Molecular Phylogenetics and Evolution 173 (2022) 107522

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Parallel decay of vision genes in subterranean water beetles

Barbara L. Langille [a,*], Simon M. Tierney [a,b], Terry Bertozzi [a,c], Perry G. Beasley-Hall [a], Tessa M. Bradford [a,c], Erinn P. Fagan-Jeffries [a,c], Josephine Hyde [a,d], Remko Leijs [c], Matthew Richardson [a], Kathleen M. Saint [a,c], Danielle N. Stringer [a,c], Adrián Villastrigo [c,e], William F. Humphreys [f,g], Andrew D. Austin [a,c], Steven J.B. Cooper [a,c]

[a] Australian Centre for Evolutionary Biology and Biodiversity, Department of Ecology and Evolution, School of Biological Sciences, University of Adelaide, South Australia 5005, Australia
[b] Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia
[c] Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia
[d] Western Australia Department of Biodiversity Conservation and Attractions, Kensington, WA 6151, Australia
[e] Institute of Evolutionary Biology, Passeig Marítim de la Barceloneta, 37-49, 08003, Spain
[f] Western Australian Museum, Locked Bag 40, Welshpool DC, WA 6986, Australia
[g] School of Animal Biology, University of Western Australia, Nedlands, Western Australia, Australia

ARTICLE INFO

Keywords:
Regressive evolution
Pseudogenisation
Neutral evolution theory
Dytiscidae
Subterranean beetles
Vision genes

ABSTRACT

In the framework of neutral theory of molecular evolution, genes specific to the development and function of eyes in subterranean animals living in permanent darkness are expected to evolve by relaxed selection, ultimately becoming pseudogenes. However, definitive empirical evidence for the role of neutral processes in the loss of vision over evolutionary time remains controversial. In previous studies, we characterized an assemblage of independently-evolved water beetle (Dytiscidae) species from a subterranean archipelago in Western Australia, where parallel vision and eye loss have occurred. Using a combination of transcriptomics and exon capture, we present evidence of parallel coding sequence decay, resulting from the accumulation of frameshift mutations and premature stop codons, in eight phototransduction genes (arrestins, opsins, ninaC and transient receptor potential channel genes) in 32 subterranean species in contrast to surface species, where these genes have open reading frames. Our results provide strong evidence to support neutral evolutionary processes as a major contributing factor to the loss of phototransduction genes in subterranean animals, with the ultimate fate being the irreversible loss of a light detection system.

1. Introduction

Regressive evolution, a process describing the evolutionary loss of phenotypic traits such as limbs in snakes, wings in birds and insects, and eyes and pigment in subterranean animals, has long intrigued evolutionary biologists (Jeffery, 2009; Wilkens and Strecker, 2017). Charles Darwin struggled with the idea of regressed features, particularly in subterranean animals and reasoned that; "*As it is difficult to imagine that eyes, though useless, could be in any way injurious to animals living in darkness, their loss may be attributed to disuse*…" (On the Origin of Species, 1872p. 139); an unexpected statement largely accepting Lamarckian theory. Incredibly, >160 years later, there is still considerable debate surrounding the evolutionary mode of trait loss, particularly eye/vision loss in subterranean animals, with major theories largely based on neo-Darwinian selection or the neutral theory of molecular evolution (Kimura, 1983; Culver and Wilkens, 2000; Jeffery, 2009; Rétaux and Casane, 2013; Wilkens, 2020).

Selectionist arguments for regressive evolution propose that an advantage is gained by the subterranean species due to the loss of the functional trait and that selection can manifest in positive or negative, direct or indirect manners (Breder, 1942; Jeffery et al., 2000; Yamamoto et al., 2003; Menuet et al., 2007; Protas et al., 2007; Jeffery, 2009). For example, direct negative selection results from the unneeded trait (e.g., eyes in subterranean animals) causing lower fitness to the individual (e.g., by increasing the chance of injury in the dark). Positive selection results in the deletion of an unneeded trait by either direct selection (e.g., reducing energy expenditure in a resource poor environment; Moran et al., 2015) or indirect selection, with positive selection to enhance the

* Corresponding author.
E-mail address: barbara.langille@adelaide.edu.au (B.L. Langille).

https://doi.org/10.1016/j.ympev.2022.107522
Received 5 January 2022; Received in revised form 19 April 2022; Accepted 20 April 2022
Available online 17 May 2022

fitness of individuals (e.g., improvement of sensory systems in an aphotic environment) indirectly leading to the loss of a trait (e.g. via pleiotropy or genetic hitchhiking; Yamamoto et al., 2003; Espinasa and Espinasa, 2008; Jeffery, 2009). Conversely, neutral theory of molecular evolution argues that traits can be lost through pseudogenisation of functional genes that are specific to the trait via accumulated random mutations and genetic drift in the associated genes, as there is no longer positive or purifying selection acting upon the gene to maintain the encoded trait (Kimura, 1968; Yokoyama et al., 1995; Wilkens, 2010, 2020). Despite widespread acceptance that many parts of the genome are under the influence of such neutral processes (Lynch, 2007; Ho et al., 2017; Kumar and Patel, 2018; Yoder et al., 2018; Zhang, 2018), selectionist interpretations remain the most strongly promoted hypotheses to account for eye regression in subterranean animals (Protas et al., 2007; Jeffery, 2009; Moran et al., 2015; Stern and Crandall, 2018; but see counter-arguments by Wilkens and Strecker, 2017; Wilkens, 2020). However, a major question is what is the evolutionary fate of genes involved in vision in subterranean animals? Are they pleiotropic and maintained by purifying selection for different essential traits, or co-opted into other signalling pathways (e.g., vibration detection) as an adaptive response to living in an aphotic environment, or do they evolve by neutral evolution and ultimately become pseudogenes via genetic drift?

The majority of eye regression research has focused upon the model cavefish *Astyanax mexicanus*, where a wealth of genetic studies have provided highly informative insights into the process of eye regression, but have also shown that genes involved in phototransduction were largely intact and appeared to be still functional (Langecker et al., 1993; Yokoyama et al., 1995; Wilkens and Strecker, 2003; Jeffery, 2009; Fumey et al., 2018; Herman et al., 2018). A study of rhodopsin genes in subterranean crayfish (Crandall and Hillis, 1997), and phototransduction genes in the troglobiont beetle *Ptomaphagus hirtus* (Friedrich et al., 2011) have also shown these 'eye' genes to be fully functional. The latter species was shown to be phototactic in its behaviour, and, therefore, maintained an intact phototransduction system (Friedrich et al., 2011). Crandall and Hillis (1997) suggested that *rhodopsin* may have an unrecognised function in subterranean crayfish, associated with circadian rhythms. An alternative explanation, however, is that these studies of subterranean crayfish, *A. mexicanus* and other cave fish species (see Policarpo et al., 2021) may have been unable to detect the actual molecular evolutionary forces operating on these 'eye' genes due to insufficient time for the accumulation and fixation of deleterious or amino acid changing mutations, indicative of a relaxation of selection (Leys et al., 2005; Niemiller et al., 2013; Tierney et al., 2018). In support of this hypothesis, independent loss-of-function mutations were found in the interphotoreceptor retinoid binding protein gene of marsupial moles (Springer et al., 1997), cone and rod-specific genes of several mole rat species (Emerling and Springer, 2014), the cinnabar eye pigment gene of subterranean water beetles (Leys et al., 2005), and rhodopsin genes of multiple ancient (10.3 My) amblyopsid cavefish lineages (Niemiller et al., 2013). A recent comparative genome study by Policarpo et al. (2021) also revealed numerous cases of pseudogenisation of eye-specific genes of *Lucifuga* cavefishes that are millions of years old. However, these studies were limited by either being focused on a single candidate gene (e.g. *rhodopsin* in Niemiller et al., 2013, and *cinnabar* in Leys et al., 2005), or on genomic analyses where very few loss-of-function mutations on genes involved with vision have been found shared among different species. In a study on *Lucifuga* cavefishes, Policarpo, et al. (2021) found ~ 28% of genes involved in vision (20 genes) have loss-of-function mutations, however, only ~ 5% of genes are shared between different species (4 genes). These studies, therefore, do not yet have the statistical power in comparative analyses among species to link the aphotic environment to the loss of these specific genes.

