using PAML and Ancestors v1.1 (Yang 2007; Diallo et al. 2009). Given that the Bayesian method (Yang 2007) implemented in PAML does not accept indels, it was used to infer the ancestral sequence in alignments that excluded indels. Ancestors v1.1, which uses, heuristic and tree-hidden Markov models, was used to reconstruct ancestral regions containing indels in the full-length SWS1 alignment. The combination of of Bayesian and Parsimony methods using PAML was performed in an alignment with only functional SWS1 sequences and in alignment where species with stop codons/indels at spectral sites were removed. Ancestral sequences were aligned with extant opsin data and spectral sites were identified. Despite possible allelic variations to the ‘five-site’ rule observed in certain vertebrate taxa, ancestral wavelength of peak sensitivity (λmax) was reconstructed using inferred λmax values based on amino acid composition of the five key sites observed across each bat species (Yokoyama et al. 2008).

Dating loss of function in SWS1

Divergence time estimates were determined to date pseudogenization events in key lineages. Although, the timing of the basal divergence within the Megadermatidae, Pteropodidae and Mormoopidae has been estimated by previous studies (Teeling et al. 2005; Miller-Butterworth et al. 2007; Meredith et al. 2011; Foley et al. 2014), the phylogenetic relationships and timing of diversification events within these families have not yet been resolved. Publicly available Cyt b sequence data from representatives of the Megadermatidae, Pteropodidae and Mormoopidae were downloaded and gaps in taxonomic representation were assessed. New sequence data was generated for key species which were not present in the publicly available Cyt b dataset using methods as described above. Novel Cyt b data was aligned with publically downloaded sequences using MAFFT v6 (Katoh et al. 2002). jModelTest 3.7 (Posada 2008) was used to determine the best fitting model of sequence evolution for the data, according to the AIC or BIC scores. ML analysis was performed as previously described with the GTR model for the Pteropodidae dataset (outgroup Rhinolophoidea) and GTR+I+G for the Mormoopidae dataset (outgroup: Phyllostomidae) and Rhinolophoidea dataset (outgroup: Pteropodidae). Phylogenetic reconstruction and divergence time estimates for these datasets were estimated simultaneously using BEAST v1.7 (Drummond and Rambaut 2007). The uncorrelated, lognormal, relaxed-clock model, where rates were allowed to vary among branches without the a priori assumption of autocorrelation between adjacent branches, was used. Default priors were used for GTR substitution parameters (0, 100), gamma shape parameter (0, 100) and proportion of invariant sites parameter (0, 1). The uncorrelated lognormal-clock was estimated with uniform priors on the mean (0, 100) and standard deviation (0, 10). The Yule process of speciation was used as the tree prior and the starting tree was estimated with UPGMA. The ingroup was assumed to be monophyletic with respect to the outgroup. Phylogenetic calibrations were applied to several branches of the tree to constrain the analysis. 1) the split between the Pteropodidae and all other yinpterochiropters was constrained following hard bound maximum age estimates (63.79MY) obtained in Meredith et al (2011) and the crown (24MY) group Pteropodidae was constrained following estimates obtained in the analysis of Teeling et al (2005). The calibration point for the Megadermatidae was based on the divergence times of the split between Craseonycteris + Megaderma (44MY) estimated in the study of Foley et al (2014). For Mormoopidae, the calibration point was based on the hard-bound maximum age divergence time estimate for the split between Phyllostomidae + Mormoopidae (37.11MY), from Meredith et al (2011).

For each dataset, three independent Markov chain Monte Carlo (MCMC) were run for 30 million generations to ensure sufficient sampling of estimated sample size (ESS) values, with auto
optimize operators. Trees were saved every 1000 generations. Log files from each run were imported into Tracer v1.7 (Rambaut and Drummond 2007), with trees sampled from the first 1 million generations discarded as burn-in and stationarity was assessed. Tree files from the individual runs were combined using LogCombiner v1.7 (Drummond and Rambaut 2007) to find the optimal tree.

**Dating pseudogenization events and estimating ecological correlations.**

In lineages where the *SWSI* gene is pseudogenized, we estimated when functional and evolutionary constraints were relaxed using the method described in Zhao et al (2010). This method assumes a scenario in which the typical evolutionary constraint (as observed in other closely related bat lineages) was suddenly and completely relaxed ($\omega = 1$) at some time point in evolutionary history ($t$ MY), as such our analyses used: 1) the rate of nucleotide substitution increase in lineages with a pseudogenized *SWSI* gene and 2) the $\omega$ value prior to the relaxation of the functional constraint in closely related functional lineages. The dataset was subdivided into four subsets to limit the exclusion of codons due to the presence of indels in the alignments: 1) *Pteronotus*; 2) Mormoopidae + Phyllostomidae; 3) Rhinolophoida; 4) Pteropodidae. For each of these datasets we used the formula detailed in Zhao et al (2010) based on functional and non-functional lineage $\omega$ estimates, assuming divergence estimates obtained as part of this study.

