

SCIENTIFIC REPORTS



OPEN

Genome sequences identify three families of Coleoptera as morphologically derived click beetles (Elateridae)

Dominik Kusy¹, Michal Motyka¹, Matej Bocek¹, Alfried P. Vogler^{2,3} & Ladislav Bocak¹ 

Plastoceridae Crowson, 1972, Drilidae Blanchard, 1845 and Omalisidae Lacordaire, 1857 (Elateroidea) are families of the Coleoptera with obscure phylogenetic relationships and modified morphology showing neotenic traits such as soft bodies, reduced wing cases and larviform females. We shotgun sequenced genomes of *Plastocerus*, *Drilus* and *Omalisus* and incorporated them into data matrices of 66 and 4202 single-copy nuclear genes representing Elateroidea. Phylogenetic analyses indicate their terminal positions within the broadly defined well-sclerotized and fully metamorphosed Elateridae and thus Omalisidae should now be considered as Omalissinae stat. nov. in Elateridae Leach, 1815. The results support multiple independent origins of incomplete metamorphosis in Elateridae and indicate the parallel evolution of morphological and ecological traits. Unlike other neotenic elateroids derived from the supposedly pre-adapted aposematically coloured and unpalatable soft-bodied elateroids, such as fireflies (Lampyridae) and net-winged beetles (Lycidae), omalisids and drilids evolved from well-sclerotized click beetles. These findings suggest sudden morphological shifts through incomplete metamorphosis, with important implications for macroevolution, including reduced speciation rate and high extinction risk in unstable habitats. Precise phylogenetic placement is necessary for studies of the molecular mechanisms of ontogenetic shifts leading to profoundly changed morphology.

The elateroid beetles (Coleoptera: Elateroidea) are a morphologically heterogeneous group and include two different body plans: the well sclerotized ‘true elateroids’ (Elateridae, Eucnemidae, Throscidae), many of them able to jump for predator evasion using the long prosternal process fitting in the mesosternal cavity, and the soft-bodied ‘cantharoids’, some of them with larviform or incompletely metamorphosed females, all with short prosternum and unable to jump (Cantharidae, Lycidae, Lampyridae, Phengodidae, etc.). Taxonomists have accepted the paradigm of two reciprocally monophyletic groups corresponding to these types, the Cantharoidea and Elateroidea, until the late 20th century when the cantharoid families were recognized as a sub-lineage within the more broadly defined Elateroidea^{1,2}. Nevertheless, the concept of the ‘cantharoid’ clade was not questioned even after they were subsumed in the elateroids, on the assumption that soft-bodiedness originated only once^{3–5}. When molecular data became available, multiple origins of soft-bodied elateroid families were proposed⁶. Nowadays, both the soft-bodied and the hard-bodied groups are known to be composed of distantly related families^{7–11}.

The confusion about the relationships of soft-bodied groups also produced uncertainty for the placement of three small elateroid families, Omalisidae Lacordaire, 1857, Drilidae Blanchard, 1845 and Plastoceridae Crowson, 1972, which traditionally have been placed in the cantharoid clade^{3,4}. In the latest morphological study, Omalisidae, a family of only about twenty known species were recovered as closely related to Lampyridae, Phengodidae and Telegeusidae (the fireflies and glow-worms)³, whereas molecular studies placed them in various positions either as a sister group to the hard-bodied Elateridae or as a terminal lineage within Elateridae, but in all cases very distant from Lampyridae and Telegeusidae^{6,8,12}. Drilidae, the ‘false fireflies’, are a small group of predatory soft-bodied beetles. They have incompletely metamorphosed females with only head and appendages expressing the adult traits. Molecular data placed them in a derived position within Elateridae, specifically

¹Laboratory of Molecular Systematics, Department of Zoology, Faculty of Science, Palacky University, 17. listopadu 50, 771 46, Olomouc, Czech Republic. ²Department of Life Science, Natural History Museum, Cromwell Road, London, SW7 5BD, UK. ³Department of Life Science, Silwood Park Campus, Imperial College London Ascot, London, SL5 7BD, UK. Correspondence and requests for materials should be addressed to L.B. (email: ladislav.bocak@upol.cz)