Here we investigate a relatively old and species rich subterranean invertebrate system comprised of ~ 100 described species with an estimated origin and diversification of ~ 3–5 Mya: subterranean beetles (Dytiscidae) from calcrete aquifers in Western Australia (Cooper et al., 2002; Leys et al., 2003; Leijs et al., 2012; Watts and Humphreys, 2009). These calcrete aquifers represent closed environmental systems, with>200 known calcrete aquifer bodies resembling a subterranean archipelago (Cooper et al., 2002). Each calcrete aquifer hosts a unique suite of aquatic subterranean taxa (stygobionts), including between one and three diving beetle species (Leys et al., 2003; Leijs et al., 2012). The majority (71%) of species have independently evolved typical cave troglomorphies (i.e. complete loss of eyes, wings and pigment, with the exception of *Limbodessus microocular* and *L. microommatoion*, which appear to have rudimentary eyes), which collectively provide an unrivalled opportunity for comparative genomic-scale analyses of regressive trait evolution (Tierney et al., 2018). The remaining 29% of the species which have also evolved typical cave troglomorphies, are sympatric sister pairs or triplets, of which each lineage appears to have independently evolved from a single subterranean (stygobiotic) ancestor (Leijs et al., 2012; Langille et al., 2021).

Our study is the first of its kind to investigate the molecular evolution of multiple phototransduction genes from a wide sampling of closely related surface and subterranean species to test whether they evolve by neutral molecular evolution in an aphotic environment. For phototransduction genes experiencing neutral evolutionary processes, our predictions are: i) genes should show evidence of loss-of-function mutations (frameshift mutations and premature stop codons) in the encoded proteins, as previously found for the eye pigment gene *cinnabar* (Leys et al., 2005), and/or increases in the rate of evolution of amino acid changes; ii) there should be evidence for neutral processes in phototransduction genes from phylogenetically (and geographically) independent subterranean lineages; iii) neutral processes should only occur in functional genes that are specific to the regressed phenotypic trait (i.e. pleiotropic genes involved in other essential functions should remain under purifying selection); and 4) more recently evolved subterranean species may show relatively less or no evidence for loss-of-function mutations in their phototransduction genes (*cf*. older subterranean species) due to an insufficient generational time for mutations to accumulate and become fixed by genetic drift.

2. Methods

2.1. Calcrete aquifer sampling

A total of 31 subterranean diving beetle species from the genera *Limbodessus* and *Paroster* were collected from 20 calcrete aquifers in central Western Australia (Fig. 1a) utilising pre-drilled bore holes ~ 10 m deep. The species were chosen to obtain a good representation of independently-evolved subterranean species across the phylogenies of *Limbodessus* and *Paroster*. Within those 31 species, six sympatric sister pairs or triplets of species were included, as each of these most likely have evolved from a stygobiotic common ancestor and represent evolution entirely underground (Langille et al., 2021). Six surface species (*A. bistrigatus*, *Limbodessus compactus* (Clark, 1862), *L. rivulus* (Larson, 1994), *Neobidessodes mjobergi* (Zimmermann, 1922), *P. nigroadumbratus,* and *P. gibbi* (Watts, 1978)), that are closely related to the subterranean species, were sampled from surface pools (SI2). Collected specimens were preserved in RNAlater, 100% ethanol or snap frozen in liquid nitrogen.

2.2. Transcriptome analyses

We used transcriptome data for five species (two surface species: *A. bistrigatus* and *P. nigroadumbratus*, and three subterranean species: *P. macrosturtensis, N. gutteridgei* and *L. palmulaoides*) from Tierney et al. (2015). Transcriptome data were also derived from an additional five subterranean species using single whole-body RNA extractions. RNA (and DNA) was extracted using an AllPrep DNA/RNA Micro kit (Qiagen,

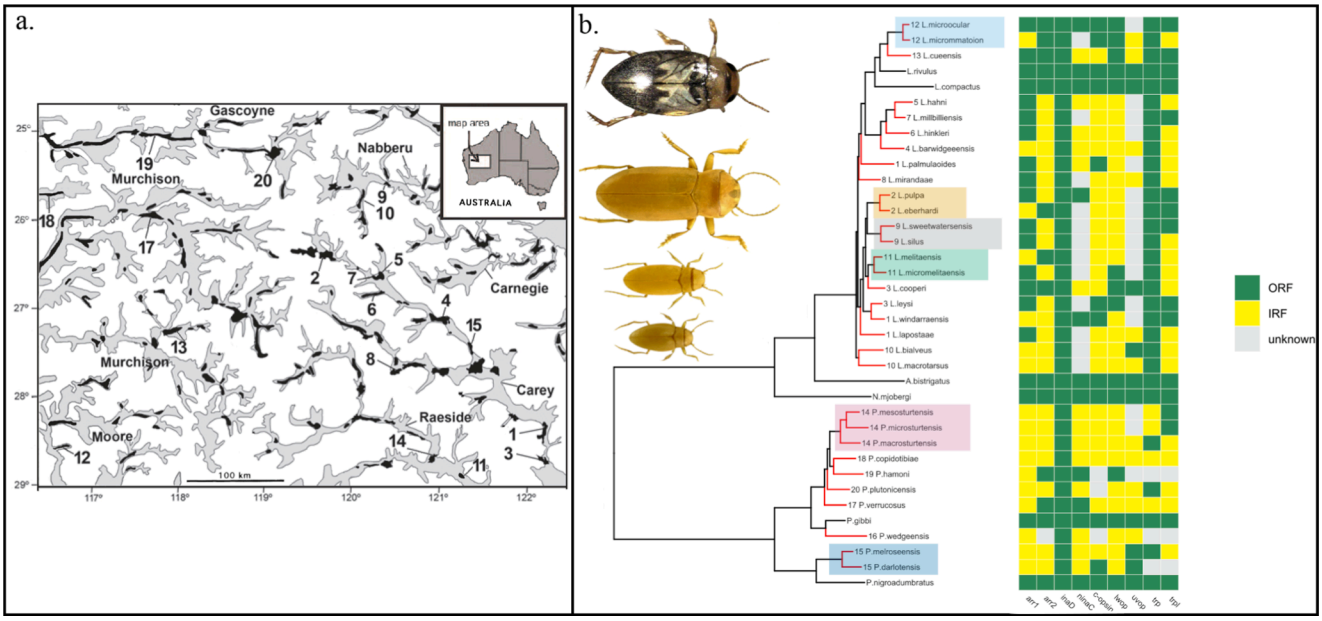

Fig. 1. a. Map of the Yilgarn Region based on a previously published map from Cooper at al. 2008 of Western Australia showing locations of calcretes (black) sampled for this study. Numbers correspond to names in SI2. Central Mt. Wedge is found well outside of this map at the coordinates 22.85° S, 131.82° E. **b.** Phylogenetic tree of dytiscid beetles used in the current study based on maximum likelihood analyses of *COI* and nuclear data, and heatmap displaying details of the nine phototransduction genes. Red branches in the phylogenetic tree represent subterranean lineages, black branches represent surface lineages. Shaded boxes indicate sister species found in the same calcrete aquifers, with different colors for each sister group. Numbers in front of species names correspond to calcretes on the map. Heat map: yellow (IRF interrupted reading frame) indicates the presence of deleterious mutations (frameshift and stop mutations) indicative of pseudogenisation; green indicates intact genes with open reading frames (ORFs); grey indicates genes that failed to be captured and sequenced. Example beetles are given to highlight the differences between subterranean and surface species. From top down: *Allodessus bistrigatus* (surface species), *Paroster macrosturtensis*, *P. mesosturtensis*, and *P. microsturtensis*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Chadstone, VIC, AUS), initially homogenizing each beetle in lysis buffer containing β-mercaptoethanol plus a steel bead on a Mixer Mill (Qiagen). Species identification was checked by PCR amplifying a *COI* molecular barcode from the DNA extract (primers M202/M70; Leys et al., 2003), followed by BLASTn analysis to compare the sequence to the NCBI nucleotide database. Stranded Illumina sequencing libraries were prepared with the NuGEN Universal Plus mRNA-Seq kit (Tecan, Redwood City, CA, USA). Input concentrations ranged from 15 ng to 93 ng. These sequencing libraries were sequenced on two lanes of a NovaSeq 6000 SP at the Australian Genomics Research Facility Melbourne, generating 150 bp paired-end reads. A library preparation blank was included in the sequencing run.

