Trends in loss and retention of vision-related genes are similar between two trechine beetles independently colonizing caves

Takuma Niida (丹伊田 拓磨)[1] Hisashi Ashida (芦田 久)[2] Shigeyuki Koshikawa (越川 滋行)[1,3]

[1]Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan
[2]Faculty of Biology-Oriented Science and Technology, Kindai University, Wakayama, Japan
[3]Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan

Corresponding authors
Takuma Niida, Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan. Email: ymbnw8bwgmxikb91sqck@gmail.com
Shigeyuki Koshikawa, Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan. Email: koshi@ees.hokudai.ac.jp

Present address
Takuma Niida, Laboratory of Insect Ecology, Graduate School of Agriculture, Kyoto University, Kyoto, Japan.

ORCID
Takuma Niida https://orcid.org/0009-0005-9678-0391
Hisashi Ashida https://orcid.org/0000-0001-5844-4075
Shigeyuki Koshikawa https://orcid.org/0000-0003-2831-555X

Abstract

Whether evolution is predictable has been tested in evolutionary biology by comparing lineages that experienced parallel evolution. For example, the repeatability of gene expression between strains was examined in the experimental evolution of bacteria. However, whether it is possible to predict the evolutionary fate of a gene (i.e., loss or retention) after an organism colonizes a new habitat and experiences a long period is not sufficiently clear. Here, we investigate a visual gene set in two species of eyeless trechine beetles (Coleoptera: Carabidae: Trechinae), which are thought to have colonized caves independently, and show that many of the lost genes and retained genes are common between them. We also estimate the pleiotropy which represents the extent to which these genes act in several tissues, using gene expression data in a model organism, and show that commonly lost genes have low pleiotropy. Our results suggest that the loss and retention of a visual gene set are relatively easy to predict in cave-dwelling trechine beetles. Furthermore, this study supports the possibility that even evolutionary fates of genes, which occur after a long period, are influenced by the functional constraints of these genes.

Keywords

predicting evolution, visual system, gene loss, pleiotropy, parallel subterranean colonization, Carabidae

Introduction

The predictability of evolution is one of the hotly debated topics in evolutionary biology (Lässig et al., 2017; Nosil et al., 2020; Mas et al., 2020). Evolution is generally difficult to predict due to stochastic factors such as genetic variations and drift. However, parallel evolution, in which analogous phenotypes occur in similar environments, has been confirmed (Losos, 1998; Rundle, 2000; Muschick et al., 2012), and the evolution of particular phenotypes under certain conditions is thought to be predictable.

Parallel evolution occurs in some organisms that have colonized underground from the surface (Jeffery, 2009; Friedrich, 2013). Subterranean organisms have often lost or show reduction of traits that are useless under aphotic conditions, namely, eyes and pigmentation, although they have sometimes enhanced some sensory systems (e.g., mechanoreception or chemoreception; Soares & Niemiller, 2013). As genetic signatures related to the regressive phenotypes, some genes have been found to have turned into pseudogenes or to have been lost from genomes. For example, in some cavefish, loss-of-function mutations have been detected in coding regions of genes related to vision and pigmentation (Policarpo et al., 2021; Zhao et al., 2022). In blind arthropod species belonging to the classes Chilopoda and Entognatha, the *opsin* subfamily involved in photoreception and the *gustatory receptor 28b* gene involved in light avoidance behavior (Xiang et al., 2010) are thought to have been lost, as these genes are not found in their assembled genome sequences (Chipman et al., 2014; Manni et al., 2020).

The group of vision-related genes includes genes encoding photoreceptor proteins, transporters and enzymes required for phototransduction. In insects, ommochrome pigments in each ommatidium filter and absorb light, playing a role in image formation (Dontsov et al., 2020), and thus genes involved in the biosynthesis of these pigments are required for vision (Shirai & Daimon, 2020; Heu et al., 2022). In a study comparing some of these visual genes across lineages that independently colonized underground, it remained unclear whether lost genes or retained genes were common (Langille et al., 2022). Clarifying this will be important for understanding the predictability of gene loss.

Here we focused on trechine beetles (Coleoptera: Carabidae: Trechinae) that colonized the subterranean (also called “hypogean”) environment from the surface. Cave-dwelling trechine beetles show the loss or reduction of compound eyes and less-pigmented bodies. These features are often observed in subterranean insects and are referred to as troglomorphism (Vandel, 1964; Barr, 1968; Chen et al., 2021). Based on morphological classification, cave-dwelling trechine beetles were classified into several genera, for example *Nipponaphaenops* Uéno, 1971 (a group with a relatively larger body

and long legs) or *Kurasawatrechus* Yoshida et Nomura, 1952 (one with a smaller body and short legs). Although evolutionary scenarios are sometimes reevaluated by molecular phylogenetic studies (Faille et al., 2023), some lineages among the cave-dwelling species are thought to have colonized caves independently. Thus, trechine beetles are well suited for testing the commonality of loss and retention of each visual gene.

This study aimed to reveal whether trends of loss and retention of visual genes are similar among cave-dwelling trechine beetles. We compared two cave species (*Nipponaphaenops erraticus* Uéno, 1971 and *Kurasawatrechus hirakei hirakei* Uéno, 1979) and two surface species (*Trechiama lewisi* (Jeannel, 1924) and *Blemus discus* (Fabricius, 1792)). First, their DNA and RNA sequencing were conducted, and phylogenetic relationships among the species were analyzed using the assembled transcripts. Second, we performed BLAST searches for the assembled genomes using multiple visual genes, identified pseudogenes and examined the commonality of lost genes and retained genes between two cave species. Finally, we tested the correlation between commonly lost genes and their pleiotropy, which has been proposed to be a constraint that causes genetic parallel evolution (Gompel & Prud'homme, 2009; Stern, 2013). This research provides the first insight into the predictability of loss of genes for cave-dwelling trechine beetles.

Materials and methods

Insects

Four species of trechine beetles were collected at various sites in Japan (Figure 1a). Adults of a cave-dwelling species, *Nipponaphaenops erraticus* Uéno, 1971, walking on gravel were collected in Hiura-dô Cave, which is a limestone cave at Kumakogen Town, Ehime Prefecture (collection date, 18 June 2022) (Figure 1b). Adults of another cave-dwelling species, *Kurasawatrechus hirakei hirakei* Uéno, 1979, walking on wet soil were collected in Koya-no-kômori-ana Cave, which is a limestone cave at Taiki Town, Mie Prefecture (24 June 2023) (Figure 1b). The two species apparently lack any eye structure or eye pigmentation. Adults of a surface-dwelling species, *Trechiama lewisi* (Jeannel, 1924), walking under rocks and moss were collected at Mount Norikura, Matsumono City, Nagano Prefecture (1 September 2022) (Figure 1b). Adults of another surface-dwelling species, *Blemus discus* (Fabricius, 1792), walking in soil near grass were collected at a riparian zone of the Ishikari River, Ebetsu City, Hokkaido (23 July 2023) (Figure 1b). The two surface-dwelling species have well-developed eye structures and eye pigmentation.

Genome and transcript sequencing

One adult specimen stored in 99.5% ethanol was used for genome sequencing for each species (males of *N. erraticus* and *T. lewisi*, and females of *K. hirakei hirakei* and *B. discus* were used). One live adult (fully matured, i.e., non-teneral, meaning an individual whose exoskeleton has sufficiently hardened over a certain period since emergence) was used for transcript sequencing for each species: males of *T. lewisi* (11 days after collection) and *B. discus* (9 days), and females of *N. erraticus* (32 days) and *K. hirakei hirakei* (5 days) were used. As preprocessing for DNA and RNA extraction, mites adhering to the body surface were removed (mites were observed only on *B. discus*) and guts were removed using tweezers (DU-5SPSF; DUMONT, Montignez, Switzerland) and a microscope (CX-43; OLYMPUS, Tokyo, Japan). When males were used, genitalia were preserved in 99.5% ethanol for identification based on the morphology.

Genomic DNA was extracted from the whole body using a Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). DNA concentrations were checked using a Qubit fluorometer (Life Technologies, Paisley, UK). For extracted samples including 1000 ng or more of DNA, libraries were constructed using a TruSeq DNA PCR-free Library Prep Kit (Illumina, San Diego, CA, USA). For extracted samples containing less than 1000 ng of DNA, libraries were constructed using a TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA). All libraries were sequenced on the NovaSeq

6000 platform (Illumina) and Paired-end reads (2 × 151 bp) were generated by Macrogen Service (Macrogen, Seoul, South Korea) (Table S1).

Total RNA was extracted from the whole body using an RNeasy Micro Kit (Qiagen, Hilden, Germany). RNA concentrations were checked using a Qubit fluorometer. For extracted samples containing 1000 ng or more RNA, libraries were constructed using a TruSeq stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA). For extracted samples containing less than 1000 ng of RNA, libraries were constructed using a SMARTer stranded RNA Library Prep Kit (Illumina, San Diego, CA, USA). All libraries were sequenced on the NovaSeq 6000 platform and Paired-end reads (2 × 151 bp) were generated by Macrogen Service (Table S1).

De novo assembly

Summary statistics of raw reads and adapter contamination were checked using FastQC v0.11.9 (Babraham Institute). Quality control was performed using fastp v0.20.1 (Chen et al., 2018) to trim off one base from the 3′ end, low quality sequences (--cut_right_window_size 4, --cut_right_mean_quality 15) and adapter sequences (adapters can be trimmed by overlap analysis). Then, summary statistics were rechecked using FastQC to confirm. Genome size was estimated using KmerGenie v1.7051 (Chikhi & Medvedev, 2014). For the reads that passed the quality control, genome and transcript assemblies were conducted with Platanus v1.2.4 (Kajitani et al., 2014) and Trinity v2.8.4 (Grabherr et al., 2011). Before the scaffolding step of genome assembly, contigs smaller than 500 bp were excluded. Completeness of the assembled genomes and transcripts was assessed using BUSCO_v5 for 1367 Insecta core genes in gVolante web server (Nishimura et al., 2017). Summary statistics of the assembled sequences were calculated using SeqKit v.0.16.1 (Shen et al., 2016).