Additionally, the method developed by Meredith et al (2009), in which CodeML (Yang 2007) was used to estimate the dN/dS ratios or $\omega$ in four branch categories: Functional (pre-existing to functional branches); Pre-mutation (pre-dating internal nodes where stop-codons appear); mixed (presence of first-appearance stop codons and anteceding non-functional branches) and non-functional (post-date first-appearance stop codons). This analysis was applied to 3 datasets: a) Mormoopidae; b) Rhinolophoidea and c) Pteropodidae. Subsequently, we used $\omega$ estimates from each branch category to resolve the mixed branches into their functional and pseudogenic components. The times of divergence for each lineage were estimated in the previous section.

To estimate possible association of between the loss of the SWS1 opsin gene and the acquisition of primarily cave roosting habits in Pteropodids, we used the Brunch algorithm in Caper (Orme, 2012) and implemented in R (R Development Core Team, 2013) using a Pteropodid species tree. However, it is known that there are no accurate models to estimate nodal values of a categorical variable (Orme et al., 2012).

**Estimates of convergent evolution**

Alternative topology tests were used to assess the relative support for the convergent gene tree topology versus the species tree topology (where Yinpterochiroptera was constrained) using the functional *SWSI* dataset. Site-wise likelihood values were then used to conduct an approximately unbiased (AU) test (Shimodaira 2002) in Consel (Shimodaira and Hasegawa 2001). Methods described by Liu et al (2011) and Castoe et al (2009) were used to identify convergent amino acids in the echolocating bat species. Amino acid changes per branch were estimated based on the marginal ancestral sequence reconstructions generated by PAML through comparison of states at ancestral and descendant nodes in the Rhinolophoidea and Yangochiroptera. For amino acid sites at which changes occurred along the two compared branches, sites with different amino acids in the descendants were defined as divergent, and those with the same amino acid in the descendant were defined as convergent.
Acknowledgments

We thank Alisha Goodbla (SBES/UCD) for help in the lab and the ICHEC staff for bioinformatic support. We want also thank ICHEC, for the use of their cluster and Alexandre de Menezes for comments on early versions of this paper. Samples were generously donated by the Ambrose Monnel Cryo Collection (American Museum of Natural History), Field Museum of Natural History of Chicago, Royal Ontario Museum, Smithsonian National Museum of Natural, Senckenberg Museum of Natural History, Australian Museum of Natural History, Bob Swanepoel (National Institute for Communicable Diseases), Maarten Vonhof (Western Michigan University), Pavel Hulva (Charles University), Maria João Pereira (CESAM, Portugal) and Hugo Rebelo (CIBIO/UP, Portugal). BFS was supported by a Doctoral Grant SFRH/BD/36369/2007 from the FCT Portuguese Science and Technology Foundation and the work was supported by a SFI President of Ireland Young Research Award (PIYRA) 06/Y13/B932 awarded to ECT. BFS is currently supported by a EU H2020 Marie Skłodowska-Curie Global Fellowship GA703438. ECT and NMF are currently supported by a European Research Council Research Grant ERC-2012-StG311000 awarded to ECT. SJR is supported by European Research Council Grant ERC-2012-StGr310482. This work was also supported by the National Natural Science Foundation of China (grant 31570382) and the Ministry of Science and Technology of the People’s Republic of China (The National Key Research and Development Program, grant 2016YFC1200100) awarded to SZ.
References