The short read RNA-seq data (308.82 Gb) were examined for quality, overrepresented sequences and adapter contamination in FASTQC, and visualized in R v.3.6.2 (R core team, 2020) using the ngsReports package (Ward et al., 2019). Adapters were trimmed from the right hand side of each read and then paired reads were merged using BBDuk and BBmerge (Bushnell, 2015) respectively. Over 82% of the reads could be joined and average sequence length was between 155 and 162 bp. Merged reads were assembled using the Trinity RNA-Seq De novo Assembler (https://github.com/trinityrnaseq/trinityrnaseq/wiki; Trinity-v2.8.5), with the sense strand set to forward.

De novo assemblies of putative transcripts of all ten species were used to identify and annotate 18 phototransduction genes (as per methods in Tierney et al., 2015). Briefly, we used the *Tribolium castaneum* isoform with the longest coding sequence available on GenBank for each phototransduction gene of interest as a query sequence to conduct tBLASTx searches of each transcriptome. The dytiscid sequence corresponding to the top hit (by E-value) from these searches was retained and validated against the NCBI's GenBank database using BLASTn, with matches considered "good" at > 50% coverage and > 70% sequence identity to a sequence corresponding to the gene of interest from an insect species. Sequences were discarded if their top GenBank hit in BLASTn corresponded to a non-insect species or an incorrect gene of interest. Eighteen photoreceptor genes were identified using the above methods (Table 1). To detect the possible presence of transcribed pseudogenes, the transcriptome assemblies for subterranean species were further scrutinised by BLASTn analyses in Geneious (version 10.2.4) using query sequences derived from surface dytiscid orthologues for each of the 18 photoreceptor genes. All transcript sequences were then further assessed for the presence of full or partial open reading frames (herein referred collectively as ORF) using Geneious. Here, we assume a full ORF encodes a functional protein, and where there was no evidence of a frameshift mutation or premature stop codon in a partially sequenced gene, we have conservatively called the gene functional, although additional sequence data in the future may show it to be non-functional. Cleaned reads were mapped back to each sequence that showed evidence of pseudogenisation (frameshift mutations and stop codons) using BWA v.0.7.12 (Li and Durbin, 2009). The resultant BAM files were examined in IGV v.2.3.92 (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) to confirm each frameshift or nonsense mutation was not the result of assembly errors caused by low/excess sequence coverage, poor consensus calling or incorrect read joining. We have referred to frameshift mutations or the presence of premature stop codons as "loss-of-function" mutations in line with other studies (e.g. Policarpo et al., 2021), however, it was beyond the scope of the study and the limits of the study system to experimentally test the functionality of the proteins.

Eight genes showed the presence of transcripts in surface species that were either absent or contained frameshift mutations or stop codons in the subterranean species (Table 1). Given that the absence of a transcript may not necessarily indicate the gene is absent or non-functional we used targeted exon capture to enrich and sequence these genes from genomic DNA. We also included *inactivation no afterpotential D* (*inaD*) as a control for a gene likely to be evolving under purifying selection due to its known pleiotropic role in *Drosophila* (from flybase; SI1). RNA baits were designed from the orthologous transcript sequences of the surface

Table 1
Photoreceptor genes detected in transcriptomes retrieved from Tierney et al. (2015) and newly obtained data. Green cells indicate the transcript had an open reading frame, yellow cells indicate the presence of a disrupted open reading frame due to frameshift mutations and/or stop codons, and white cells with a dash indicate the transcript was not detected. Additionally, '*' indicates a partial transcript detected and '‡' indicates the transcript likely contains retained intronic sequence.

	Surface species		Subterranean species							
Photoreceptor genes	*A.bis*	*P.nig*	*L.cue*	*L.ebe*	*L.hin*	*L.pal*	*L. pul*	*N.gut*	*P.mac*	*P.mes*
Arrestin 1			*	*‡	*	-	*	*		-
Arrestin 2			-	-	-	-	-	-	*	*
Chaoptin	*	-	-	*	-	*	-	*	*	-
G protein gamma 30A				*		*	-		*	*
G protein alpha 49B							*		*	*
G protein beta 76C			*			*	*			*
G protein-coupled receptor kinase 1		*	*	*	-		-	*		*
inaC			-	-	*		-		*	*
inaD	*	*	*	*	-	*	-		*	-
ninaC	*	*	-	-	-	-	-	-	*‡	-
No receptor potential A type 1					*	*	*		*	-
Opsin c-opsin			-	-	-	-	-	*	-	-
Opsin ultraviolet			-	-	-	-	-	-	-	-
Opsin long-wavelength	*	*	-	-	-	-	-	-	-	-
Rab-protein 6a					*		*			*
Spacemaker	-	-	-	-	-	*	-	*	-	-
Transient receptor potential	*	*	-	-	-	-	*‡	-	-	-
Transient receptor potential-like	*	*	-	-	-	-	-	-	*	-

A.bis (Allodessus bistrigatus), P.nig (Paroster nigroadumbratus), L.cue (Limbodessus cueensis), L.ebe (L. eberhardi), L.hin (L. hinkleri), L.pal (L. palmulaoides), L.pul (L. pulpa), N. gut (Neobidessodes gutteridgei), P. mac (Paroster macrosturtensis), P. mes (Paroster mesosturtensis).

species *A. bistrigatus* and *P. nigroadumbratus* and synthesized by Arbor Biosciences (Ann Arbor, MI).

2.3. Library preparation

We extracted DNA from whole beetles following the Gentra whole genomic DNA from tissue protocol (Gentra Systems, Inc.) with the following modifications: 0.5 μL of glycogen was used instead of 2 μL and precipitated DNA was centrifuged for 15 min instead of 10 min at 4 °C. We measured DNA concentrations using a Qubit 2.0 fluorometer (Life Technologies) using a dsDNA quantification kit, and we pooled multiple samples when<1 ng/μL. For *P. nigroadumbratus*, we used previously extracted DNA, which had been stored at −80 °C. Starting material for sonication ranged from 100 ng to 500 ng. We sheared DNA to an average fragment distribution of 400–600 bp using an on/off setting of 30/30 for 2 cycles, in a Diagenode Bioruptor sonicator. DNA fragment sizes were subsequently verified using a High Sensitivity D1000 screen tape following the accompanying kit protocol for an Agilent 4200 TapeStation System.

Following sonication, we constructed sequencing libraries using the Meyer and Kircher protocol (Meyer and Kircher, 2010) using double indexing primers (Hugall et al., 2015; Glenn et al., 2016). We assessed the success of library preparation by measuring DNA concentration using a standard qPCR run in a LightCycler 96 Real-Time PCR System, following the manufacturer's protocol. All samples underwent a 0.9–1.8X Ampure XP bead clean-up after each step and as a final step following Meyer and Kircher (2010).

2.4. Hybridization of baits to libraries and sequencing

Prior to capture we divided the target baits into ¼ capture reactions. We enriched each species in its own ¼ capture or added a maximum of two species together in ¼ capture to ensure maximum targeting of each species. We performed the enrichment following the Arbor Biosciences myBaits user manual v2. The enrichment was performed at 65 °C for a period of 44 h for the first enrichment which contained *Limbodessus palmulaoides*, and 18 h for all other enrichments, as we determined that 18 h was sufficient for hybridisation.