Data collection

As a previously sampled trechine beetle, we utilized genome and transcript sequences of *Trechiama kuznetsovi* Uéno et Lafer, 1992 (Niida et al., 2023; DDBJ/EMBL/GenBank accession numbers: DRR415324, DRR415325, PRJDB15628, PRJDB15666, BSYM01000001-BSYM01055616 and ICTK01000001-ICTK01071292). *Trechiama kuznetsovi* adults inhabit the upper hypogean zone of Yûbari mountains, Yûbari City, Hokkaido. They have a vestigial eye structure but no eye pigmentation. We also collected publicly available transcripts or genome of other Coleoptera species (*Calosoma frigidum* Kirby, 1873, *Gyrinus marinus* Gyllenhal, 1808 and *Pogonus chalceus* (Marsham, 1802)) as outgroups from NCBI databases (Accession numbers:

SRR2083640, SRR921604 and PRJNA381601).

Phylogenetic analysis

To estimate the phylogeny among sampled trechine beetles, we used transcript orthologs (Ramesh et al., 2021). For the assembled and collected transcripts, the longest isoforms were retained using TransDecoder.Longorfs (https://github.com/TransDecoder/TransDecoder). Obtained amino acid sequences were searched using DIAMOND (Buchfink et al., 2021) BLASTP (-sensitive -max-target-seqs 1 -evalue 1 × E-10) against arthropod sequences (taxonomy_id:6656) of UniProtKB/Swiss-Prot. For the BLAST-hit amino acid sequences, 395 single-copy orthogroups were identified using Orthofinder v2.5.5 (Emms & Kelly, 2019) among all seven species (sampled five trechine species, *C. frigidum* and *G. marinus*) and the orthogroups were aligned using MAFFT v7.520 (all parameters were default) (Nakamura et al., 2018). Then, aligned orthogroups were imported into StarBEAST2 template (Ogilvie et al., 2017) in BEAST2 v2.7.5 (Bouckaert et al., 2014). The site models were not linked for each orthogroup. The clock models and the tree models were linked among all orthogroups. For a population model, analytical population size integration was used by default. For sites models, the Blosum62 substitution model was set with estimated substitution rate, zero gamma category count and zero proportion invariant. For a clock model, the amino acid substitution rate of 6.43E-4 ± 1.5E-5 per site per million years (Savard et al., 2006) was set with an uncorrelated lognormal. For a tree model, a Yule tree prior was set, and a fossil record of Adephaga, which is 254.2–252.2 million years ago (Mya) (Yan et al., 2018; McKenna et al., 2019), was used for a calibration point. BEAST analysis was run on 10,000,000 generations of Markov chain Monte Carlo (MCMC) with sampling every 5000 generations. The initial 10% of runs were discarded as burn-in. For the remaining 1,801 trees, the estimated sample size (ESS) was checked using Tracer v1.7.2 (Rambaut et al., 2018). Here, 16 orthogroups with ESS < 200 were removed and BEAST analysis was performed again for remaining 379 orthogroups (See Results). A maximum clade credibility (MCC) tree with median heights was calculated using TreeAnnotator v2.7.5 (https://www.aclweb.org/anthology/L18-1308.pdf). A phylogeny and divergence time were visualized using FigTree v1.4.4 (https://github.com/rambaut/figtree/releases).

We also reconstructed the phylogenetic relationship among the sampled trechine beetles, using orthologs obtained from their genomes. Amino acid sequences of single-copy genes were obtained from BUSCO outputs for the assembled and collected genomes. A total of 940 single-copy orthogroups were identified using Orthofinder among all six

species (the five sampled trechine beetles and *P. chalceus*), and the orthogroups were aligned using MAFFT. Their aligned amino acid sequences were concatenated into single sequences for each species. A maximum likelihood (ML) tree was reconstructed using IQ-TREE v2.3.6 (Minh et al., 2020) with the LG amino acid substitution model (Le & Gascuel, 2008) and 1000 bootstrap replications. The tree was midpoint-rooted and displayed in MEGA v10 (Kumar et al., 2018).

Similarity search

Twenty-four visual genes were chosen for similarity searches. Out of the 24 visual genes, three genes encode photoreceptor proteins: Long wavelength opsin (Lw opsin), Ultraviolet opsin (Uv opsin) and C-opsin, 15 genes encode phototransduction proteins: Arrestin 1 (Arr1), Arrestin 2 (Arr2), G protein α-subunit 49B (Gα49B), G protein β-subunit 76C (Gβ76C), G protein γ-subunit 30A (Gγ30A), G protein-coupled receptor kinase 1 (Gprk1), Inactivation no afterpotential D (InaD), Neither inactivation nor afterpotential A (NinaA), Neither inactivation nor afterpotential C (NinaC), No receptor potential A (NorpA), Protein C kinase 53E (Pkc53E), Rab-protein 6 (Rab6), Retinal degeneration C (RdgC), Transient receptor potential (Trp) and Trp-like (Trpl), and six genes encode eye pigmentation proteins: Vermilion (V), Kynurenine formamidase (KFase), Cinnaber (Cn), Scarlet (St), White (W) and Cardinal (Cd). The photoreceptor proteins and phototransduction proteins were chosen from the queries of similarity searches in previous studies (Friedrich et al., 2011; Langille et al., 2022; Niida et al., 2023), and the choice of the eye pigmentation proteins referred to the ommochrome biosynthetic pathway in insects (Shirai & Daimon, 2020; Heu et al., 2022).

We conducted TBLASTN searches with the amino acid sequences of the 24 visual genes as queries against the assembled genomes as databases. Sequences with an e-value < 1 × E-20 and identity > 50%, and the best-hit sequences were treated as BLAST-hit sequences. Next, to secure the searches of genes, we conducted BLASTP searches with the BLAST-hit sequences as queries against non-redundant protein sequences from NCBI as databases.

We determined the amino acid sequences of the 24 visual genes for surface-dwelling species. For their assembled transcripts, the most highly expressed isoforms were extracted using Trinity perl scripts: align_and_estimate_abundance.pl and filter_low_expr_transcripts.pl. Next, TBLASTN search was conducted against the filtered transcripts. From the BLAST-hit transcripts, amino acid sequences and coding sequences were predicted using Transdecoder.Predict (https://github.com/TransDecoder/TransDecoder), in which the results of BLASTP

searches against UniProtKB/Swiss-Prot database and hmmscan searches against Pfam-A databases (Finn et al., 2011) were retained. Finally, we again conducted TBLASTN searches with the amino acid sequences of the genes for surface-dwelling species as queries against the assembled genomes as databases.

Identifying pseudogenes

To estimate coding regions of the examined genes in the trechine beetles, the amino acid sequences of surface-dwelling species were aligned against the BLAST-hit genome scaffolds in cave-dwelling species and surface-dwelling species, using EXONERATE (Slater & Birney, 2005). When premature stop codons or frameshift mutations were found in the estimated coding regions, the visual genes were regarded as pseudogenes. For the identified pseudogenes, it was checked whether no transcript was observed or the same mutations were observed in their transcripts.

Estimating pseudogenization age

In cave-dwelling *N. erraticus*, we found several pseudogenes (See Results). The ages when their visual genes turned into pseudogenes were inferred by methods based on the ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide substitution rates (Zhang et al., 2010). Firstly, all premature stop codons or frameshift mutations were removed in estimated coding sequences for pseudogenes of *N. erraticus*. Coding sequences for the genes of three trechine beetles: *T. kuznetsovi*, *T. lewisi* and *B. discus,* were also prepared from the assembled transcripts. As one outgroup, coding sequences for the genes of *Pogonus chalceus* or *Tribolium castaneum* were collected from NCBI. A total of five entries were translated into amino acid sequences and aligned with MUSCLE implemented in MEGA. The maximum likelihood phylogenetic tree was constructed for the five nucleotide sequences of each of the genes, where GTR + G + I models were used and topologies were fixed as species trees (See Results). Ancestral and divergent sequences obtained by the phylogenetic analyses in MEGA were used to calculate Ka/Ks ratios on each branch. The calculations were based on the maximum likelihood method (Goldman & Yang, 1994) and performed using KaKs_Calculator v3.0 (Zhang, 2022).

The pseudogenization age was calculated using the equation below:

$$Tn = T \times \frac{\omega_x - \bar{\omega}}{1 - \bar{\omega}}$$

in which ω_x was Ka/Ks ratio of *N. erraticus* lineage and $\bar{\omega}$ was the average of Ka/Ks ratios of all surface lineage. T was a constant number, i.e., the divergence time between the ancestor of *N. erraticus* and that of the closest related surface-dwelling species.

Because the divergence time could not be estimated from the phylogeny of the sampled species, we tentatively designated T as 10 Mya and calculated the hypothetical pseudogenization age as Tn. Here, $\omega_x$will be close to one if enough time has passed since one gene turned into a pseudogene, because mutations in both synonymous and non-synonymous sites are expected to accumulate at the neutral mutation rate (Zhang et al., 2010).

Evaluating the pleiotropy of visual genes

While there are several interpretations of pleiotropy, in this study, we focus on pleiotropy at the tissue level (Artieri et al., 2009; Williams et al., 2023). Namely, the pleiotropy of one gene is defined as the number of tissues in which the gene is expressed. To calculate the pleiotropy of examined visual genes, expression level data were obtained throughout the various tissues of *Drosophila melanogaster* males at the adult stage from FlyAtlas2 project reports (Leader et al., 2018) on FlyBase (https://flybase.org/). This was derived from 15 tissues: eye, head, brain, thoracicoabdominal ganglion, crop, midgut, hindgut, Malpighian tubule, fat body, salivary gland, heart, rectal pad, testis, accessory gland, and carcass. Another type of expression-level data that could be available is BeetleAltas project reports (Naseem et al., 2023). However, that project did not examine the transcripts for eyes. At present, the FlyAtlas2 project is considered to be a better choice for investigating the pleiotropy of visual genes based on the transcripts for eyes and other tissues.