in the subfamily Agrypninae, as the tribe Drilini^{7,13}. A recent analysis of basal coleopteran relationships that sampled both Drilidae and Omalidae recovered them together with cardiophorine click beetles, but the conclusions were problematic for splitting the core groups of Elateridae⁸. The third ‘family’ under consideration here, Plastoceridae, represented by *Plastocerus* Schaum, 1852, have been firmly placed within the soft-bodied cantharoids ever since their formal recognition¹. However, molecular data tentatively placed them within Elateridae near *Oxynopterus* Hope, 1842, leading to a proposed ranking as a subfamily of Elateridae¹⁴, in conflict with the earlier morphology-based analyses^{4,5}. All molecular studies have been using only Sanger data, i.e., a low number of mitochondrial and rRNA markers. The molecular findings are counterintuitive and have been met with skepticism. Therefore, most taxonomists still assign Drilidae the status of a family and do not accept the results of the molecular studies and proposed changes in the formal classification^{5,8,15–18}.

Among soft-bodied groups, several lineages exhibit larviform or incompletely metamorphosed females, which also occur in a few other beetles, e.g., *Thylocladius* Motschulsky, 1839, *Micromalthus*, Leconte, 1878 and *Ozopemon* Hagedorn, 1910, but they are most prevalent in Elateroidea^{6,19–24}. The distinct appearance generally led to their taxonomic recognition at family rank^{1,5,15}. Only females show the most extreme cases of larviform or partially metamorphosed morphology, although some traits may appear also in males^{25,26}. The point at which metamorphosis is prematurely terminated across elateroid lineages is variable. The neotenic females of net-winged beetles (Lycidae), glow-worm beetles (Rhagophthalmidae) and some genera of fireflies (Lampyridae) are completely larviform, except for the fully developed reproductive organs. Various fireflies have only an adult-like pronotum and head, in drilids only the head and appendages are adult-like, omalids exhibit a larviform abdomen, modified thorax and short elytra, and modifications in plastocerids are limited to the free abdominal sternites and lack the click mechanism^{14,27–34}. The neotenic differ not only in their morphology, but also in ecological traits, which affect macroevolution: the lineages with brachelytrous, wingless or completely larviform females occupy small ranges due to the low ability to disperse, they usually move only slowly, and are often unpalatable and aposematically coloured. Due to limited dispersal capacity, they are mostly limited to environmentally stable habitats and regions where populations persist with low risk of extinction. The lineages exhibiting neotenic females usually represent only a fraction of the species diversity compared to their fully winged sister groups. Most neotenic are uncommon and, therefore, often only males are known while the females are rarely seen or may not be known at all, but inferred to be neotenic based on certain modifications in the males^{1,7,19,20,34}. Taxa with incompletely metamorphosed females have generally been found to be related with soft-bodied lineages. Their characteristic traits, such as low dispersal capacity, restricted ranges and chemical protection as an alternative anti-predatory strategy, have been considered as pre-adaptations which increase the profitability of the shift to incomplete metamorphosis and higher investment in offspring^{19,20}.

The aim of this study is to produce genomic data for three enigmatic neotenic and morphologically aberrant lineages, Omalidae, Drilidae and Plastoceridae and use them to investigate their relationships to other elateroid families. As the previous studies provided ambiguous phylogenetic signal, whole genome data are the ultimate source of information which could shed light on their phylogenetic relationships. These elateroids are unique by their divergent morphology and relictual distribution. Their robust placement is crucial for future studies on the molecular mechanisms of incomplete metamorphosis leading to weakly sclerotized bodies and winglessness.

Results

The Fig. 1 shows the ML tree topology inferred from the 66-gene analysis at nucleotide level. The three focal taxa were found within Elateridae: *Plastocerus* associated with *Pectocera* (Oxynopterinae), *Drilus* associated with *Agrypnus* (Agrypninae) and *Omalisus* as a sister to the latter clades combined. The AU tests rejected alternative topologies (Table 1).

Newly generated shotgun genomic sequencing data provided high coverage of genomes at a sequencing depth of 69–312×. Data were used to create an ortholog set of 4202 genes from publicly available transcriptome and genome data of Coleoptera. The genome size was estimated to ~536 million base pairs (Mbp) for *D. mauritanicus*, 367 Mbp for *P. angulosus* and 270 Mbp for *O. fontisbellaquei* (Fig. S1). The genome completeness is summarized in Fig. 2A and the completeness of datasets in Figs S2–S7. The dataset based on 4202 orthologs recovers an almost identical topology at amino acid and nucleotide levels or filtering approach (Figs S8–S14). Taking into account the lower taxon sampling, *Omalisus* was recovered as sister to *Plastocerus*, unlike in the 66-gene dataset which placed it as sister to a wider clade that also includes *Drilus*, but with poor support. The genomic dataset includes only two elaterids, *Melanotus* and *Ignelater*, but they represent the two major clades of the family (Fig. 2). Because of the early branching of *Melanotus*, the three neotenic lineages are safely placed within the Elateridae (Fig. 2), with 100% support at all nodes.