We verified enrichment success with qPCR as previously mentioned and used a TapeStation for visualization of fragments to confirm they were within the correct size range for sequencing. Size selection was performed using Ampure XP beads in order to reduce the small bp size fragments (Meyer and Kircher, 2010). All samples were pooled in equal concentrations and subsequently concentrated to 30 μL. The first pool contained six pooled samples (four different species), the second contained eight pooled samples (eight different species), the third contained 22 pooled samples (18 different species) and the final pool contained 15 pooled samples (12 different species). Each pool was run on its own lane on the Illumina MiSeq platform (AGRF facility in Adelaide, Australia), obtaining 300 bp paired-end reads for captures one, two and four, and 150 bp paired-end reads for capture three. We chose the smaller read return for capture three because starting material was fragmented into sizes too low for 300 bp paired end reads.

2.5. Bioinformatics

2.5.1. Quality assessment and mapping

Raw sequencing reads for each species were assessed using FASTQC v.0.11.3 (Babraham Institute). Poor quality bases and Illumina adapter sequences were then trimmed using BBDuk v.2 with the following parameters: literal = AGATCGGAAGAGCAC,AGATCGGAAGAGCGT ktrim = r k = 15 mink = 15 hdist = 0 tbo qtrim = rl trimq = 20 minlength = 30 threads = 10. Cleaned reads were mapped to transcriptome generated genes of either *A. bistrigatus* (for *Limbodessus* species), or

P. nigroadumbratus (for *Paroster* species), using BWA v.0.7.12 with default parameters and the resulting alignments converted to BAM format, only retaining mapped reads, using SAMTools v.1.3.1 (Li et al,. 2009). Trimming and mapping steps were implemented in a Unix shell script (https://github.com/babslangille/NGS-workthrough/blob/main/indexing-and-mapping) on a 12-core virtual machine on the NeCTAR research cloud (National Research Infrastructure for Australia) under an Ubuntu 16.04 LTS image. BAM files were viewed in IGV v.2.3.92 to visually identify exon-intron junctions (i.e. presence of softclips to the left in some sequences and to the right in others of a particular position), which were manually separated. Subsequently, each intron position was padded with 100 'Ns', and genes were subsequently re-mapped to the new reference sequences which allowed reads to span intron/exon boundaries.

Cleaned reads for each species were *de novo* assembled using four different assemblers to optimize successful coverage of exons across genes and to verify that the mapping approach detected all sequences: IDBA v.1.1.1 (Peng et al., 2010), RAY (Boisvert et al., 2010), SPAdes v.3.13.0 (Bankevich et al., 2012) and Celera v.8.3 (WGS; Denisov et al., 2008). All assemblies were viewed in Geneious (Kearse et al., 2012) and subsequently compared to a custom BLASTn database containing the photoreceptor genes from the transcriptome data of the two surface species. Geneious was also used to map cleaned reads to the same reference sequences as mentioned above (the reference was chosen based on relatedness to species being mapped and was either *L. compactus*, *A. bistrigatus*, or *P. nigroadumbratus*), as a comparison to our BAM files, in order to verify mapping quality, but also to extend final sequences if possible. We wanted to be thorough when verifying gene sequences to eliminate all possibility of base errors in assemblies that would give the appearance of a pseudogene (Sharma et al., 2020). The general coverage of exons was higher in the first and second pool as less samples were pooled, resulting in hundreds to thousands of per base sequencing depth. The general coverage was consistently lower in the remaining two pools, as we would expect due to the pooling of samples for sequencing, resulting in less than a hundred sequence coverage. The coverage was consistently lower near the edges of the exons, due to an issue termed the 'edge effect' (Bi et al., 2012), consistent with other studies (Bragg et al., 2016; Portik et al., 2016; Puritz and Lotterhos, 2018).

2.5.2. Orthology of genes

We used the BLASTn feature within GenBank to verify that the exon sequences we captured were orthologous to the target genes. An orthologous match was considered positive when identities were greater than or equal to 70% (Tommaso et al., 2011) and the best BLAST hits were to verified orthologous genes from coleopteran taxa (Ensembl Metazoa; https://metazoa.ensembl.org/). Exon sequences were concatenated and then the gene sequences were further scrutinised for their orthology to target genes using phylogenetic analyses. Sequences of each gene (*arr1*, *arr2*, *c-opsin*, *lwop*, *uvop*, *inaD*, *ninaC*, *trp*, and *trpl*) were aligned independently with CLUSTALW, or using the global pairwise alignment function in Geneious, and manually adjusted by eye. We conducted maximum likelihood (ML) analyses for each different gene alignment using RAxML-HPC Blackbox v.8.2.12 (Stamatakis, 2014) through the CIPRES Science Gateway (Miller et al., 2010), using default parameters (GTR + G model of evolution), with orthologous reference sequences obtained from GenBank (SI3). Final gene trees were viewed and edited in FigTree v.1.4.3 (Rambaut, 2012). To produce a comparative independently-derived phylogenetic tree, we also carried out a maximum likelihood phylogenetic analysis, as given above, of the 37 dytiscid species used in exon capture analyses, utilising a concatenated alignment comprising cytochrome *c* oxidase subunit I gene (*COI*) and the nuclear genes wingless (*wg*), topoisomerase (*TOP1*) and arginine kinase (*argk*). We avoided using the entire gene set (i.e. arrestins, opsins and trps) as these genes may incorrectly skew the topology due to pseudogenisation. Methods for PCR-amplification and sequencing of these genes are given in Langille et al. (2021), with the exception that full length *COI* sequence data were derived from bi-catch assemblies of exon capture sequence data from 21 species. GenBank accession numbers for new sequences are ON099067 - ON099397, and ON064371 - ON064383.

2.5.3. Pseudogene assessment

All sequences of each phototransduction gene were translated and visually assessed for ORFs to determine whether the sequences had frameshift mutations (insertions or deletions that result in a codon frameshift) and premature stop codons indicative of pseudogenisation. We assessed the read quality of these sites for sequencing or assembly errors using methods given above for the transcriptome analyses, or by mapping raw reads onto the sites using the Geneious Map to Reference function, with medium sensitivity – fast, and three iterations. Mutations were considered verified if they were supported by at least five independent (non-PCR duplicate) reads.

2.5.4. Tests of selection

We would expect genes evolving under neutral evolution to show increased rates of amino acid changing (nonsynonymous) substitutions (dN), with the ratio (ω)of nonsynonymous to synonymous substitutions (d*S*) approaching one (i.e. $\omega = dN/dS = 1$), whereas under purifying selection, ω values would be significantly less than one. It is usually unclear at which exact point along a terminal branch leading to a subterranean species, that a surface species first colonised the calcrete and evolved in darkness. Therefore, these branches involve periods of evolution on the surface, with vision genes under purifying selection ($\omega < 1$), and a period underground with vision genes potentially evolving neutrally ($\omega = 1$) Hence, overall ω will on average be significantly less than one. To overcome this problem, we utilised a unique feature of the subterranean dytiscid system: clades containing sympatric sister species, for which there is strong evidence that these species evolved from subterranean ancestors (Leijs et al., 2012; Langille et al., 2021). In these cases, the entire branch for these taxa would not comprise any period of evolution on the surface (i.e. $\omega = 1$ is the expectation for neutrally evolving genes).

We employed Datamonkey v.2.0 (datamonkey.org; Weaver et al., 2018) for phylogenetic hypothesis testing. We computed the fixed effect likelihood (FEL; Pond and Frost, 2005), which is a site-specific method that calculates ω independently at each codon position. We also carried out branch by branch analyses (RELAX in HyPhy; Wertheim et al., 2015b) comparing the terminal branches of surface species to subterranean sympatric sister species terminal branches. Under this model, $k > 1$ is indicative of purifying selection, while $k < 1$ is indicative of relaxed selection. RELAX requires proper codon structure; therefore, all insertions or deletions that were not a multiple of three were removed to acquire the correct reading frame, an 'N' was added to the third position of stop codons, and deletions were filled with Ns until the expected reading frame was obtained. Both the null and alternative model estimate ω for each branch of the tree; however, the null model does not transform the branches, whereas the alternative model estimates *k* which transforms ω for two different branch classes.