Finally, the pleiotropy (tau, τ) was calculated using the equation:

$$\tau = \frac{\sum_{j=1}^{n} 1 - \log\left(A_j\right)/\log\left(A_{max}\right)}{n-1}$$

in which n is the number of tissues, A_j is the expression level in tissue j, and A_{max} is the maximum expression level of tissues. Tau value varies from 0 (broadly expressed across tissues) to 1 (expressed in only specific tissues) (Fraïsse et al., 2019).

Statistical analyses

Statistical analyses were performed using the statistical program R (R Core Team, 2018). We categorized non-detected genes and pseudogenes as lost genes, and made a 2 × 2 contingency table about lost genes and retained genes in two cave-dwelling species. Thus, the difference between observed and expected values in the table was tested by Fisher's exact test.

Two generalized linear models (GLMs) were fitted with glm function for analysis below. In cave-dwelling *N. erraticus* and *K. hirakei hirakei*, we gave each visual

gene one of three values, which were 2 (lost in both species), 1 (lost in one species) or 0 (not lost, retained in both species). A GLM was used to investigate the influence of pleiotropy (an explanatory variable) on the genes (a response variable). The response variable with 0 and positive discrete values was treated as a Poisson distribution (a log link function).

Additionally, in cave-dwelling *N. erraticus*, a GLM was used to investigate the influence of pleiotropy (an explanatory variable) on the hypothetical pseudogenization age (a response variable). The response variable with positive continuous numbers was treated as a Gamma distribution (a log link function).

Results

Genome size, genome and transcriptome assembly

The assembled genome contained 26,854 scaffolds with a total length of 466,561,169 bp (N50S: 74,992) for *N. erraticus*, 170,334 scaffolds with that of 658,245,750 bp (N50S: 7,083) for *K. hirakei hirakei*, 42,268 scaffolds with that of 389,813,926 bp (N50S: 31,542) for *T. lewisi* and 108,567 scaffolds with that of 476,552,327 bp (N50S: 9,372) for *B. discus* (Table 1). These assembled genomes had 98.39%, 87.27%, 97.07% and 92.03% BUSCO completeness for *N. erraticus*, *K. hirakei hirakei*, *T. lewisi* and *B. discus*, respectively. Genome size was estimated to be about 580 Mbp for *N. erraticus*, 856 Mbp for *K. hirakei hirakei*, 567 Mbp for *T. lewisi* and 641 Mbp for *B. discus*.

Assembled transcripts contained 64,296 contigs for *N. erraticus*, 61,818 contigs for *K. hirakei hirakei*, 55,949 contigs for *T. lewisi* and 79,009 contigs for *B. discus* (Table 1). These assembled transcripts had 97.88%, 80.76%, 97.51% and 98.39% BUSCO completeness for *N. erraticus*, *K. hirakei hirakei*, *T. lewisi* and *B. discus*, respectively. The reason why the BUSCO completeness was relatively low for *K. hirakei hirakei* might be that the amount of mRNA was low because the adult body size was small compared to those of the other species.

Phylogenetic relationships

Among seven species consisting of five trechine beetles and two adephagan beetles as outgroups, 379 single-copy orthogroups were identified from their transcripts. The aligned amino acid sequences of the orthogroups were used for constructing a Bayesian phylogenetic tree (Figure S1, See supplemental data for details). The cladogram of the Bayesian tree showed that the five trechine beetles were grouped as one clade, and cave-dwelling *N. erraticus* and *K. hirakei hirakei* did not form a monophyletic group (Figure 2). This cladogram was consistent with a maximum likelihood tree, which was constructed using 940 single-copy orthogroups from the genomes of the five trechine beetles and *P. chalceus* (Figure S2).

Photoreceptor and phototransduction genes in the genomes

Three photoreceptor genes, namely *lw opsin*, *uv opsin* and *c-opsin*, and 15 phototransduction genes, namely *Arr1*, *Arr2*, *Gα49B*, *Gβ76C*, *Gγ30A*, *Gprk1*, *inaD*, *ninaA*, *ninaC*, *norpA, Pkc53E*, *Rab*6, *rdgC*, *trp* and *trpl*, were searched in the assembled genomes of each trechine beetle (Table 2; See Table S2 and supplementary data for searching in the transcripts). In the genome of cave-dwelling *N. erraticus*, BLAST

searches detected one photoreceptor gene (*uv opsin*) and all 15 phototransduction genes, but did not detect two photoreceptor genes (*lw opsin* and *c-opsin*). In the genome of cave-dwelling *K. hirakei hirakei*, BLAST searches detected nine phototransduction genes (*Gα49B*, *Gγ30A*, *Gprk1*, *inaD*, *ninaA*, *norpA*, *Pkc53E*, *Rab*6 and *rdgC*), but none of three photoreceptor genes nor six phototransduction genes (*Arr1*, *Arr2*, *Gβ76C*, *ninaC*, *trp* and *trpl*). In the genomes of both surface-dwelling *T. lewisi* and *B. discus*, BLAST searches detected all three photoreceptor genes and all 15 phototransduction genes.

Eye pigmentation genes in the genomes

Six eye pigmentation genes, consisting of *v*, *KFase*, *cn*, *st*, *w* and *cd*, were searched in the assembled genomes of each trechine beetle (Table 2; See Table S2 and supplementary data for searching in the transcripts). In the genome of cave-dwelling *N. erraticus,* BLAST searches detected five eye pigmentation genes (*v*, *KFase*, *st*, *w* and *cd*), but did not detect one eye pigmentation gene (*cn*). In the genome of cave-dwelling *K. hirakei hirakei*, BLAST searches detected all six eye pigmentation genes. In the genomes of both surface-dwelling *T. lewisi* and *B. discus*, BLAST searches detected all six eye pigmentation genes. They were also examined in the genome of *T. kuznetsovi*, which inhabits the upper hypogean zone (Niida et al., 2023). BLAST searches detected five eye pigmentation genes (*v*, *KFase*, *st*, *w* and *cd*), but did not detect one eye pigmentation gene (*cn*) (Table S3).

Pseudogenes

Mutations causing pseudogenization, which includes insertions and deletions leading to frameshifts and nonsense mutations, were observed in eight genes (*uv opsin*, *Arr1*, *Gβ76C*, *ninaC*, *norpA*, *trp*, *trpl* and *cd*) of *N. erraticus*, one gene (*cd*) of *K. hirakei hirakei*, and one gene (*Rab*6) of *B. discus* (Table 2). These genes were regarded as pseudogenes. One eye pigmentation gene (*cd*) was also identified as a pseudogene in the upper hypogean zone-dwelling *T. kuznetsovi* (Table S3).

Loss and retention of visual genes

Among 24 visual genes examined, 12 genes were retained in both cave-dwelling *N. erraticus* and *K. hirakei hirakei*, nine genes were lost in both species, and only three genes were lost in one but not the other species (Figure 3). The numbers of retained genes and lost genes were larger than the expected values ($p = 0.000516$, odds ratio = 40.62338, Fisher's Exact Test), showing that most of these genes were common in both cave-dwelling species.

Pleiotropic effects

Pleiotropy (tau) was calculated for 22 visual genes (Table S4). One photoreceptor gene (*c-opsin*) was removed from analyses due to the absence of the orthologous gene in the referred data sets of *D. melanogaster*. The pleiotropy of one eye pigmentation gene (*cn*) could not be calculated as it was only expressed in one tissue. Thus, analyzed genes were eight visual genes (*lw opsin*, *uv opsin*, *Arr1*, *Gβ76C*, *ninaC*, *trp*, *trpl* and *cd*) lost in both cave-dwelling *N. erraticus* and *K. hirakei hirakei* , two genes (*Arr2* and *norpA*) lost in only one species and 12 genes (*Gα49B*, *Gγ30A*, *Gprk1*, *inaD*, *ninaA*, *Pkc53E*, *Rab6*, *rdgC*, *v*, *KFase*, *st* and *w*) retained in both species. The visual genes that were lost in both species showed higher tau (Figure 4a) ($p = 0.0318$, estimate = 5.774, GLM).

The hypothetical pseudogenization ages of the eight visual genes in the *N. erraticus* lineage were estimated. When the ancestor of *N. erraticus* was assumed to diverge at 10 Mya, the pseudogenization of *uv opsin* was assumed to have occurred at 7.26 Mya, and the pseudogenizations of *Arr1* at 4.79 Mya, *Gβ76C* at 4.85 Mya, *ninaC* at 2.38 Mya, *norpA* at 0.74 Mya, *trp* at 1.73 Mya, *trpl* at 9.46 Mya and *cd* at 4.82 Mya. The visual genes with higher tau turned into pseudogenes at earlier times (Figure 4b) ($p = 0.0144$, estimate = 9.773, GLM).

Discussion

This study focused on vision-related gene sets for cave-dwelling trechine beetles, in which their compound eyes are lost and their body is less pigmented. We revealed that trends of loss and retention of genes were similar even between lineages with different origins of cave colonization.

Our phylogenomic analyses showed that cave-dwelling *N. erraticus* and *K. hirakei hirakei* do not constitute a monophyletic group. The ancestors of these two cave trechine beetles could have colonized caves independently, which is consistent with the classification based on morphological characters (Uéno, 1971, 1979). Previous studies on the phylogenies of subterranean insects, such as diving beetles inhabiting calcrete aquifers under arid zones in inland Australia (Leys et al., 2003; Toussaint et al., 2016), trechine beetles and the tribe Leptodirini inhabiting limestone caves in Europe (Faille et al., 2013; Balart-García et al., 2023), noted that the evolutionary transition from surface to underground habitats had occurred multiple times within related taxonomic groups. Considering these results, it is not surprising that multiple subterranean colonization events have occurred in trechine beetles across various regions of Japan. The number of times that independent underground colonization has occurred in the past should be tested with more species in the future.