Fig 1. Pseudogenization of SWS1 opsin gene (branches in red) mapped in the bat species tree.
Figure 1: Ancestral state reconstruction spectral tuning sites and inferred $\lambda_{\text{max}}$ of the MWS/LWS opsin gene for major bat clades. Inferred $\lambda_{\text{max}}$ values are shown for each species in addition to the observed changes in the five MWS/LWS spectral sites for each species. The spectral sites for the MWS/LWS opsin and estimated $\lambda_{\text{max}}$ for the Laurasiatherian ancestor is based in Lui et al., 2018.
Figure 2: Schematic depicting the shared and private indels leading to frame-shift mutations (insertion ▲, deletion▼), premature stop codons (◼) and loss-of-function mutations in the SWS1 opsin in 26 bats where this gene is non-functional.
Figure 3: Phylogram inferred from ML analysis of the SWS1 gene. Nodal support for both the ML and BA analysis are shown at nodes corresponding to major clades. Species using HDC echolocation are highlighted in blue. Red branches correspond to tips in which the SWS1 gene is non-functional. The Pteropodidae are highlighted in green, where brown is used to denote cave-dwelling taxa.
<table>
<thead>
<tr>
<th>Models</th>
<th>$\omega$ (dN/dS)</th>
<th>LnL</th>
<th>NP</th>
<th>Models Compared</th>
<th>$2\Delta$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MWS/LWS Opsin Gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. One-Ratio</td>
<td>$\omega = 0.0827$</td>
<td>3200.90</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Free-Ratio</td>
<td>Variable by branch</td>
<td>3136.94</td>
<td>173</td>
<td>1 vs. 2</td>
<td>127.9</td>
<td>0.002</td>
</tr>
<tr>
<td>3. M1a</td>
<td>$p_1= 0.9465; p_2= 0.0535$</td>
<td>3128.68</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. M2a</td>
<td>$p_1= 0.9465; p_2= 0.0281; p_3= 0.0254$</td>
<td>3128.68</td>
<td>91</td>
<td>3 vs. 4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5. M7Beta</td>
<td>$P = 0.1495; q = 1.2597; \omega = 0.1020$</td>
<td>3136.69</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. M8Beta</td>
<td>$p_0 = 0.9624; p_1 = 0.4673; q = 9.1088; (p_2 = 0.0376)$</td>
<td>3125.14</td>
<td>91</td>
<td>5 vs. 6</td>
<td>23.1</td>
<td>9.64^-06</td>
</tr>
<tr>
<td>7. Two-Ratio ((\omega_1) – SWS1 Functional; (\omega_2) – SWS1 Non-Functional)</td>
<td>$\omega_1= 0.1194 ; \omega_2= 0.0710$</td>
<td>3256.47</td>
<td>89</td>
<td>1 vs. 7</td>
<td>111.1</td>
<td>5.51^-26</td>
</tr>
<tr>
<td>8. Two-Ratio ((\omega_1) – HDC; (\omega_2) – LDC)</td>
<td>$\omega_1= 0.1504; \omega_2= 0.0736$</td>
<td>3248.89</td>
<td>89</td>
<td>1 vs. 8</td>
<td>95.9</td>
<td>1.20^-22</td>
</tr>
<tr>
<td>9. Two-Ratio ((\omega_1) – Yinpterochiroptera; (\omega_2) – Yangochiroptera)</td>
<td>$\omega_1= 0.0943; \omega_2= 0.0759$</td>
<td>3258.46</td>
<td>89</td>
<td>1 vs. 9</td>
<td>115.1</td>
<td>7.40^-27</td>
</tr>
<tr>
<td>10. Two-Ratio ((\omega_1) – Non-Echolocating; (\omega_2) – Echolocating)</td>
<td>$\omega_2= 0.0940; \omega_1= 0.0831$</td>
<td>3258.85</td>
<td>89</td>
<td>1 vs. 10</td>
<td>115.9</td>
<td>4.99^-27</td>
</tr>
<tr>
<td>11. Two-Ratio ((\omega_1) – Primarily Cave Roosters; (\omega_2) – Other)</td>
<td>$\omega_1= 0.081; \omega_2= 0.090$</td>
<td>3259.15</td>
<td>89</td>
<td>1 vs. 11</td>
<td>116.5</td>
<td>3.69^-27</td>
</tr>
</tbody>
</table>
### SWS1 Opsin Gene – Chiroptera

<table>
<thead>
<tr>
<th>Models</th>
<th>$\omega$ (dN/dS)</th>
<th>LnL$^1$</th>
<th>NP$^2$</th>
<th>Models Compared</th>
<th>$2\Delta^3$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Two-Ratio ($\omega_1$ – Insectivore/Carnivore; $\omega_2$ – Frugivore/Nectivore)</td>
<td>$\omega_1$ = 0.082; $\omega_2$ = 0.086</td>
<td>3258.93</td>
<td>89</td>
<td>1 vs. 12</td>
<td>116.06</td>
<td>4.61$^{-27}$</td>
</tr>
</tbody>
</table>

#### SWS1 Opsin Gene – Chiroptera

1. **One-Ratio**
   - $\omega = 0.1947$
   - LnL = 9713.82
   - NP = 220

2. **Free-Ratio**
   - Variable by branch
   - LnL = 9577.64
   - NP = 437

3. **M1a**
   - $p_1 = 0.8021; p_2 = 0.19790$
   - $\omega_1 = 0.1127; \omega_2 = 1$
   - LnL = 9529.13
   - NP = 221

4. **M2a**
   - $p_1 = 0.7978; p_2 = 0.1313; p_3 = 0.0709$
   - $\omega_1 = 0.1116\; \omega_1 = 1; \omega_2 = 1$
   - LnL = 9529.13
   - NP = 223