3. Results

3.1. Transcriptome analyses

Transcriptome data from two surface species (*Allodessus bistrigatus* (Clark 1862), *Paroster nigroadumbratus* (Clark 1862)) and eight subterranean species (*Limbodessus cueensis* (Watts and Humphreys 2000), *Limbodessus eberhardi* (Watts and Humphreys 1999), *Limbodessus hinkleri* (Watts and Humphreys 2000), *Limbodessus palmulaoides* (Humphreys 2006), *Limbodessus pulpa* (Watts and Humphreys 1999), *Neobidessoides gutteridgei* (Watts and Humphreys 2003), *Paroster macrosturtensis* (Watts and Humphreys 2006), and *Paroster mesosturtensis* (Watts and

Humphreys 2006)), were used to identify 18 phototransduction genes (SI1; Tierney et al., 2015). A comparison of surface and subterranean species revealed that eight of these genes; *arrestin 1* (*arr1*), *arrestin 2* (*arr2*), *neither inactivation nor afterpotential C* (*ninaC*), invertebrate *c-opsin* (*c-opsin*), *long wavelength opsin* (*lwop*), *ultraviolet opsin* (*uvop*), *transient receptor potential* (*trp*), and *transient receptor potential-like* (*trpl*)), showed either (i) evidence for pseudogenisation via frameshifts leading to aberrant stop codons in the encoded sequence, or (ii) no detectable transcription in subterranean species (Table 1). The remaining 10 phototransduction genes all had open reading frames (ORFs) in both subterranean and surface species, each encoding highly conserved amino acid sequences.

3.2. Targeted capture of candidate phototransduction genes

We further evaluated the nucleotide and amino acid sequences of the phototransduction genes displaying evidence of pseudogenisation, or an absence of transcription in subterranean species, using targeted exon capture from nuclear DNA (Gnirke et al., 2009) with RNA baits designed from the gene sequences of surface species *A. bistrigatus* and *P. nigroadumbratus* (Tierney et al., 2015). We also included *inactivation no afterpotential D* (*inaD*) as a positive control of a gene we predicted to be fully functional based on: (i) the finding of ORFs for five subterranean species and (ii) due to its pleiotropic role in sound detection (SI1, Senthilan et al., 2012). We obtained exon sequence data for the above nine genes for six surface species and 31 subterranean species (21 *Limbodessus* and 10 *Paroster*), including five groups of sympatric sister species from 20 calcrete aquifers (Fig. 1; SI2). Capture success was high in the surface species, but it was variable in the subterranean species (SI2). Two genes (*ninaC* and *UV opsin*) had no captured sequences in 12 and 16 of the 31 subterranean species respectively. However, the levels of missing data were not consistent across all nine genes for one species, or across one gene for all 31 species. These results suggest that the capture success was most likely influenced by sequence divergence from surface ancestors, rather than the quality of the RNA-baits, DNA samples, or the sequencing depth.

3.3. Pseudogene assessment

When we compared the nine genes (SI3; SI4) from the subterranean species to the orthologous genes in the surface species, we detected mutations, indicative of pseudogenisation (i.e. insertions and deletions, leading to frameshifts or nonsense mutations), which resulted in a truncation of the encoded proteins in eight of the genes: *arr1, arr2, ninaC, c-opsin, lwop, uvop, trp*, and *trpl* (Fig. 1b; SI5 for alignments and SI6 for table of deleterious mutations associated with each alignment). In general, each pseudogene from each subterranean species contained its own unique combination of protein-altering mutations (see SI6), with pseudogenisation detected as follows: *arr1* (17 pseudogenes from 31 subterranean species where exon capture was successful), *arr2* (22/30), *ninaC* (14/19), *c-opsin* (22/28), *lwop* (24/31), *uvop* (11/15), *trp* (4/28), and *trpl* (18/28) (Fig. 1b). In most of these cases, the deleterious mutations occurred independently in the terminal lineages leading to each subterranean species. Exceptions to this observation were detected for several sympatric sister species pairs and triplets, where some mutations may have occurred in the common ancestor of each pair or triplet of subterranean species. In addition to the shared mutations reported for *Paroster* taxa by Langille et al. (2021), we observed two shared deletion mutations in the phototransduction genes of sympatric sister species of *Limbodessus* taxa, including *L. micromelitaensis* (Watts and Humphreys, 2009) and *L. melitaensis* (Watts and Humphreys 2006) (30 bp deletion in *trpl*) and *L. sweetwatersensis* (Watts and Humphreys 2003) and *L. silus* (Watts and Humphreys 2003) (3 bp deletion in *Arr2*), suggesting these mutations occurred in the common ancestor branch of each pair of species. Two additional shared mutations (a 1 bp deletion and a 24 bp deletion; see SI6) in *Arr2* were found for the sister species *L. leysi* (Watts and Humphreys 2006) and *L. windarraensis* (Watts and Humphreys 1999) from adjacent calcretes, further suggesting these occurred in their common ancestor, which was likely to have been a stygobiont. Last, several shared insertion (1–2 bp) mutations at position 25 of *Arr2* in *Limbodessus* species (*lapostaae* (Watts and Humphreys 1999), *macrotarsus* (Watts and Humphreys, 2003), *millbilliensis* (Watts and Humphreys, 2006), *palmulaoides* and *pulpa*) were also detected, but these occurred in a 7 bp homopolymer run of As, and the distant phylogenetic position of the taxa involved (see Fig. 1) suggest these were cases of independently evolved insertion mutations. In contrast to the eight above genes, we recovered ORFs for *inaD* sequences in all surface and subterranean species. However, we also detected ORFs for *trp* from all subterranean *Limbodessus* species, in contrast to *Paroster* species where four *trp* pseudogenes were detected.

3.4. Tests of purifying selection and neutral evolution

HyPhy site by site fixed effect likelihood analyses (Pond and Frost, 2005; SI7)) showed that for all eight candidate genes, subterranean species exhibited elevated ω values (ω range 0.461–1.519) compared to surface species (ω range 0.006–0.229). The positive control gene (*inaD*) exhibited a similar but less pronounced trend: subterranean ω range 0.243–0.371; *cf.* surface ω range 0.042–0.085). We also ran the HyPhy branch model to test for relaxed selection (RELAX; Kosakovsky et al., 2015a), and in all cases an alternative model, allowing transformation to a different branch class, fitted the data better, based on likelihood ratio values (Table 2). Most genes had a significant ($p < 0.001$) k value of less than one (range 0.00 to 0.59), indicating a relaxation of selection of these genes in subterranean species. The exception was *Limbodessus* and *Paroster inaD* and *trp*, which showed non-significant k values (Table 2).

4. Discussion

The evolutionary processes that lead to the loss of vision in subterranean animals, either via selection or neutral evolution have been widely debated (Culver and Wilkens, 2000; Jeffery, 2009; Rétaux and Casane, 2013; Wilkens, 2020). Using a unique system comprising independently evolved subterranean water beetles we have explored the molecular evolutionary forces that are operating on genes involved in phototransduction. Our results revealed that eight genes responsible for a variety of functions along the phototransduction pathway (*arr1*, *arr2*, *ninaC*, *c-opsin*, *lwop*, *uvop*, *trp*, and *trpl*) showed direct evidence of neutral molecular evolutionary processes leading to parallel pseudogenisation among 31 subterranean species of the dytiscid beetle genera *Paroster* and *Limbodessus* (Fig. 1) and one from the genus *Neobidessodes* (*N. gutteridgei*, transcriptome data only; Table 1). In contrast, 10 phototransduction genes had open reading frames in all species studied and were found to be transcribed in both subterranean and surface species. Based on gene expression studies in *Drosophila melanogaster* Meigen, 1830 (http://flybase.org/), it is likely that these 10 genes are pleiotropic and are maintained by purifying selection because they have other biological functions in addition to phototransduction (SI1).