In this study, we obtained draft genome sequences of two cave-dwelling and two surface-dwelling trechine beetle species and examined 24 of their vision-related genes. Pseudogenes and genes not detected in the assembled genome sequences can be functionally lost genes, following definitions used in other studies (Policarpo et al., 2021; Langille et al., 2022; Chipman et al., 2014; Manni et al., 2020). For cave-dwelling *N. erraticus*, eight pseudogenes (*uv opsin*, *Arr1*, *Gβ76C*, *ninaC*, *norpA*, *trp*, *trpl* and *cd*) were found and three genes (*lw opsin*, *c-opsin* and *cn*) were not found. For cave-dwelling *K. hirakei hirakei*, one pseudogene (*cd*) was found and nine genes (*lw opsin*, *uv opsin*, *c-opsin*, *Arr1*, *Arr2*, *Gβ76C*, *ninaC*, *trp* and *trpl*) were not found. These visual genes could have been functionally lost in association with subterranean colonization because they became useless under aphotic conditions. Loss of visual genes has been observed in some cavefish and groundwater diving beetles (Langille et al., 2021, 2022; Policarpo et al., 2021; Zhao et al., 2022). In the cave population of *Astyanax mexicanus*, which diverged relatively recently from its closely related surface population, the loss of visual genes has rarely been observed (Zhao et al., 2022). However, changes in the expression levels and patterns of certain candidate genes appear to be involved in eye reduction (McGaugh et al., 2014). Additional analyses revealed that the *cardinal* gene (*cd*) and *cinnabar* gene (*cn*), which encode ommochrome biosynthetic enzymes, had turned into pseudogenes and

were not found in the genome sequences for the upper hypogean zone-dwelling *T. kuznetsovi*. The *cardinal* knockout mutants show the loss of eye color in *Tribolium castaneum* adults (Shirai & Daimon, 2020), and thus the pseudogenization of the *cardinal* gene is most likely related to the lack of pigmentation in compound eyes of *Trechiama kuznetsovi* adults (Niida et al., 2023). Unexpectedly, we also found that the *Rab6* gene had turned into a pseudogene for surface-dwelling *B. discus*. The *Rab6* gene is involved in photoreceptor synaptic transmission in *Drosophila melanogaster* (Shetty et al., 1998; Huang et al., 2015), and thus there might be some kind of abnormality in the phototransduction system of *B. discus*. However, *B. discus* adults have well-developed compound eyes and no troglomorphic characters, and the pseudogenization of the *Rab6* gene is not likely associated with subterranean colonization. As future directions, functional analyses such as knockout experiments of visual genes or rescue of the pseudogenes will help to understand which genes are particularly involved in the visual loss.

When 24 visual genes were compared between cave-dwelling *N. erraticus* and *K. hirakei hirakei*, it was found that nine genes had been lost in both, and 12 genes had been retained in both. Thus, even though the two cave beetles have different origins of subterranean colonization, lost genes and retained genes are mostly common between them. This is regarded as genetic parallel evolution, following the definition by Stern (2013) and Cerca (2023). Genetic parallel evolution emerges at several levels (Mas et al., 2020): (1) the same families of genes, (2) orthologous genes and (3) the same nucleotide substitutions of a given gene. As an example of (1) in subterranean organisms, some cavefish have lost several genes categorized into gene sets related to vision, pigmentation and circadian clock; however, the specific genes lost vary in each lineage of the cavefish (Policarpo et al., 2021). As an example of (2) in subterranean organisms, *opsin* genes involved in photoreception were lost in groundwater diving beetles (Langille et al., 2022), and the expansion of orthogroups annotated as sensory perception and gene duplications of ionotropic receptors IR25a and IR8a, which are involved in thermal and humidity sensing, were observed in the tribe Leptodirini (Balart-García et al., 2023, 2024). To our knowledge, there is no supporting data for (3) in subterranean organisms. Some research suggested that gene loss had arisen from the accumulation of random mutations (Wilkens, 2020; Langille et al., 2022), and parallelly lost genes might not have experienced the same nucleotide substitutions (Protas et al., 2006; Langille et al., 2021). The results of this study can be classified into parallel evolution at the orthologous genes, namely category (2) above. We focused on multiple orthologous genes from the vision-related gene set, not just a single orthologous gene. Consequently, we showed for the first time that the trends

of gene loss and retention are similar in two genera of trechine beetles that had colonized caves independently.

It should be carefully considered why commonality was seen in the loss and retention of visual genes between lineages. If no selective pressures or constraints had acted on the visual gene set at all, most genes might have been lost in both lineages, or lost genes might have been different between lineages under stochastic effects. Here we focused on pleiotropy, calculated based on the number of tissues in which a certain gene is expressed and the extent of its expression, as an index of functional constraint on genes (Williams et al., 2023). It has been proposed that gene pleiotropy can be a factor affecting genetic parallel evolution (Gompel & Prud'homme, 2009; Stern, 2013). We estimated pleiotropies of each visual gene from *D. melanogaster* transcriptome data, and showed that lost genes in both cave-dwelling *N. erraticus* and *K. hirakei hirakei* had low pleiotropy and retained genes in both species had high pleiotropy. Thus, the commonality of gene loss and retention between lineages is likely related to the pleiotropy of their genes. Additional tests found that a gene with the highest pleiotropy turned into a pseudogene at the youngest age for cave-dwelling *N. erraticus*, although the number of analyzed pseudogenes was limited. Taken together, these findings indicate that the pleiotropy of genes could be an important factor affecting parallel loss of genes.

Parallel evolution has received considerable attention in attempts to understand the predictability of evolution (Lässig et al., 2017; Nosil et al., 2020; Mas et al., 2020). Because the available genomic information for cave insects has been scarce, whether there was parallel loss of their genes has not been well known. We obtained the draft genome sequences for trechine beetles and revealed that trends of gene loss and retention are similar in two genera of trechine beetles that colonized caves independently. Therefore, this study highlights that the evolutionary loss of visual genes can be predictable to some extent. We also referred to the gene expression data for a model organism and showed that parallelly lost genes tended to be less pleiotropic. Future studies will need to verify the relationship between parallel loss of genes and their pleiotropy in detail, using annotation information based on long-read genome sequencing or incorporating expression data for compound eyes of trechine beetles. Gene loss related to other selective pressures than darkness in the subterranean environment (e.g., hypoxia or food limitation; Trajano, 2001) should also be examined. Our approach of focusing on the pleiotropy may be useful for comprehending the predictability of gene loss.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

Sequence reads and assembled genomes and transcripts were deposited to DDBJ/EMBL/GenBank under the BioProject accession number PRJNA1237515.

Conflicts of interest

None declared.

Funding

This work was supported by KAKENHI (22KJ0085), Kurita Water and Environment Foundation, and The Asahi Glass Foundation.

Acknowledgments

We thank Dr. Takashi Hayakawa, Dr. Masahiro Ôhara and Dr. Shunsuke Utsumi for their comments and technical help. We are grateful to Mr. Keita Kuroda, Dr. Yuta Nakase, Mr. Tomohiro Yanone, Mr. Daisuke Sumikawa, Mr. Yusuke Hara, Mr. Kazuki Sugaya for field sampling and their information. *T. lewisi* samples were collected under the permit of Ministry of the Environment. We also thank colleagues of Koshikawa Laboratory for their helpful discussions and Dr. Elizabeth Nakajima for English editing. The super-computing resource was provided by Human Genome Center (The Univ. of Tokyo). TN was supported by a JSPS research fellowship.

References

Artieri, C. G., Haerty, W., & Singh, R. S. (2009). Ontogeny and phylogeny: Molecular signatures of selection, constraint, and temporal pleiotropy in the development of *Drosophila*. *BMC Biology*, *7*, 42. https://doi.org/10.1186/1741-7007-7-42

Balart-García, P., Aristide, L., Bradford, T. M., … Fernández, R. (2023). Parallel and convergent genomic changes underlie independent subterranean colonization across beetles. *Nature Communications*, *14*, 3842. https://doi.org/10.1038/s41467-023-39603-1

Balart-García, P., Bradford, T. M., Beasley-Hall, P. G., … Fernández, R. (2024). Highly dynamic evolution of the chemosensory system driven by gene gain and loss across subterranean beetles. *Molecular Phylogenetics and Evolution*, *194*, 108027. https://doi.org/10.1016/j.ympev.2024.108027

Barr, T. C. (1968). Cave ecology and the evolution of troglobites. *Evolutionary Biology*, *2*, 35–102.