Further exploration of the 66-gene dataset based on individual gene trees from individual loci and supernetwork analysis, using various subsets of taxa in addition to the Elateridae, always grouped the three focal groups within Elateridae, in positions very compatible with the supermatrix approach on the full taxon set (Fig. 2C–F). The Four cluster Likelihood Mapping (FcLM) analysis identified predominant support for the monophyly of all Elateridae including the three focal taxa and Rhagophthalmidae/Phengodidae + Lampyridae as a sister to them (Fig. 2J,K). Further ML topologies recovered from 66-taxa dataset, the coalescent trees recovered from both datasets and the tests of the alternative relationships are shown in Figs S15–S23.

Discussion

The current study is based on the most extensive transcriptomic dataset of ‘cantharoid’ families to date. At the base of the tree, the results recover three ‘cantharoid’ lineages, i.e. Cantharidae, Lycidae and the bioluminescent clade, composed of the Lampyridae, Rhagophthalmidae, and Phengodidae. The monophyly of the latter is in contrast to some studies that separated these three families and suggested that bioluminescence in adult Lampyridae has a separate evolutionary origin from the other two families with exclusively larval bioluminescence^{6,7}. It seems now

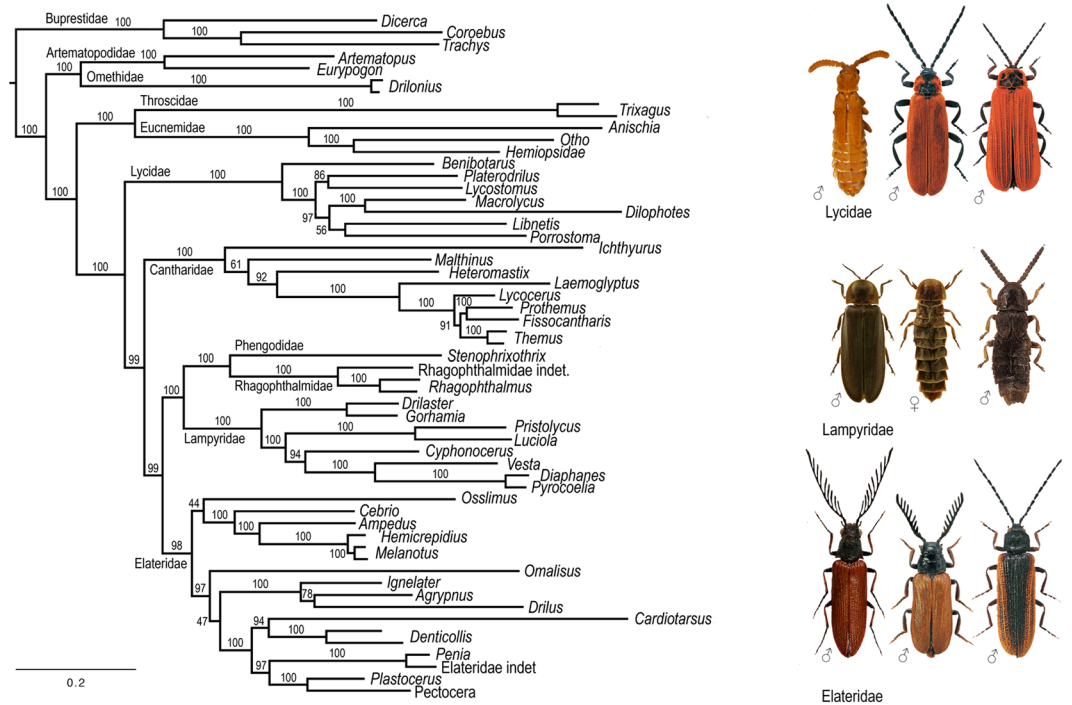


Figure 1. The maximum likelihood tree recovered from the 66-single copy protein coding genes at nucleotide level. Photographs of general appearance © authors.