These findings provide strong support for the hypothesis that genes which are specific in their function for traits that are no longer required (e.g., eyes and pigment in subterranean animals) ultimately develop into pseudogenes and that, given sufficient time, the loss of a light detection system will be irreversible. Our study cannot rule out that selection may operate via antagonistic pleiotropy (Jeffery, 2009) or energy conservation (Moran et al., 2015) in the initial loss of vertebrate eye phenotypes (see Wilkens, 2020 for the neutral evolution case). However, we contend that the absence of selective processes are a major contributing factor to the loss of photoreception from a functional molecular evolutionary perspective. Our finding of expressed pseudogenes in relatively anciently evolved (>3–5 my; Leijs et al., 2012) subterranean dytiscid lineages (e.g. *Arr1, Arr2, ninaC,* and *trpl* in *P. macrosturtensis*) suggests that the switching off of the light detection system may be a slow

Table 2
Branch corrected (RELAX), independent comparisons of surface and subterranean sympatric sister species branches for determination of selection strength (k).

Species	Gene	Model	logL	AICc	np	k	p-value	LR
Limbodessus	*arr1*	null	−3830.5	7746.8	65	1.00	–	–
		alternative	−3819.7	7746.7	66	0.00	< 0.001	20.51
	arr2	null	−4692.4	9514.4	71	1.00	–	–
		alternative	−4680.6	9514.4	72	0.22	< 0.001	23.49
	inaD	null	−4461.6	9068.2	72	1.00	–	–
		alternative	−4460.5	9067.9	73	4.57	0.125	2.35
	c-opsin	null	−4344.7	8842.5	71	1.00	–	–
		alternative	−4356.5	8842.5	72	0.17	< 0.001	23.59
	lwop	null	−3141.5	6408.7	69	1.00	–	–
		alternative	−3133.8	6408.7	70	0.28	< 0.001	15.40
	trp	null	−6027.6	12199.7	72	1.00	–	–
		alternative	−6027.5	12201.7	73	1.10	0.762	0.09
	trpl	null	−6207.8	12508.4	67	1.00	–	–
		alternative	−6185.0	12508.4	68	0.00	< 0.001	45.62
Paroster	*arr1*	null	−3490.1	7030.8	45	1.00	–	–
		alternative	−3466.1	7030.8	46	0.00	< 0.001	47.99
	arr2	null	−3218.1	6492.3	39	1.00	–	–
		alternative	−3204.1	6492.3	40	0.05	< 0.001	27.94
	inaD	null	−4237.8	8572.0	45	1.00	–	–
		alternative	−4235.6	8572.0	46	0.59	0.035	4.45
	c-opsin	null	−2940.7	5964.3	41	1.00	–	–
		alternative	−2937.2	5964.3	42	0.05	0.008	7.00
	lwop	null	−3873.1	7787.4	43	1.00	–	–
		alternative	−3854.0	7787.4	44	0.00	< 0.001	38.18
	trp	null	−1049.2	2155.8	35	1.00	–	–
		alternative	−1047.8	2155.8	36	0.00	0.089	2.90

np = number of parameters.
k = selection intensity parameter, where a significant $k > 1$ indicates intensification of selection and a significant $k < 1$ indicates a relaxation of selection.
Bonferroni corrected p-value, where $p < 0.007$ and $p < 0.008$ indicates significance for *Limbodessus* and *Paroster*, respectively.
LR = likelihood ratio.

process, involving genetic drift and the random accumulation of mutations in regulatory regions (e.g., promoter regions) leading to the absence of transcription, or major loss-of-function mutations (e.g., frameshift, nonsense and key missense mutations) in exon regions, which truncate or modify important active sites in the encoded proteins. Interestingly, behavioural studies suggest that *P. macrosturtensis* has retained phototactic ability (i.e. avoids lighted zones; Langille et al., 2019), most likely from an interstitial ancestor (Langille et al., 2021), despite the species having a complete absence of eyes/eye structures and living for millions of years in darkness (Leijs et al., 2012). Evidence for this behaviour further suggests that the genes *Arr1*, *Arr2*, *ninaC* and *trpl* may have retained some level of functionality in *P. macrosturtensis*, despite the evidence for their pseudogenisation. Overall, the above findings of retained phototactic behaviour in a permanently aphotic environment, and transcription of phototransduction pseudogenes in the subterranean dytiscid species, suggest that loss-of-function mutations are unlikely to have been driven to fixation by natural selection. A remaining question, however, is how many 'vision' genes are on the pseudogene trajectory in the subterranean dytiscids?

There is mounting evidence from analyses of *Drosophila* that phototransduction genes could be involved in a range of other signalling pathways in insects. Arrestins *Arr1* and *Arr2* were found to be expressed in olfactory neurons, suggesting that these visual genes are important in olfaction (Merrill et al., 2002). Merrill et al. (2002) also found detectable alterations in electropalpogram (EPG) and electroantennogram (EAG) recordings when mutations to these arrestins were present, providing strong evidence for their direct function in *Drosophila* olfaction. Auditory defects have been found linked to various visual genes (as well as others) including *trp*, *trpl*, and rhodopsins (Senthilan et al., 2012). Several studies also suggest that visual genes (e.g. rhodopsin) may play a role in thermosensation in *Drosophila* (Shen et al., 2011; Sokabe et al., 2016). We realize *Drosophila* is a very different order from those in the beetle system; however, these findings compared to our own raise the intriguing possibility that the subterranean beetles have also lost one or more of these additional signalling systems that may no longer be required for living in a thermally stable and dark groundwater environment. The finding that *trp* has ORFs and appears to be functional in all subterranean *Limbodessus* species studied to date, but not all *Paroster* species (see Fig. 1), suggests that *trp* may have retained a pleiotropic role in a signalling pathway, unrelated to vision, within the former genus.

In many study systems, the ability to detect the mode of evolution operating on genes, specifically associated with regressed traits, is potentially hampered by insufficient time for the accumulation and fixation of mutations that result in pseudogenisation (Podlaha and Zhang, 2010). The beetle species *Limbodessus microocula* (Watts and Humphreys, 2004) and *Limbodessus micrommatoion* (Watts and Humphreys, 2006) are the only subterranean dytiscid species found so far with rudimentary eyes, are more recently evolved (~1.5 my, Leijs et al., 2012) from a surface ancestor, and showed only two cases of pseudogenisation in the vision genes studied here. Conflicting evidence has been found in the opsin genes of subterranean animals; some studies have found opsin genes with full open reading frames (e.g. crayfish: Crandall and Hillis, 1997; Amphipoda: Carlini et al., 2013; *Astyanax* cave fish: Yokoyama et al., 1995; *Lucifuga* cave fish: Policarpo et al., 2021), while other cave fish studies have shown evidence of pseudogenisation of opsin genes (*Phreatichthys andruzzii*: Calderoni et al., 2016; amblyopsid species: Niemiller et al., 2013). On the other end of the spectrum, there are also cases in which opsin genes can simply not be found in the genome. Such is the case in centipedes: *Strigamia* (Minelli, 2015) and bathynelids: *Allobathynella bangokensis* (Kim et al., 2017), which may be due to the extreme length of time in which they have been underground. We have shown that in order to acquire clear evidence to support the mode of gene evolution in cave animals, one needs a system that has evolved in the dark for millions of years. However, too much time may also erase our ability to detect evolutionary mechanisms.

The subterranean diving beetle system comprises relictual taxa that have been underground for millions of years, with multiple independently evolved subterranean species, and additional sympatric sister species that have likely speciated underground (Leijs et al., 2012; Langille et al., 2021). These attributes make the study system both robust

and statistically powerful for understanding the evolutionary forces that may, or may not, be operating on the genome of organisms evolving in an aphotic environment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Horst Wilkens, Markus Friedrich, and two anonymous reviewers for reviewing our manuscript and providing invaluable feedback. We thank the numerous field assistants who contributed to the collection of beetles or provided technical assistance, particularly J. Waldock. We also thank the Axford family for kindly providing accommodation during fieldwork, and C. Watts and H. Hamon for providing images of the beetles. This research was supported by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy (NCRIS). This work was funded by ARC Discovery grants (DP120102132 and DP180103851) to S.J.B.C., T.B., W.F.H. and A.D.A.