Bouckaert, R., Heled, J., Kühnert, D., … Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*, 366–368. https://doi.org/10.1038/s41592-021-01101-x

Cerca, J. (2023). Understanding natural selection and similarity: Convergent, parallel and repeated evolution. *Molecular Ecology*, *32*, 5451–5462. https://doi.org/10.1111/mec.17132

Chen, M., Guo, W., Huang, S., … Liu, W. (2021). Morphological adaptation of cave-dwelling ground beetles in china revealed by geometric morphometry (Coleoptera, Carabidae, Trechini). *Insects*, *12*, 1002. https://doi.org/10.3390/insects12111002

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chikhi, R., & Medvedev, P. (2014). Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics*, *30*, 31–37. https://doi.org/10.1093/bioinformatics/btt310

Chipman, A. D., Ferrier, D. E. K., Brena, C., … Richards, S. (2014). The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biology*, *12*, e1002005. https://doi.org/10.1371/journal.pbio.1002005

Dontsov, A. E., Sakina, N. L., Yakovleva, M. A., … Ostrovsky, M. A. (2020). Ommochromes from the compound eyes of insects: physicochemical properties and antioxidant activity. *Biochemistry (Moscow)*, *85*, 668–678. https://doi.org/10.1134/s0006297920060048

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 238. https://doi.org/10.1186/s13059-019-1832-y

Faille, A., Casale, A., Balke, M., & Ribera, I. (2013). A molecular phylogeny of Alpine subterranean Trechini (Coleoptera: Carabidae). *BMC Evolutionary Biology*, *13*, 248. https://doi.org/10.1186/1471-2148-13-248

Faille, A., Fresneda, J., & Bourdeau, C. (2023). Reconciling morphological and molecular data in a highly convergent group: the Pyrenean radiation of hypogean Trechini (Coleoptera: Carabidae). *Integrative Systematics: Stuttgart Contributions to Natural History*, *6*, 9–37. https://doi.org/10.18476/2023.609967

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*, W29–37. https://doi.org/10.1093/nar/gkr367

Fraïsse, C., Puixeu Sala, G., & Vicoso, B. (2019). Pleiotropy modulates the efficacy of selection in *Drosophila melanogaster*. *Molecular Biology and Evolution*, *36*, 500–515. https://doi.org/10.1093/molbev/msy246

Friedrich, M. (2013). Biological clocks and visual systems in cave-adapted animals at the dawn of speleogenomics. *Integrative and Comparative Biology*, *53*, 50–67. https://doi.org/10.1093/icb/ict058

Friedrich, M., Chen, R., Daines, B., … Peck, S. B. (2011). Phototransduction and clock gene expression in the troglobiont beetle *Ptomaphagus hirtus* of Mammoth cave. *Journal of Experimental Biology*, *214*, 3532–3541. https://doi.org/10.1242/jeb.060368

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*, 725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153

Gompel, N., & Prud'homme, B. (2009). The causes of repeated genetic evolution. *Developmental Biology*, *332*, 36–47. https://doi.org/10.1016/j.ydbio.2009.04.040

Grabherr, M. G., Haas, B. J., Yassour, M., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. https://doi.org/10.1038/nbt.1883

Heu, C. C., Gross, R. J., Le, K. P., … Fabrick, J. A. (2022). CRISPR-mediated knockout of *cardinal* and *cinnabar* eye pigmentation genes in the western tarnished plant bug. *Scientific Reports*, *12*, 4917. https://doi.org/10.1038/s41598-022-08908-4

Huang, Y., Xie, J., & Wang, T. (2015). A fluorescence-based genetic screen to study retinal degeneration in *Drosophila*. *PLoS One*, *10*, e0144925. https://doi.org/10.1371/journal.pone.0144925

Jeffery, W. R. (2009). Regressive evolution in *Astyanax* cavefish. *Annual Review of Genetics*, *43*, 25–47. https://doi.org/10.1146/annurev-genet-102108-134216

Kajitani, R., Toshimoto, K., Noguchi, H., … Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, *24*, 1384–1395. https://doi.org/10.1101/gr.170720.113

Kumar, S., Stecher, G., Li, M., … Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*,

1547–1549. https://doi.org/10.1093/molbev/msy096

Langille, B. L., Hyde, J., Saint, K. M., … Cooper, S. J. B. (2021). Evidence for speciation underground in diving beetles (Dytiscidae) from a subterranean archipelago. *Evolution*, *75*, 166–175. https://doi.org/10.1111/evo.14135

Langille, B. L., Tierney, S. M., Bertozzi, T., … Cooper, S. J. B. (2022). Parallel decay of vision genes in subterranean water beetles. *Molecular Phylogenetics and Evolution*, *173*, 107522. https://doi.org/10.1016/j.ympev.2022.107522

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, *25*, 1307–1320. https://doi.org/10.1093/molbev/msn067

Leader, D. P., Krause, S. A., Pandit, A., … Dow, J. A. T. (2018). FlyAtlas 2: A new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Research*, *46*, D809–D815. https://doi.org/10.1093/nar/gkx976

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. *Evolution*, *57*, 2819–2834. https://doi.org/10.1111/j.0014-3820.2003.tb01523.x

Losos, J. B., Jackman, T. R., Larson, A., … Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, *279*, 2115–2118. https://doi.org/10.1126/science.279.5359.2115

Lässig, M., Mustonen, V., & Walczak, A. M. (2017). Predicting evolution. *Nature Ecology & Evolution*, *1*, 0077. https://doi.org/10.1038/s41559-017-0077

Manni, M., Simao, F. A., Robertson, H. M., … Zdobnov, E. M. (2020). The genome of the blind soil-dwelling and ancestrally wingless dipluran *Campodea augens*: A key reference hexapod for studying the emergence of insect innovations. *Genome Biology and Evolution*, *12*, 3534–3549. https://doi.org/10.1093/gbe/evz260

Mas, A., Lagadeuc, Y., & Vandenkoornhuyse, P. (2020). Reflections on the predictability of evolution: Toward a conceptual framework. *Iscience*, *23*, 101736. https://doi.org/10.1016/j.isci.2020.101736

McGaugh, S. E., Gross, J. B., Aken, B., ... Warren, W. C. (2014). The cavefish genome reveals candidate genes for eye loss. *Nature Communications*, *5*, 5307. https://doi.org/10.1038/ncomms6307

McKenna, D. D., Shin, S., Ahrens, D., … Beutel, R. G. (2019). The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 24729–24737. https://doi.org/10.1073/pnas.1909655116

Minh, B. Q., Schmidt, H., Chernomor, O., … Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Muschick, M., Indermaur, A., & Salzburger, W. (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, *22*, 2362–2368. https://doi.org/10.1016/j.cub.2012.10.048

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, *34*, 2490–2492. https://doi.org/10.1093/bioinformatics/bty121

Naseem, M. T., Beaven, R., Koyama, T., … Halberg, K. V. (2023). NHA1 is a cation/proton antiporter essential for the water-conserving functions of the rectal complex in *Tribolium castaneum*. *Proceedings of the National Academy of Sciences of the United States of America*, *120*, e2217084120. https://doi.org/10.1073/pnas.2217084120

Niida, T., Terashima, Y., Aonuma, H., & Koshikawa, S. (2023). Photoreceptor genes in a trechine beetle, *Trechiama kuznetsovi*, living in the upper hypogean zone. *Zoological Letters*, *9*, 9. https://doi.org/10.1186/s40851-023-00208-7

Nishimura, O., Hara, Y., & Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, *33*, 3635–3637. https://doi.org/10.1093/bioinformatics/btx445

Nosil, P., Flaxman, S. M., Feder, J. L., & Gompert, Z. (2020). Increasing our ability to predict contemporary evolution. *Nature Communications*, *11*, 5592. https://doi.org/10.1038/s41467-020-19437-x

Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, *34*, 2101–2114. https://doi.org/10.1093/molbev/msx126

Policarpo, M., Fumey, J., Lafargeas, P., … Casane, D. (2021). Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals. *Molecular Biology and Evolution*, *38*, 589–605. https://doi.org/10.1093/molbev/msaa249

Protas, M. E., Hersey, C., Kochanek, D., … Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, *38*, 107–111. https://doi.org/10.1038/ng1700

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rambaut, A., Drummond, A. J., Xie, D., … Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*, 901–904. https://doi.org/10.1093/sysbio/syy032

Ramesh, B., Firneno, T. J., & Demuth, J. P. (2021). Divergence time estimation of genus *Tribolium* by extensive sampling of highly conserved orthologs. *Molecular Phylogenetics and Evolution*, *159*, 107084. https://doi.org/10.1016/j.ympev.2021.107084

Rundle, H. D., Nagel, L., Boughman, J. W., & Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science*, *287*, 306–308. https://doi.org/10.1126/science.287.5451.306

Savard, J., Tautz, D., & Lercher, M. J. (2006). Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Ecology and Evolution*, *6*, 7. https://doi.org/10.1186/1471-2148-6-7

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, *11*, e0163962. https://doi.org/10.1371/journal.pone.0163962

Shetty, K. M., Kurada, P., & O'Tousa, J. E. (1998). Rab6 regulation of rhodopsin transport in *Drosophila*. *Journal of Biological Chemistry*, 273, 20425–20430.

Shirai, Y., & Daimon, T. (2020). Mutations in *cardinal* are responsible for the *red-1* and *peach* eye color mutants of the red flour beetle *Tribolium castaneum*. *Biochemical and Biophysical Research Communications*, *529*, 372–378. https://doi.org/10.1016/j.bbrc.2020.05.214

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Soares, D., & Niemiller, M. L. (2013). Sensory adaptations of fishes to subterranean environments. *Bioscience*, *63*, 274–283. https://doi.org/10.1525/bio.2013.63.4.7

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, *14*, 751–764. https://doi.org/10.1038/nrg3483

Toussaint, E. F. A., Hendrich, L., Escalona, H. E., … Balke, M. (2016). Evolutionary history of a secondary terrestrial Australian diving beetle (Coleoptera, Dytiscidae) reveals a lineage of high morphological and ecological plasticity. *Systematic Entomology*, *41*, 650–657. https://doi.org/10.1111/syen.12182

Trajano, E. (2001). Ecology of subterranean fishes: an overview. *Environmental Biology of Fishes*, *62*, 133–160. https://doi.org/10.1023/A:1011841913569

Uéno, S.-I. (1971). Occurrence of an aphaenopsoid trechine beetle in Japan (1). *Annales de Spéléologie*, *26*, 451–462.

Uéno, S.-I. (1979). New cave-dwelling trechine beetles from the eastern part of the Kii peninsula, central Japan. *Bulletin of the National Museum of Nature and Science. Series A, Zoology*, *5*, 115–126.

Vandel, A. (1964). Biospéologie. La biologie des animaux cavernicoles. Gauthier-Villars, Paris.