Topology	logL	deltaL	p-AU
ML topology as in Fig. 1	-1035875	0.000	0.7247
Constraint:			
<i>Omalisus</i> + <i>Plastocerus</i> + <i>Drilus</i>			
sister to Elateridae	-1038925	3046.7	0.0001
<i>Drilus</i> sister to Cantharidae	-1037003	1124.9	0.0000
<i>Drilus</i> sister to Elateridae	-1036926	1048.2	0.0000
<i>Drilus</i> sister to Lampyridae	-1037133	1254.5	0.0000
<i>Drilus</i> sister to Lycidae	-1037070	1192.1	0.0000
<i>Omalisus</i> sister to Cantharidae	-1035963	84.5	0.0114
<i>Omalisus</i> sister to Elateridae	-1035942	63.8	0.0045
<i>Omalisus</i> sister to Lampyridae	-1036076	197.6	0.0000
<i>Omalisus</i> sister to Lycidae	-1036033	154.6	0.0004
<i>Plastocerus</i> sister to Cantharidae	-1038735	2857.3	0.0000
<i>Plastocerus</i> sister to Elateridae	-1037933	2054.8	0.0006
<i>Plastocerus</i> sister to Lampyridae	-1038902	3023.6	0.0000
<i>Plastocerus</i> sister to Lycidae	-1038975	3096.6	0.0000

Table 1. Approximately unbiased test of alternative relationships of *Drilus*, *Omalisus*, and *Plastocerus* recovered by the analysis of the 66-taxa dataset.

unlikely and the current topology is in agreement with eight-gene analysis⁸ and with morphology⁴. The widely defined Elateridae, including the three small ‘cantharoid’ lineages Drilidae, Omalisidae and Plastoceridae is the sister of the bioluminescent clade (Figs 1 and 2). The families traditionally assigned to the morphology-based Cantharoidea¹ acquired soft-bodiedness independently, or alternatively reverted to hard-bodied forms in the Elateridae (Fig. 1). Whatever the ancestral state, neotenic lineages may be derived both from soft-bodied and hard-bodied ancestors.

The morphology-based classifications have emphasized morphological divergence of Omalisidae, Drilidae, and Plastoceridae, and thus assigned them family rank^{4,5,8,15–18} although the possible relationships of *Omalisus* and Elateridae was mentioned already in the 19th century³⁵. These taxa share some characters with the ‘cantharoid’ soft-bodied elateroids (Fig. 2H,I) and even if not all traits are present¹⁴, morphological phylogenetic analyses have never found the proximity of these taxa to the well-sclerotized click beetles^{4,5}.

when numerous length-variable sequences were combined in a single dataset which contained a broad set of Coleoptera, mis-aligning the comparatively short loops in the rRNA genes of elaterids⁶. Similarly, low genetic divergence is characteristic for the protein coding nuclear genes of Elateridae, which might have contributed to the failure to recover click beetle monophyly in the recent analysis of Zhang *et al.*¹⁰. Therefore, more conclusive evidence was sought in the genomic datasets. We investigate two contradicting phylogenetic hypotheses: (1) drilids, plastocerids and omalids evolved within the elaterid clade, hence, they are in fact modified click beetles or (2) these lineages are deeply nested in Elateroidea and should be designated as families, as is still widely held.