Author contributions

B.L.L. was responsible for laboratory work, data editing, bioinformatics, and writing the bulk of the manuscript. S.M.T. carried out transcriptome analyses, sequence selection for bait design, and provided guidance with genetic software and supervision. T.B. provided guidance with bioinformatics, and grant funding. P.B-H and M.R provided bioinformatic analyses of the transcriptome data. T.M.B. and K.M.S. provided technical support in the lab. E.P.F.-J., J.H. and D.N.S. provided molecular and phylogenetic analyses. R.L. provided conceptual development of the project, field collections, and molecular analyses of beetle species. A.V. provided bioinformatic analyses of the exon capture data. W.F.H. led field collections, contributed grant funding. A.D.A. provided grant funding and supervision of B.L.L. S.J.B.C. provided conceptual development of the project, field collections, grant funding, guidance with genetic software, paper direction and scope, and writing of some parts of the manuscript. All authors contributed to the editing of the manuscript.

Author Information

GenBank accession numbers for new sequences are ON099067 - ON099397, and ON064371 - ON064383. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.L.L (barbara.langille@adelaide.edu.au) or S.J.B.C. (steve.cooper@samuseum.sa.gov.au). This research was supported by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy (NCRIS). This work was funded by ARC Discovery grants (DP120102132 and DP180103851) to S.J.B.C., T.B., W.F.H. and A.D.A.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107522.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477.

Bi, K.e., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., Good, J.M., 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13 (1), 403.

Boisvert, S., Laviolette, F., Corbeil, J., 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17 (11), 1519–1533.

Bragg, J.G., Potter, S., Bi, K.e., Moritz, C., 2016. Exon capture phylogenomics: efficacy across scales of divergence. Mol. Ecol. Resour. 16 (5), 1059–1068.

Breder, C.M., 1942. Descriptive ecology of La Cueva Chica, with especial reference to the blind fish. *Anoptichthys. Zool.* 27, 7–15.

Bushnell, B., 2015. BBMap. Available at: http://sourceforge.net/projects/bbmap/.

Calderoni, L., Rota-Stabelli, O., Frigato, E., Panziera, A., Kirchner, S., Foulkes, N.S., Kruckenhauser, L., Bertolucci, C., Fuselli, S., 2016. Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization. Heredity 117 (5), 383–392.

Carlini, D.B., Satish, S., Fong, D.W., 2013. Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of *Gammarus minus* (Crustacea: Amphipoda). BMC Evol. Biol. 13, 89.

Cooper, S.J.B., Hinze, S., Leys, R., Watts, C.H.S., Humphreys, W.F., 2002. Islands under the desert: molecular systematics and evolutionary origins of stygobitic water beetles (Coleoptera: Dytiscidae) from central Western Australia. Invertebr Syst. 16, 589–598.

Crandall, K.A., Hillis, D.M., 1997. Rhodopsin in the dark. Nature 387, 667–668.

Culver, D.D., Wilkens, H., 2000. Critical review of the relevant theories of the evolution of subterranean animals. In: Wilkens, H., Culver, D.C., Humphreys, W.F., (Eds), Ecosystems of the world. Subterranean ecosystems. Vol. 30. Elsevier; Amsterdam: 381–398.

Darwin, C.R., 1872. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life, 6th edition. John Murray, London.

Denisov, G., Walenz, B., Halpern, A.L., Miller, J., Axelrod, N., Levy, S., Sutton, G., 2008. Consensus generation and variant detection by Celera assembler. Bioinfor. 24 (8), 1035–1040.

Emerling, C.A., Springer, M.S., 2014. Eyes underground: regression of visual protein networks in subterranean mammals. Mol. Phylo. Evol. 78, 260–270.

Espinasa, M., Espinasa, L., 2008. Losing sight of regressive evolution. Evo. Edu. Outreach. 1 (4), 509–516.

Friedrich, Markus, Chen, Rui, Bao, Riyue, Caravas, Jason, Rai, Puneet, Zagmajster, Maja, Peck, Stewart, 2011. Phototransduction and clock gene expression in the troglobiont beetle Ptomaphagus hirtus of Mammoth cave. The Journal of Experimental Biology 214, 3532–3541.

Fumey, J., Hinaux, H., Noirot, C., Thermes, C., Retaux, S., Casane, D., 2018. Evidence for late Pleistocene origin of *Astyanax mexicanus* cavefish. BMC Evol. Biol. 18 (43) https://doi.org/10.1186/s12862-018-1156-7.

Glenn, T.C., Nilsen, R.A., Kieran, T.J., Sanders, J.G., Bayona-Vasquez, N.J., Finger, J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., et al. 2016. Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru and iNext). *bioRxiv*, doi: 10.1101/049114.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotech. 27 (2), 182–189.

Jeffery, W.R., 2009. Regressive evolution in *Astyanax* cavefish. Annu. Rev. Genet. 43 (1), 25–47.

Jeffery, W.R., Strickler, A.G., Guiney, S., Heyser, D.G., Tomarev, S.I., 2000. *Prox1* in eye degeneration and sensory organ compensation during development and evolution of the cavefish *Astyanax*. Dev. Genes. Evol. 210 (5), 223–230.

Herman, A., Brandvain, Y., Weagley, J., Jeffery, W.R., Keene, A.C., Kono, T.J.Y., Bilandžija, H., Borowsky, R., Espinasa, L., O'Quin, K., Ornelas-García, C.P., Yoshizawa, M., Carlson, B., Maldonado, E., Gross, J.B., Cartwright, R.A., Rohner, N., Warren, W.C., McGaugh, S.E., 2018. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra. *Astyanax mexicanus*. *Mol Ecol.* 27 (22), 4397–4416.

Ho, W.-C., Ohya, Y., Zhang, J., 2017. Testing the neutral hypothesis of phenotypic evolution. PNAS 114 (46), 12219–12224.

Hugall, A.F., O'Hara, T.D., Hunjan, S., Nilsen, R., Moussalli, A., 2015. An exon-capture for the entire class Ophiuroidea. Mol. Biol. Evol. 33, 281–294.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinfor. 28 (12), 1647–1649.

Kim, B.-M., Kang, S., Ahn, D.-H., Kim, J.-H., Ahn, I., Lee, C.-W., Cho, J.-L., Min, G.-S., Park, H., 2017. First insights into the subterranean crustacean Bathynellacea transcriptome: transcriptionally reduced opsin repertoire and evidence of conserved homeostasis regulatory mechanisms. *PLoS ONE*. 12: e0170424.

Kimura, M., 1968. Evolutionary rate at the molecular level. Nature 217 (5129), 624–626.

Kimura, M. (Ed.), 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press.

Kosakovsky, SL., Murrell, B., Weaver, S., Temple iGEM/UCSD viral evolution group, 2015a. RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.*, 32: 820–832.