Wilkens, H. (2020). The role of selection in the evolution of blindness in cave fish. *Biological Journal of the Linnean Society*, *130*, 421–432. https://doi.org/10.1093/biolinnean/blaa054

Williams, A. M., Ngo, T. M., Figueroa, V. E., & Tate, A. T. (2023). The effect of developmental pleiotropy on the evolution of insect immune genes. *Genome Biology and Evolution*, *15*, evad044. https://doi.org/10.1093/gbe/evad044

Xiang, Y., Yuan, Q., Vogt, N., … Jan, Y. N. (2010). Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature*, *468*, 921–926. https://doi.org/10.1038/nature09576

Yan, E. V., Beutel, R. G., & Lawrence, J. F. (2018). Whirling in the late Permian: Ancestral Gyrinidae show early radiation of beetles before Permian-Triassic mass extinction. *BMC Ecology and Evolution*, *18*, 33. https://doi.org/10.1186/s12862-018-1139-8

Zhang, Z. (2022). KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics, Proteomics & Bioinformatics*, *20*, 536–540. https://doi.org/10.1016/j.gpb.2021.12.002

Zhang, Z. D., Frankish, A., Hunt, T., … Gerstein, M. (2010). Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates. *Genome Biology*, *11*, R26. https://doi.org/10.1186/gb-2010-11-3-r26

Zhao, Q., Shao, F., Li, Y., … Peng, Z. (2022). Novel genome sequence of Chinese cavefish (*Triplophysa rosa*) reveals pervasive relaxation of natural selection in cavefish genomes. *Molecular Ecology*, *31*, 5831–5845. https://doi.org/10.1111/mec.16700

Legends

Figure 1.
The habitats and morphology of sampled trechine beetles. (a) Collection sites are indicated as I to IV on the map. In each, photographs I to IV below the map were taken at the collection sites as follows: I, inside of a limestone cave at Kumakogen Town, Ehime Prefecture where *Nipponaphaenops erraticus* adults were collected; II, inside of a limestone cave at Taiki Town, Mie Prefecture where *Kurasawatrechus hirakei hirakei* adults were collected; III, the slope of Mount Norikura at Matsumoto City, Nagano Prefecture where *Trechiama lewisi* adults were collected; IV, a riparian zone of the Ishikari River at Ebetsu City, Hokkaido where *Blemus discus* adults were collected. (b) Photographs show dorsal views of adults: I, *N. erraticus*; II, *K. hirakei hirakei*; III, *T. lewisi*; IV, *B. discus*.

Figure 2.
Simplified cladogram showing the habitats and eyes of trechine beetles. The cladogram was based on a Bayesian tree in Figure S1. It includes five trechine beetles (Nerr, *Nipponaphaenops erraticus*; Tkuz, *Trechiama kuznetsovi*; Tlew, *Trechiama lewisi*; Bdis, *Blemus discus*; Khir, *Kurasawatrechus hirakei hirakei*) and two other beetles (Cfri, *Calosoma frigidum*; Gmar, *Gyrinus marinus*) as outgroups. Habitats of trechine beetles are shown next to OTU names as colored circles. Photographs next to OTU names are lateral views of adult heads of trechine beetles, showing that Nerr and Khir lack eyes, Tlew and Bdis have well-developed eyes, and Tkuz have reduced eyes with no pigmentation.

Figure 3.
A contingency table for 24 vision-related genes in ca v e-dwelling *Nipponaphaenops erraticus* and *Kurasawatrechus hirakei hirakei*. The top left panel shows genes retained in both species; the top right and bottom left panels show genes lost in only one species; and the bottom right panel shows genes lost in both species. Digits inside cells indicate the number of genes.

Figure 4.
Association between visual gene loss and the pleiotropy of these genes in cave-dwelling species. (a) Effects of the pleiotropy on gene loss in *Nipponaphaenops erraticus* and *Kurasawatrechus hirakei hirakei* experiencing cave colonization independently. X-axis

represents tau values as the index of pleiotropy (0 to 1). It varies from 0 (broadly expressed across tissues) to 1 (expressed only in specific tissues). Y-axis represents the number of sampled cave-dwelling species. Grey points indicate lost genes (n = 22). Two genes were excluded because these tau values could not be calculated. The line means Poisson regression predicted by a generalized linear model. (b) Effects of the pleiotropy on the age of pseudogenization in cave-dwelling *Nipponaphaenops erraticus*. X-axis represents tau values (0.5 to 1). Y-axis represents hypothetical ages of pseudogenization. It ranges from 0 (present) to 10 MYA (hypothetical divergence time when *N. erraticus* diverged from its ancestor). Points indicate pseudogenized visual genes (n = 8).

Table 1.
The summary statistics for *de novo* assemblies of genomes and transcripts in *Nipponaphaenops erraticus*, *Kurasawatrechus hirakei hirakei*, *Trechiama lewisi* and *Blemus discus*.

Table 2.
Visual genes detected in the assembled genomes of *Nipponaphaenops erraticus*, *Kurasawatrechus hirakei hirakei*, *Trechiama lewisi* and *Blemus discus*.

'yes' indicates that the genes were detected in BLAST-hit sequences. Among the genes described as 'yes', 'pseudo' means that the genes were regarded as pseudogenes. 'no' indicates that the genes were not detected in BLAST-hit sequences. Species name abbreviation: Tcas, *Tribolium castaneum*; Dmel, *Drosophila melanogaster*.

Figure S1.
A time calibrated phylogeny of seven beetles based on a Bayesian method. The tree was estimated using 379 single-copy orthologs of amino acid sequences. These orthologs were obtained from the transcripts of five trechine beetles (Nerr, *Nipponaphaenops erraticus*; Tkuz, *Trechiama kuznetsovi*; Tlew, *Trechiama lewisi*; Bdis, *Blemus discus*; Khir, *Kurasawatrechus hirakei hirakei*) and two other beetles (Cfri, *Calosoma frigidum*; Gmar, *Gyrinus marinus*) as outgroups. Posterior probabilities of divergence times are indicated to the right of the nodes. 95% highest posterior densities of divergence times are indicated by blue error bars on the nodes. The scale bar below the phylogeny indicates million years ago (MYA).

Figure S2.

A maximum likelihood phylogeny of six trechine beetles. The tree was reconstructed using 940 single-copy orthologs of amino acid sequences. These orthologs were obtained from the genomes of five trechine beetles (Nerr, *Nipponaphaenops erraticus*; Tkuz, *Trechiama kuznetsovi*; Tlew, *Trechiama lewisi*; Bdis, *Blemus discus*; Khir, *Kurasawatrechus hirakei hirakei*) and one beetles (Pcha, *Pogonus chalceus*), which was assigned as an outgroup by mid-point rooting. Bootstrap probabilities are provided on the right of the nodes. The scale bar means the number of amino acid substitutions per site.

Table S1.
Information of sequencing reads before and after trimming in *Nipponaphaenops erraticus*, *Kurasawatrechus hirakei hirakei*, *Trechiama lewisi* and *Blemus discus*.

Table S2.
Visual genes detected in the assembled transcripts of *Nipponaphaenops erraticus*, *Kurasawatrechus hirakei hirakei*, *Trechiama lewisi* and *Blemus discus*.

‘yes’ indicates that the genes were detected in BLAST-hit sequences. ‘no’ indicates that the genes were not detected in BLAST-hit sequences. Species name abbreviation: Tcas, *Tribolium castaneum*; Dmel, *Drosophila melanogaster*.

Table S3.
Eye pigmentation genes detected in the genome and transcripts of *Trechiama kuznetsovi*.

‘yes’ indicates that the genes were detected in BLAST-hit sequences. Among the genes described as ‘yes’, ‘pseudo’ means that the genes were regarded as pseudogenes. ‘no’ indicates that the genes were not detected in BLAST-hit sequences. Species name abbreviation: Tcas, *Tribolium castaneum*.

Table S4.
Tau values of 22 visual genes that were estimated form the transcriptome database of *Drosophila melanogaster* male at adult stage.

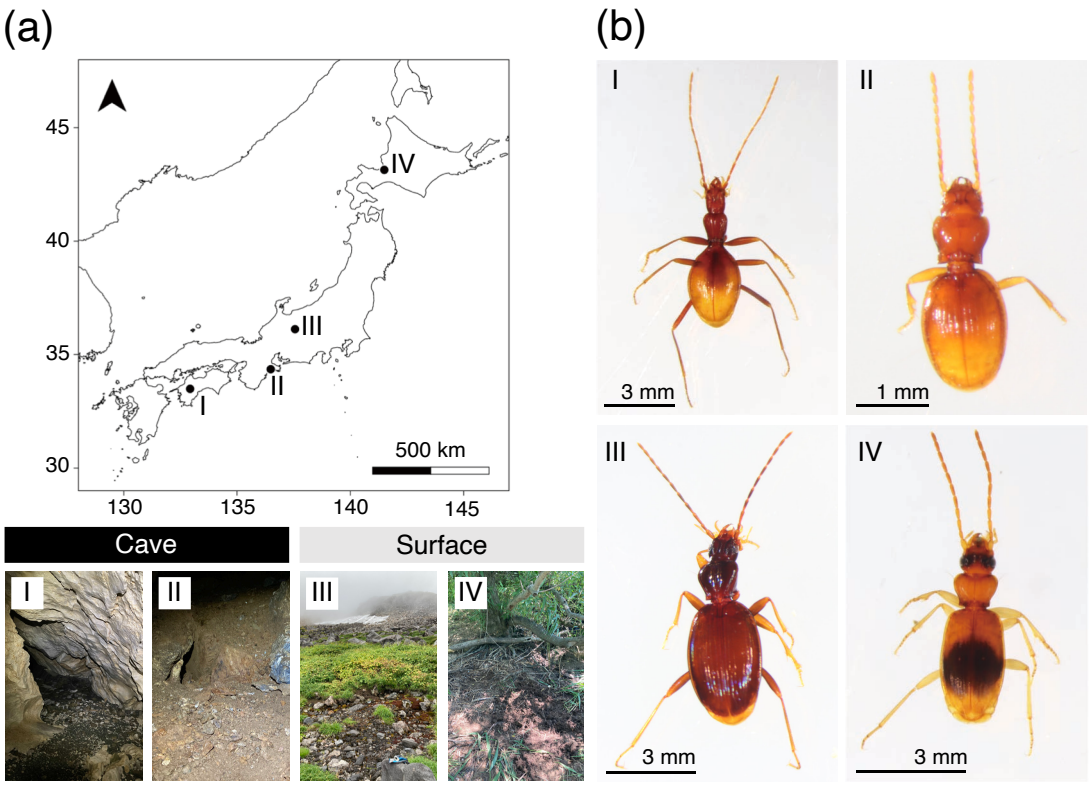
(a)
45
40
35
30
130
135
140
145
IV
III
II
I
500 km
Cave
Surface
I
II
III
IV
(b)
I
3 mm
II
1 mm
III
3 mm
IV
3 mm