The current study, based on tens of thousands nucleotide positions in the densely sampled 66-gene Elateroidea dataset and millions of positions in the genomic dataset, now confirms the placement of former Omalidae, Drilidae and Plastoceridae in Elateridae with high support (Figs 1 and 2) and under a wide array of analytical approaches (Figs S8–S20) which were performed to test their impact on the resulting topology³⁸. The topologies set *Drilus* and *Plastocerus* near *Agrypnus* and *Pectocera*, respectively, corroborating the earlier studies^{13,14}, and *Omalisus* is robustly recovered as a deeply rooted branch in Elateridae, although in variable positions (Figs 1 and 2). Deep splits¹¹ and rapid radiations^{39,40} often represent extremely difficult phylogenetic problems even when large datasets are analyzed (here, the unstable position of *Omalisus* or a conflicting signal for *Melanotus*; Figs S11–S21). We found that although the genomic and 66-gene datasets place *Omalisus* in Elateridae, the later cannot robustly identify its sister clade within this family (Figs 2, S15–S17, S20). Therefore, we studied in detail alternative positions of focal taxa and we found that the alternative hypotheses that force well-sclerotized elaterids to be monophyletic, or *Drilus*, *Omalisus*, and *Plastocerus* to be a sister to either Lampyridae, Cantharidae, Lycidae or Elateridae are all rejected by AU tests (Table 1). Similarly, the monophyly of all Elateridae including three focal families got predominant support (Fig. 2J,K). To sum up, it becomes difficult to dispute the possibility that the lineages with modified morphology (*Plastocerus*) and even with incompletely metamorphosed females (*Drilus*, *Omalisus*) are nested within the Elateridae. Thus, these taxa do not deserve the family rank despite their morphological uniqueness. Based on the current analysis, we propose to lower Omalidae from the family rank to a subfamily Omalidae Lacordaire, 1857 in Elateridae and we confirm the status of Drilini and Plastocerinae^{13,14}.

The shift to incompletely metamorphosed females (Fig. 2H,I) is known to affect the macroevolution of a lineage¹⁹. The well-sclerotized elaterids contain the substantial part of the extant Elateroidea diversity, i.e. ~10,000 species³⁷, and the morphologically modified elaterids are represented only by Plastocerinae (2 sp), Omalidae (22 spp) and Drilini (>100 spp). Their evolutionary trajectory is determined by low species numbers, limited geographic ranges, and small population sizes^{14,34,41,42}. Most omalids and *Plastocerus angulosus* occur in the Mediterranean only and their ranges are usually restricted to coastal refugia where broad leaf forests persisted during the last glacial maximum, indicating the constraints to their subsequent dispersal^{14,34}. Drilini are widespread in the Afrotropical region, the Mediterranean and along the South Asian coast to Thailand, but similarly to omalids they are known to have limited species ranges and are generally rare in ecosystems⁴¹. Hence, omalids, drilids and plastocerids represent examples of lineages which diverted from a successful body-plan of click beetles whose morphology did not substantially change since the Late Triassic⁴³ and which has been reproduced in thousands of extant species. Yet, they survive for a long time in stable ecosystems similarly to other neotenic elateroids^{19,32,33}.

The secure placement of these three lineages within the true click beetles has important implications for the origin of neoteny and macroevolution of modified lineages. First, the derived positions within Elateridae demonstrates that astonishing differences in morphology can be achieved through what appears to be shifts in ontogenetic programs. The general appearance and individual morphological traits changed over short evolutionary time scales, and these differences lead to convergent morphologies observed in several unrelated groups throughout the Elateroidea. Second, these morphological shifts also have profound macroevolutionary consequences. The traditional placement suggested relationships of Elateridae, Eucnemidae and Throscidae², and Elateridae became the dominant part of Elateroidea diversity with >40% of species, world-wide distribution and high dominance in many beetle communities since the Jurassic⁴³. Strong body sclerotization and an effective escape mechanism apparently proved to be an evolutionarily successful design, to make Elateridae an example of a morphologically conservative lineage whose great species diversity was not accompanied by morphological diversity^{43,44}. Yet, among these conservative groups some lineages arose which were very different morphologically and resemble distant neotenic relatives in Lycidae, Lampyridae and Rhagophthalmidae. It has generally been assumed that soft-bodiedness in Elateroidea might be the first symptom of unfinished metamorphosis^{19,30} raising the possibility for the more fully neotenic lineages raised within them. Soft-bodied elateroids move slowly, are commonly unpalatable^{45,46} and aposematically coloured⁴⁷. These factors might increase the trade-off gains if dispersal capacity is further lowered in favor of higher fecundity due to incomplete metamorphosis²⁰. Neotenic nested in click beetles now falsify the general validity of this hypothesis about ecological pre-adaptations required for the origin of arrested or prematurely terminated metamorphosis²⁰. If there are life-history pre-adaptations for incomplete metamorphosis present in Elateroidea, they differ in individual groups and might not have only the ecological character. Based on the origins of neoteny in click beetles, the alternative hypothesis can be formulated: a rapid modification of the regulatory system of insect metamorphosis at molecular level^{48,49} might be a trigger for a subsequent improvement by the natural selection, such as large-bodied females, higher fecundity and evolution of alternative defensive strategies. To sum up, we can say that convergent evolution has produced multiple lineages that 'replay life's tape'^{50,51} exploring a life strategy that is favored by stable climatic and ecological conditions, and which can apparently start from different ancestral states. As this shift is triggered by prematurely arrested metamorphosis, the repeated origin may indicate the low stability of the molecular regulatory system in Elateroidea. The identification of the closest relatives of neotenic may therefore be the first step for a comparative biology of the mechanisms of metamorphosis.