5. **M7Beta**
   - $p = 0.5843; q = 1.7967$
   - LnL = 9577.89
   - NP = 221

6. **M8Beta**
   - $p_0 = 0.9958; p_1 = 0.6198; q = 2.0180; (p_2 = 0.0042)$
   - $\omega = 4.0557$
   - LnL = 9565.24
   - NP = 222

7. **Two-Ratio ($\omega_1$ – Functional; $\omega_2$ – Non-Functional)**
   - $\omega_1 = 0.1111; \omega_2 = 0.5116$
   - LnL = 9775.90
   - NP = 221

8. **Two-Ratio ($\omega_1$ – HDC; $\omega_2$ – Others)**
   - $\omega_1 = 0.4646; \omega_2 = 0.1638$
   - LnL = 9844.01
   - NP = 221

9. **Two Ratio ($\omega_1$ – Yinpterochiroptera; $\omega_2$ – Yangochiroptera)**
   - $\omega_1 = 0.2448; \omega_2 = 0.1334$
   - LnL = 9855.28
   - NP = 221

10. **Two-Ratio ($\omega_1$ – Echolocating; $\omega_2$ – Non-Echolocating)**
    - $\omega_1 = 0.1836; \omega_2 = 0.2047$
    - LnL = 9868.35
    - NP = 221
<table>
<thead>
<tr>
<th>Models</th>
<th>$\omega$ (dN/dS)</th>
<th>LnL$^1$</th>
<th>NP$^2$</th>
<th>Models Compared</th>
<th>$2\Delta^3$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Two-Ratio ((\omega_1) – Primarily Cave Roosters; (\omega_2) – Other)</td>
<td>(\omega_1) = 0.2957; (\omega_2) = 0.1219</td>
<td>9834.15</td>
<td>221</td>
<td>1 vs 11</td>
<td>42.24</td>
<td>8.07$^{11}$</td>
</tr>
<tr>
<td>10. Two-Ratio ((\omega_1) – Insectivore/Carnivore; (\omega_2) – Frugivore/Nectivore)</td>
<td>(\omega_1) = 0.1845; (\omega_2) = 0.2014</td>
<td>9868.51</td>
<td>221</td>
<td>1 vs 12</td>
<td>26.48</td>
<td>2.66$^7$</td>
</tr>
</tbody>
</table>

SWS1 Opsin Gene – Pteropodidae

| 1. One-Ratio            | \(\omega\) = 0.2247 | 4918.21 | 92     |
| 2. Free-Ratio           | Variable by branch  | 4823.60 | 181    | 1 vs 2        | 189.2       | 6.50$^{08}$ |
| 3. M1a                  | \(p_1\) = 0.8135 \(p_2\) = 0.1865 \(\omega_1\) = 0.1086; \(\omega_2\) = 1.0000 | 4859.52 | 93     |
| 4. M2a                  | \(p_1\) = 0.8135 \(p_2\) = 0.0489; \(p_3\) = 0.1376 \(\omega_1\) = 0.1086; \(\omega_2\) = 1; \(\omega_3\) = 1 | 4859.91 | 95     | 3 vs 4        | 0.78        | 0.677     |
| 5. M7Beta               | \(p\) = 0.38351 \(q\) = 1.13584 | 4859.97 | 93     |
| 6. M8Beta               | \(p_0\) = 0.9499; \(p_1\) = 0.5836; \(q\) = 2.3549 \(p_2\) = 0.0501; \(\omega\) = 1.601 | 4855.76 | 95     | 5 vs 6        | 8.42        | 0.015     |
| 7. Two-Ratio (\(\omega_1\) – Non-Functional; \(\omega_2\) – Functional) | \(\omega_1\) = 0.5173; \(\omega_2\) = 0.1154 | 4941.03 | 93     | 1 vs 7        | 45.6        | 1.41$^{11}$ |
| 8. Two-Ratio (\(\omega_1\) – Echolocating; \(\omega_2\) – Non-Echolocating) | \(\omega_1\) = 0.6660; \(\omega_2\) = 0.1684 | 4962.70 | 93     | 1 vs 8        | 89          | 3.95$^{21}$ |
| 9. Two-Ratio (\(\omega_1\) – Cave; \(\omega_2\) – Tree) | \(\omega_1\) = 0.5384; \(\omega_2\) = 0.1132 | 4937.97 | 93     | 1 vs 9        | 37.6        | 8.69$^{10}$ |

Table 1. \(\omega\) values and Likelihood Ratio Tests (LRT) of selective pressures in the bat MWS/LWS opsin gene and on the SWS1 opsin gene in Chiroptera and Pteropodidae. $^1$ The log likelihood values. $^2$ Number of parameters. $^3$ \(2\Delta = 2 \times (\text{LnL}_1 - \text{LnL}_2)\).