Figure 1

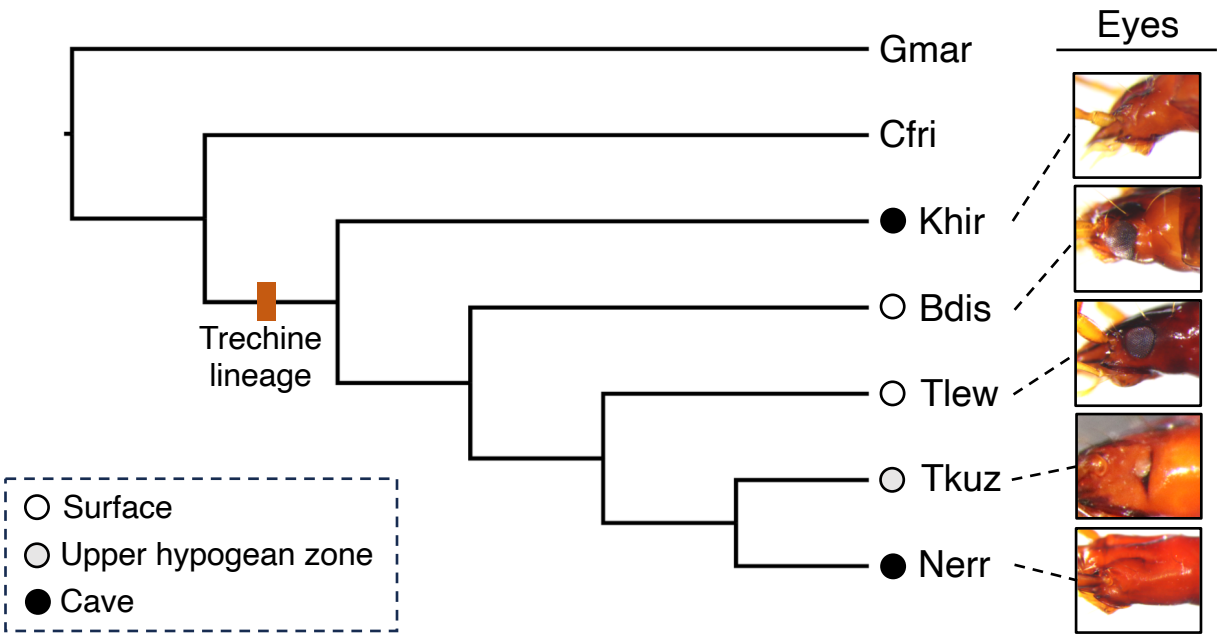
Eyes
Gmar
Cfri
Khir
Bdis
Tlew
Tkuz
Nerr
Trechine
lineage
Surface
Upper hypogean zone
Cave

Figure 2

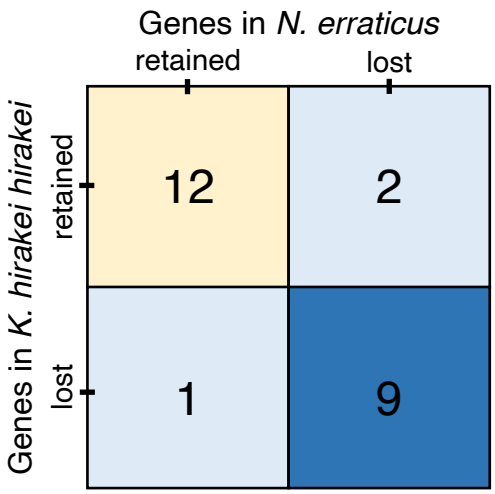
Genes in N. erraticus
retained
lost
Genes in K. hirakei hirakei
retained
lost
12
2
1
9

Figure 3

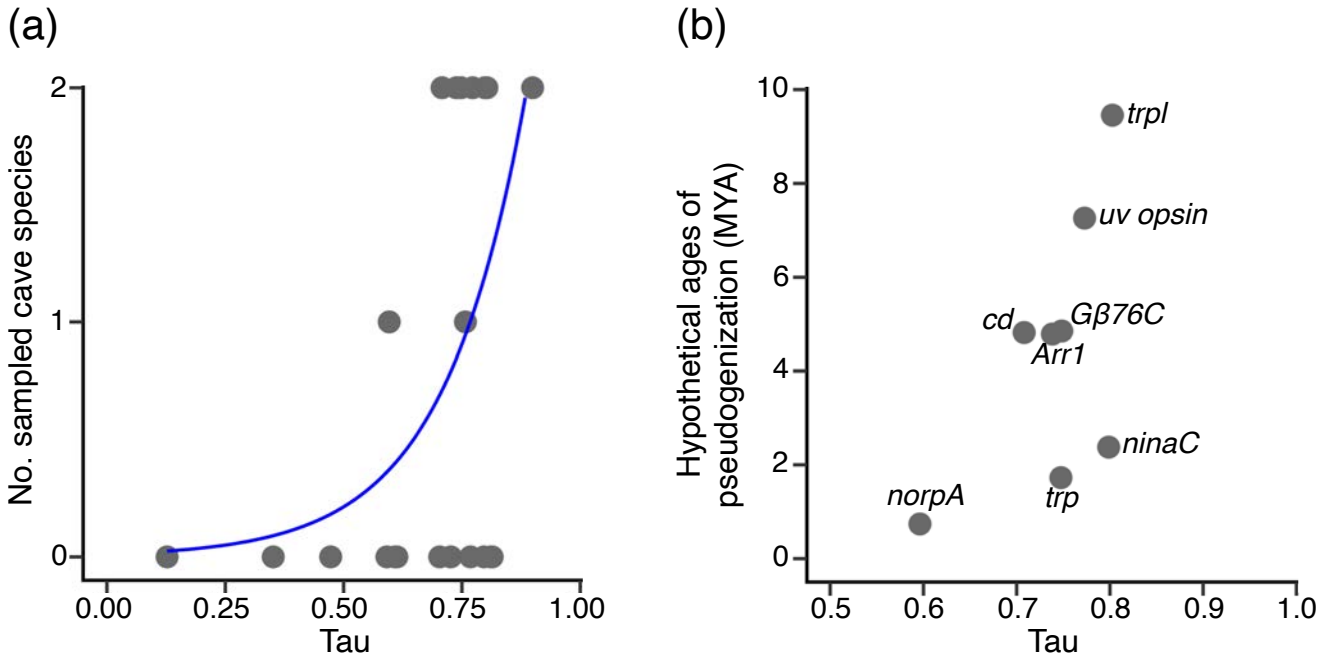
(a)
No. sampled cave species
2
1
0
0.00
0.25
0.50
0.75
1.00
Tau
(b)
Hypothetical ages of pseudogenization (MYA)
10
8
6
4
2
0
trpl
uv opsin
cd
Gβ76C
Arr1
ninaC
norpA
trp
0.5
0.6
0.7
0.8
0.9
1.0
Tau

Figure 4

Table 1

	N. erraticus	*K. hirakei hirakei*	*T. lewisi*	*B. discus*
Assembled genomes				
Total length (bp)	466,561,169	658,245,750	389,813,926	476,552,327
Scaffolds (n)	26,854	170,334	42,268	108,567
The longest scaffold (bp)	854,270	275,469	602,459	224,561
The shortest scaffold (bp)	390	408	378	390
Scaffold N50 (bp)	74,992	7,083	31,542	9,372
Estimated genome size (Mbp)	580	856	567	641
BUSCO pipeline analysis (% complete / partial)	98.39 / 99.20	87.27 / 97.00	97.07 / 99.05	92.03 / 98.39
Assembled transcripts				
Total length (bp)	76,867,119	46,810,738	68,858,705	95,318,759
Contigs (n)	64,296	61,818	55,949	79,009
The longest contig (bp)	54,033	12,989	44,022	31,941
The shortest contig (bp)	181	181	193	195
BUSCO pipeline analysis (% complete / partial)	97.88 / 98.61	80.76 / 91.95	97.51 / 98.46	98.39 / 99.20

Table 2

Query			Database: Genomes			
Name	Species	Accession number	*N. erraticus*	*K. hirakei hirakei*	*T. lewisi*	*B. discus*
Visual opsin proteins						
Lw opsin	Pcha	APY20654	no	no	yes	yes
Uv opsin	Pcha	APY20653	yes / pseudo	no	yes	yes
Non-visual opsin protein						
C-opsin	Tcas	NP_001138950	no	no	yes	yes
Phototransduction proteins						
Arr1	Tcas	XP_966595	yes / pseudo	no	yes	yes
Arr2	Tcas	NP_001164084	yes	no	yes	yes
Gα49B	Tcas	XP_966311	yes	yes	yes	yes
Gβ76C	Tcas	XP_973851	yes / pseudo	no	yes	yes
Gγ30A	Tcas	XP_972123	yes	yes	yes	yes
Gprk1	Tcas	XP_015838145	yes	yes	yes	yes
InaD	Tcas	XP_015837295	yes	yes	yes	yes
NinaA	Tcas	KYB27177	yes	yes	yes	yes
NinaC	Tcas	XP_015835531	yes / pseudo	no	yes	yes
NorpA	Tcas	XP_001812780	yes / pseudo	yes	yes	yes
Pkc53E	Dmel	NP_476682	yes	yes	yes	yes
Rab6	Tcas	XP_008200265	yes	yes	yes	yes / pseudo
RdgC	Tcas	XP_008192672	yes	yes	yes	yes
Trp	Tcas	XP_015835000	yes / pseudo	no	yes	yes
Trpl	Tcas	XP_968598	yes / pseudo	no	yes	yes
Eye pigmentation proteins						
V	Tcas	NP_001034499	yes	yes	yes	yes
KFase	Tcas	XP_008197380	yes	yes	yes	yes
Cn	Tcas	NP_001034500	no	yes	yes	yes
St	Tcas	AJD07060	yes	yes	yes	yes
W	Tcas	AAL56571	yes	yes	yes	yes
Cd	Tcas	XP_008200769	yes / pseudo	yes / pseudo	yes	yes