Methods

Material, laboratory procedures, and draft genomes. Total genomic DNA was extracted from single adult specimens of *Omalisus fontisbellaquei* Geoffroy, 1785 (from Czechia), *Drilus mauritanicus* Lucas, 1842 (Spain) and *Plastocerus angulosus* Schaum, 1852 (Turkey) using the DNeasy kit (Qiagen Inc., Hilden, Germany). The voucher specimens have been deposited in the collection of Department of Zoology, Palacký University, Olomouc. Genomic DNA of all three specimens was shotgun sequenced on the Illumina X Ten platform (Illumina Inc., San Diego, CA) for 2×150 bp paired-end reads by Novogene Co., Ltd. (Beijing, China) and each individual was sequenced for $30\text{--}107 \times 10^9$ base pairs. Raw paired-end reads were filtered using fastp 0.13.2⁵² under the following parameters `-q 5 -u 50 -l 50 -n 15` and other settings as default. The filtering steps included the removal of read pairs if either one read contains adapter contamination; if the proportion of low quality bases is over 50%; or if either one read contains more than 15 N bases. The quality of reads was visualized with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Sequence data were deposited in GenBank SRA (Accession Numbers AB123456-AB123456).

We performed k-mer counts on the filtered data in Jellyfish 2.2.7⁵³ using 31-mer sizes. Moreover, based on the distribution of k-mer occurrences, we estimated the genome size using GenomeScope⁵⁴ and assembled draft genomes of *O. fontisbellaquei*, *D. mauritanicus* and *P. angulosus* using MEGAHIT 1.1.3^{55,56}, with all parameters set to default and k-mer sizes of 31, 59, 87, 115 and 143. Additionally, the genome of *O. fontisbellaquei* was processed using the Shovill pipeline (<https://github.com/tseemann/shovill>) and assembled with SPAdes 3.12.0⁵⁷ using k-mer size 31, 51, 71, 91 and 111 under default parameters. We produced statistics of draft genomes with the assembly-stats algorithm (<https://github.com/rjchallis/assembly-stats>) and the results of both methods used for *O. fontisbellaquei* were similar and are shown in Fig. S24. The SPAdes contigs were used for further analyses of *Omalisus*. Obtained contig sequences were used to train Augustus⁵⁸ for species specific gene models with BUSCO 3⁵⁹, -long option, Endopterygota set of conserved genes ($n = 2442$) and -sp tribolium2012 as the closest relative. Predicted species specific gene models were then used for *ab initio* gene predictions in Augustus, and predicted protein coding sequences were used for subsequent analyses in Orthograph 0.6.1⁶⁰. The genome and coding gene set completeness was evaluated based on the predicted protein sets with BUSCO using expected 2442 Endopterygota single-copy orthologs as targets. BUSCO quantitatively assesses completeness using evolutionary conserved expectations of gene content. We compared completeness of predicted protein sets with *Agrilus planipennis* and *Iguelater luminosus*.

Data matrices. Two datasets were assembled for the phylogenetic analysis:

(1) Single-copy genes – 66-taxa dataset (Table S1). This dataset is based on PCR amplified sequences for 95 genes across Coleoptera¹⁰, of which 66 gene for 54 taxa were retained after the removal of 29 supposedly multi-copy genes, as described earlier¹¹. The putative homologs for *O. fontisbellaquei*, *D. mauritanicus* and *P. angulosus* were added to the earlier published data. Exons were concatenated to produce a supermatrix of 53,253 aligned positions.