Kumar, S., Patel, R., 2018. Neutral theory, disease mutations, and personal exomes. Mol. Biol. Evol. 35, 1297–1303.
Langecker, T.G., Schmale, H., Wilkens, H., 1993. Transcription of the opsin gene in degenerate eyes of cave-dwelling *Astyanax fasciatus* (Teleortei, Characidae) and of its conspecific epigean ancestor during early ontogeny. Cell Tissue Res. 273, 183–192.
Langille, B.L., Tierney, S.M., Austin, A.D., Humphreys, W.F., Cooper, S.J.B., 2019. How blind are they? Phototactic responses in stygobiont diving beetles (Coleoptera; Dytiscidae) from calcrete aquifers of Western Australia. Austral Entomol. 58 (2), 425–431.
Langille, B.L., Hyde, J., Saint, K.M., Bradford, T.M., Stringer, D.N., Tierney, S.M., Humphreys, W.F., Austin, A.D., Cooper, S.J.B., 2021. Evidence for speciation underground in diving beetles (Dytiscidae) from a subterranean archipelago. Evol. 75 (1), 166–175.
Leys, R., Watts, C.H.S., Cooper, S.J.B., Humphreys, W.F., 2003. Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. Evol. 57, 2819–2834.
Leys, R., Cooper, S.J.B., Strecker, U., Wilkens, H., 2005. Regressive evolution of an eye pigmentation gene in independently evolved eyeless subterranean diving beetles. Biol. Lett. 1, 496–499.
Leijs R, vanNes EH, Watts CH, Cooper SJB, Humphreys WF, Hogendoorn K. 2012. Evolution of blind beetles in isolated aquifers: a test of alternative modes of speciation. *PLoS ONE*. 7: e34260.
Li, H.L., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinfor. 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequences alignment/map format and SAMtools. Bioinfor. 25, 2078–2079.
Lynch, M., 2007. The origins of genome architecture. Sinauer, Sunderland, MA.
Menuet, A., Alunni, A., Joly, J.S., Jeffery, W.R., Retaux, S., 2007. Expanded expression of sonic hedgehog in *Astyanax* cavefish: multiple consequences on forebrain development and evolution. Devel. 134, 845–855.
Merrill, C.E., Riesgo-Escovar, J., Pitts, R.J., Kafatos, F.C., Carlson, J.R., Zwiebel, L.J., 2002. Visual arrestins in olfactory pathways of *Drosophila* and the malaria vector mosquito *Anopheles gambiae*. PNAS 99, 1633–1638.
Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. https://doi.org/10.1101/pdb.prot5448.
Miller, Mark, Pfeiffer, W.T, Schwartz, Terri, 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Gateway Computing Environments Workshop (GCE) 2010, 1–8. https://doi.org/10.1109/GCE.2010.5676129.
Minelli, A., 2015. Genome evolution: Groping in the soil interstices. Current Biol. 25, 194–196.
Moran, D., Softley, R., Warrant, E.J., 2015. The energetic cost of vision and the evolution of eyeless Mexican cavefish. Sci. Adv. 1, e1500363.
Niemiller, M.L., Fitzpatrick, B.M., Shah, P., Scmitz, L., Near, T.J., 2013. Evidence for repeated loss of selective constraint in rhodopsin of amblyopsid cavefishes (Teleostei: Amblyopsidae). Evol. 67, 732–748.
Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2010. IDBA – a practical iterative de bruijn graph de novo assembler. Recomb.
Podlaha, O., Zhang, J., 2010. Pseudogenes and their evolution. Encyclopedia of Life Sciences. John Wiley and Sons Ltd:, Chichester.
Policarpo, M., Fumey, J., Lafargeas, P., Naquin, D., Thermes, C., Naville, M., Dechaud, C., Volff, J.-N., Cabau, C., Klopp, C., Moller, P.R., Bernatchez, L., Garcia-Machado, E., Retaux, S., Casane, D., 2021. Contrasting gene decay in subterranean vertebrates: Insights from cavefishes and fossil mammals. Mol. Biol. Evol. 38, 589–605.
Pond, S.L.P., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.
Portik, D.M., Smith, L.L., Bi, K., 2016. An evaluation of transcriptome based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol. Ecol. Resour. 16, 1069–1083.
Protas, M., Conrad, M., Gross, J.B., Tabin, C., Borowsky, R., 2007. Regressive evolution in the Mexican cave tetra. *Astyanax mexicanus*. *Curr Biol*. 17, 452–454.
Puritz, J.B., Lotterhos, K.E., 2018. Expressed exome capture sequencing: a method for cost-effective exome sequencing for all organisms. Mol. Ecol. Resour. 18, 1209–1222.
Rambaut, A., 2012. FigTree v1.4.2. Available at http://tree.bio.ed.ac.uk/software/figtree/.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
Rétaux, S., Casane, D., 2013. Evolution of eye development in the darkness of caves: adaptation, drift, or both? EvoDevo. 4, 26.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrated genomics viewer. Nat. Biotechnol. 29, 24–26.
Senthilan, P.R., Piepenbrock, D., Ovezmyadov, G., Nadrowski, B., Bechstedt, S., Pails, S., Winkler, M., Möbius, W., Howard, J., Göpfert, M.C., 2012. *Drosophila* Auditory Organ Genes and Genetic Hearing Defects. Cell 150, 1042–1054.
Sharma, V., Hecker, N., Walther, F., Stuckas, H., Hiller, M., 2020. Convergent losses of TLR5 suggest altered extracellular flagellin detection in four mammalian lineages. Mol. Biol. Evol. 37, 1847–1854.
Shen, W.L., Kwon, Y., Adegbola, A.A., Luo, J., Chess, A., Montell, C., 2011. Function of rhodopsin in temperature discrimination in *Drosophila*. Science 331, 1333–1336.
Sokabe, T., Chen, H.C., Luo, J., Montell, C., 2016. A switch in thermal preference in *Drosophila* larvae depends on multiple rhodopsins. Cell Rep. 17, 336–344.
Springer, M.S., Burk, A., Kavanagh, J.R., Waddell, V.G., Stanhope, M.J., 1997. The interphotoreceptor retinoid binding protein gene in therian mammals: Implications for higher level relationships and evidence for loss of function in the marsupial mole. PNAS 94, 13754–13759.
Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinfo. 9, 1312–1313.
Stern, D.B., Crandall, K.A., 2018. Phototransduction gene expression and evolution in cave and surface crayfishes. Integr. Comp. Biol. 58, 398–410.
Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Itegrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14, 178–192.
Tierney, S.M., Cooper, S.J.B., Saint, K.M., Bertozzi, T., Hyde, J., Humphreys, W.F., Austin, A.D., 2015. Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches. R. Soc. Open Sci. 2, 140386.
Tierney, S.M., Langille, B., Humphreys, W.F., Austin, A.D., Cooper, S.J.B., 2018. Massive parallel regression: A précis of genetic mechanisms for vision loss in diving beetles. Integr. Comp. Biol. 58, 465–479.
Tommaso, P.D., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids. 39, W13–W17.
Ward, C.M., To, T.-H., Pederson, S.M., 2019. ngsReports: An R Package for managing FastQC reports and other NGS related log files. Bioinfor. 15, 2587–2588.
Watts, C.H.S., Humphreys, W.F., 2009. Fourteen new Dytiscidae (Coleoptera) of the genera *Limbodessus* Guignot, *Paroster* Sharp, and *Exocelina* Broun from underground waters in Australia. Trans. R. Soc. S. Aust. 133, 62–107.
Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., Pond, S.L.K., 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Mol. Biol. Evol. 35, 773–777.
Wertheim, J.O., Murrell, B., Smith, M.D., Pond, S.L.K., Scheffler, K., 2015. RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.*, 32: 820-832.
Wilkens, H., 2010. Genes, modules and the evolution of cave fish. Heredity 105, 413–422.
Wilkens, H., 2020. The role of selection in the evolution of blindness in cave fish. Biol. J. Linn. Soc. 130, 421–432.
Wilkens, H., Strecker, U., 2003. Convergent evolution in the cavefish *Astyanax* (Characidae, Teleostei): genetic evidence from reduced eye-size and pigmentation. Biol. J. Linn. Soc. 80, 545–554.
Wilkens, H., Strecker, U., 2017. Surface and cave populations of Mexican *Astyanax*. In: Evolution in the dark, 207. Springer, Berlin, p. P.37-52..
Yamamoto, Y., Espinasa, L., Stock, D.W., Jeffery, W.R., 2003. Development and evolution of craniofacial patterning is mediated by eye-dependent and –independent processes in the cavefish *Astyanax*. Evol Dev. 5, 435–446.
Yoder, A.D., Poelstra, J.W., Tiley, G.P., Williams, R.C., 2018. Neutral theory is the foundation for conservation genetics. Mol. Biol. Evol. 35, 1322–1326.
Yokoyama, S., Meany, A., Wilkens, H., Yokoyama, R., 1995. Initial mutational steps toward loss of opsin gene function in cavefish. Mol. Biol. Evol. 12, 527–532.
Zhang, J., 2018. Neutral theory and phenotypic evolution. Mol. Biol. Evol. 35, 1327–1331.