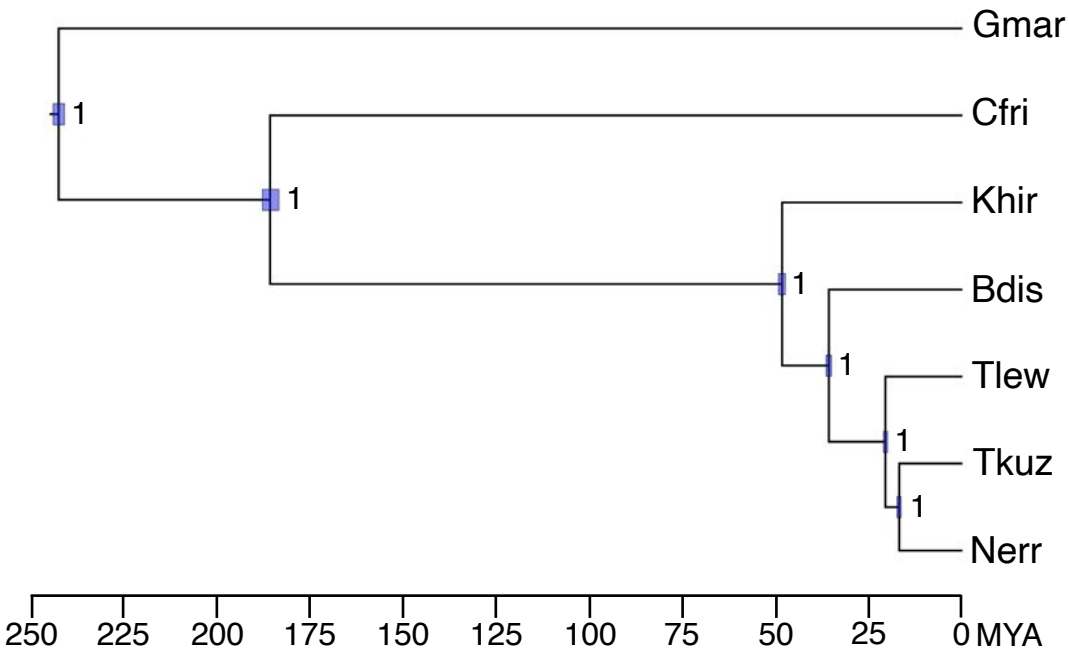
Gmar
Cfri
Khir
Bdis
Tlew
Tkuz
Nerr
1
1
1
1
1
1
250
225
200
175
150
125
100
75
50
25
0 MYA

Figure S1

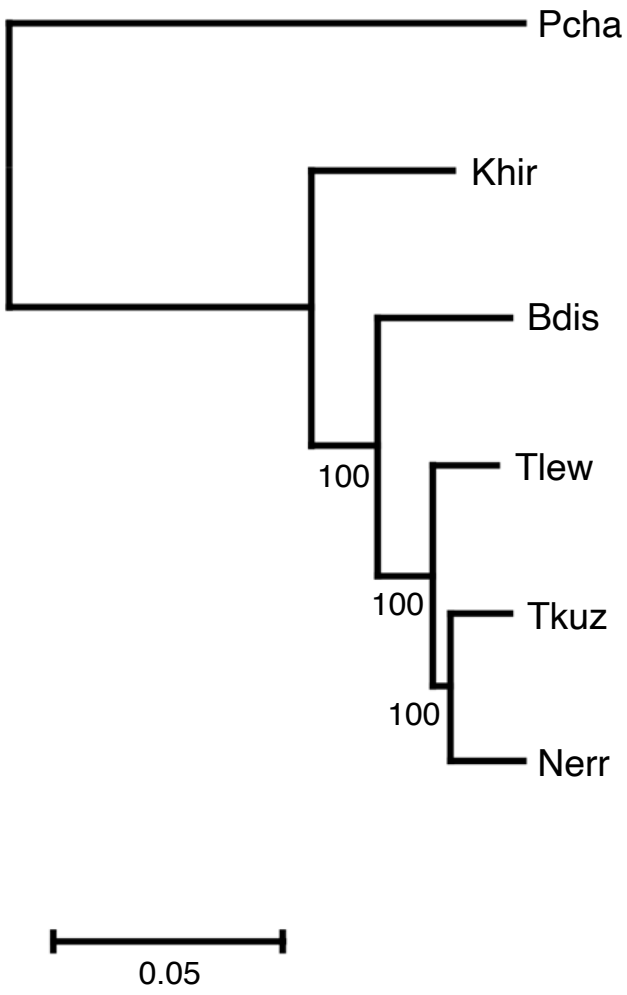

Figure S2

Table S1

	Genomic sequences				Transcript sequences			
	Forward reads		Reverse reads		Forward reads		Reverse reads	
Trimming	Before	After	Before	After	Before	After	Before	After
N. erraticus								
Total sequences (n)	384,831,742	380,994,697	384,831,742	380,994,697	26,660,748	26,531,161	26,660,748	26,531,161
Sequence length (bp)	151	15-150	151	15-150	151	15-150	151	15-150
GC (%)	31	30	31	30	43	43	43	42
K. hirakei hirakei								
Total sequences (n)	337,755,155	334,802,178	337,755,155	334,802,178	29,645,134	29,046,780	29,645,134	29,046,780
Sequence length (bp)	151	15-150	151	15-150	151	15-150	151	15-150
GC (%)	31	31	31	31	46	45	45	44
T. lewisi								
Total sequences (n)	438,139,657	432,451,653	438,139,657	432,451,653	23,023,302	22,906,877	23,023,302	22,906,877
Sequence length (bp)	151	15-150	151	15-150	151	15-150	151	15-150
GC (%)	30	30	30	30	43	43	43	42
B. discus								
Total sequences (n)	338,921,882	335,885,561	338,921,882	335,885,561	24,621,955	24,484,825	24,621,955	24,484,825
Sequence length (bp)	150	15-149	150	15-149	151	15-150	151	15-150
GC (%)	31	31	31	31	41	41	40	40

Table S2

Query			Database: Transcripts			
Name	Species	Accession number	*N. erraticus*	*K. hirakei hirakei*	*T. lewisi*	*B. discus*
Visual opsin proteins						
Lw opsin	Pcha	APY20654	no	no	yes	yes
Uv opsin	Pcha	APY20653	no	no	yes	yes
Non-visual opsin protein						
C-opsin	Tcas	NP_001138950	no	no	yes	yes
Phototransduction proteins						
Arr1	Tcas	XP_966595	yes	no	yes	yes
Arr2	Tcas	NP_001164084	no	no	yes	yes
Gα49B	Tcas	XP_966311	yes	yes	yes	yes
Gβ76C	Tcas	XP_973851	yes	no	yes	yes
Gγ30A	Tcas	XP_972123	yes	yes	yes	yes
Gprk1	Tcas	XP_015838145	no	yes	yes	yes
InaD	Tcas	XP_015837295	yes	yes	yes	yes
NinaA	Tcas	KYB27177	yes	yes	yes	yes
NinaC	Tcas	XP_015835531	no	no	yes	yes
NorpA	Tcas	XP_001812780	no	yes	yes	yes
Pkc53E	Dmel	NP_476682	yes	yes	yes	no
Rab6	Tcas	XP_008200265	yes	yes	yes	yes
RdgC	Tcas	XP_008192672	yes	yes	yes	yes
Trp	Tcas	XP_015835000	no	no	yes	yes
Trpl	Tcas	XP_968598	no	no	yes	yes
Eye pigmentation proteins						
V	Tcas	NP_001034499	yes	yes	yes	yes
KFase	Tcas	XP_008197380	yes	yes	yes	yes
Cn	Tcas	NP_001034500	no	yes	no	yes
St	Tcas	AJD07060	yes	no	yes	yes
W	Tcas	AAL56571	yes	no	yes	yes
Cd	Tcas	XP_008200769	yes	no	yes	yes

Table S3

Query			Database	
Name	Species	Accession number	Genome	Transcripts
Eye pigmentation proteins				
V	Tcas	NP_001034499	yes	yes
KFase	Tcas	XP_008197380	yes	yes
Cn	Tcas	NP_001034500	no	no
St	Tcas	AJD07060	yes	yes
W	Tcas	AAL56571	yes	yes
Cd	Tcas	XP_008200769	yes / pseudo	no

Table S4

Visual genes	Tau
lw opsin	0.899178909
uv opsin	0.772771028
Arr1	0.738277286
Arr2	0.757425177
Gα49B	0.472701056
Gβ76C	0.748336512
Gγ30A	0.612378484
Gprk1	0.35127037
inaD	0.811006321
ninaA	0.813783381
ninaC	0.798694972
norpA	0.596218624
Pkc53E	0.591337526
Rab6	0.127301308
rdgC	0.72633475
trp	0.747487312
trpl	0.802662665
v	0.607976614
KFase	0.703095672
st	0.768176184
w	0.796839318
cd	0.708014598