(2) Genome orthologs. Transcriptomes of *Melanotus cribricollis*⁶¹; *Asymmetricata circumdata*, *Aquatica ficta*, *Pyrocoelia pectoralis*, *Rhagophthalmus* sp⁶²; *Chauliognathus flavipes* and *Phrixothrix hirtus*⁶³ were downloaded from the NCBI SRA archive and assembled as described by Kusy *et al.*¹¹. Additionally, we downloaded the gene set of *I. luminosus* from fireflybase.org⁶⁴ and the transcriptome of *Photinus pyralis*⁶⁵ from NCBI Transcriptome Shotgun Assembly database (Table S2). The ortholog set was obtained by searching the OrthoDB 9.1 database⁶⁶ for one-to-one orthologs among Coleoptera in available genome sequences of *A. planipennis*⁶⁷, *Anoplophora glabripennis*⁶⁸, *Dendroctonus ponderosae*⁶⁹, *Leptinotarsa decemlineata*⁶⁷, *Onthophagus taurus*⁶⁷, and *Tribolium castaneum*^{70,71} (Tables S3, S4). OrthoDB 9.1 specified 4225 protein coding single copy genes for the above species and the Coleoptera reference node. We used Orthograph 0.6.1⁶⁰ to search the above transcriptomes and predicted protein coding gene sets for the corresponding sequences. Default settings were used. We summarized Orthograph data reporting 4202 orthologs, removed terminal stop codons and masked internal stop codons at the translational level and nucleotide levels using the perl script summarize_orthograph_results.pl⁶⁰. The amino acid sequences were aligned using MAFFT 7.394 with the *L-INS-i* algorithm⁷². Resulting alignments from each ortholog group were checked for the presence of outliers using the script https://github.com/mptrsen/scripts/blob/master/outlier_check.pl and following the methods reported by Misof *et al.*⁷³ and Peters *et al.*⁷⁴. We used Pal2Nal⁷⁵ to generate multiple sequence alignments of nucleotides corresponding to amino acids. Aliscore 2.0^{76,77} with the maximal number of pairwise comparisons, -e option and default settings were used to identify random or ambiguous similarity within alignments which were masked using Alicut 2.3 (https://github.com/mptrsen/scripts/blob/master/ALICUT_V2.3.pl) and Alinuc.pl⁷³ to apply Aliscore results to the nucleotide data. The dataset of 4202 orthologs on amino acid and nucleotide levels were concatenated into supermatrices 1 and 2 (1–amino acids; 2–all nucleotides) with FASconCAT-G⁷⁸. Additional filtered datasets were used for tree construction: (a) nucleotides, 1st + 2nd codon positions only; (b) amino acid raw data (no filtering e.g. outliers removal, Aliscore); (c) nucleotide raw data; (d) a subset of Supermatrix 1, orthologs available for all taxa; (e) a subset of Supermatrix 2 containing orthologs for all taxa. AliStat 1.7 (<https://github.com/thomaskf/AliStat>) was used to generate distributions of missing data per site in the supermatrices.

Phylogenetic analyses. IQ-TREE 1.6.6⁷⁹ and RaxML⁸⁰ were used to calculate maximum likelihood (ML) trees, with partitions identified by the Model Finder tool of IQ-TREE and using the Bayesian Information Criterion^{81,82}. The partitions, models and parameters are available upon request. The ultrafast bootstrap option was used with 3000 bootstrap iterations⁸³. The IQ-TREE analyses were run with the -spp parameter allowing each partition to have its own evolutionary rate.

To investigate alternative and/or confounding signal in the 66-taxa dataset and genomic dataset, we used FcLM analysis^{73,84} implemented in IQ-TREE to study a possibility of the occurrence of incongruent signal in

phylogenomic datasets that might not be revealed by a phylogenetic multi-species tree. Additionally, gene tree incongruence in 66 genes dataset was tested by visualizations of the dominant bipartitions among individual loci based on the individual IQ-TREE ML gene topologies by constructing supernetworks using the SuperQ method implemented in Spectre selecting the ‘balanced’ edge-weight with ‘JOptimizer’ optimization function, and applying no filter^{85,86}. This methodology decomposes all gene trees into quartets to build supernetworks where edge lengths correspond to quartet frequencies. We tested the alternative tree topologies within Elateridae and three focal taxa *Plastocerus*, *Drilus*, and *Omalisus* and further, we tested a potential ambiguity of relationships of all these taxa and three putative relatives, i.e., Cantharidae, Lycidae, and Rhagophthalmidae. Resulting supernet-works were visualized in SplitsTree 4.14.6⁸⁷. Further, we used ASTRAL 5.6.1⁸⁸ and genomic dataset to construct coalescent species trees from individual IQ-TREE ML gene topologies at amino acid and nucleotide level.

The focal taxa, *Plastocerus*, *Drilus* and *Omalisus* have been placed in relationships with soft-bodied ‘cantharoid’ families^{1–5}. We tested these alternative hypotheses using the densely sampled 66-taxa dataset and compared likelihoods of hypothesized relationships of focal taxa with Cantharidae, Lampyridae and Lycidae and alternatively the topology (*Omalisus*,(*Plastocerus*, *Drilus*)),(Elateridae) which accepts their relationships with Elateridae, but excludes them as sister lineages of Elateridae. The likelihood of these topologies was compared with the best ML topology using AU test⁸⁹ implemented in IQTREE⁷⁹ and using -au option and 1,000,000 replicates.

The click beetle *Melanotus* was earlier recovered as a lineage distantly related to other well-sclerotized Elateridae^{8,10}. We encountered a similarly contentious position of *Melanotus*. The topology recovered from the dataset at amino acid level suggested that *Melanotus* does not belong to click beetles^{8,10}. Therefore, we estimated the number of genes supporting the alternative relationships. We tested positions of *Melanotus* as (1) a sister to Lampyridae and (2) a sister to other Elateridae by evaluating which single gene partition of amino acid data favor alternative topologies by calculating log-likelihood scores for each gene partition using IQ-TREE option -wpl. As an input we provided both topologies. To interpret the results of the partition log-likelihood and to evaluate the contribution of each gene partition, we calculated differences of each pL score of topologies^{90,91}.

Data Accessibility

The DNA sequences reported in this article can be accessed in GenBank under accessions AB123456-789.

References

- Crowson, R. A. A review of the classification of Cantharoidea (Coleoptera), with definition of two new families Cneoglossidae and Omethidae. *Rev. Univ. Madrid* **21**, 35–77 (1972).
- Lawrence, J. F. Rhinorhipidae, a new beetle family from Australia, with comments on the phylogeny of the Elateriformia. *Invertebr. Taxon.* **2**, 1–53 (1988).
- Beutel, R. G. Phylogenetic analysis of Elateriformia (Coleoptera: Polyphaga) based on larval characters. *J. Zool. Syst. Evol. Res.* **33**, 145–171 (1995).
- Branham, M. A. & Wenzel, J. W. The evolution of photic behaviour and the evolution of sexual communication in fireflies (Coleoptera: Lampyridae). *Cladistics* **84**, 565–586 (2003).
- Lawrence, J. F. *et al.* Phylogeny of the Coleoptera based on morphological characters of adults and larvae. *Ann. Zool.* **61**, 1–217 (2011).
- Bocakova, M., Bocak, L., Hunt, T., Teravainen, M. & Vogler, A. P. Molecular phylogenetics of Elateriformia (Coleoptera): evolution of bioluminescence and neoteny. *Cladistics* **23**, 477–496 (2007).
- Kundrata, R., Bocakova, M. & Bocak, L. The comprehensive phylogeny of the superfamily Elateroidea (Coleoptera: Elateriformia). *Mol. Phyl. Evol.* **76**, 162–171 (2014).
- McKenna, D. D. *et al.* The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst. Entomol.* **40**, 835–880 (2015).
- Timmermans, M. J. T. N. *et al.* Family-Level Sampling of Mitochondrial Genomes in Coleoptera: Compositional Heterogeneity and Phylogenetics. *Genome Biol. Evol.* **8**, 161–175 (2016).
- Zhang, S. Q. *et al.* Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Comm.* **9**, 205 (2018).
- Kusy, D. *et al.* Genome sequencing of *Rhinorhipus* Lawrence exposes an early branch of the Coleoptera. *Front. Zool.* **15**, 21 (2018).
- Hunt, T. *et al.* A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913–1916 (2007).
- Kundrata, R. & Bocak, L. The phylogeny and limits of Elateridae (Insecta, Coleoptera): is there a common tendency of click beetles to soft-bodiedness and neoteny? *Zool. Scr.* **40**, 364–378 (2011).
- Bocak, L., Motyka, M., Bocek, M. & Bocakova, M. Incomplete sclerotization and phylogeny: The phylogenetic classification of *Plastocerus* (Coleoptera: Elateroidea). *PLoS One* **13**, e0194026 (2018).
- Beutel, R. G. & Leschen, R. A. B. *Coleoptera, Beetles*; Volume 1: Morphology and Systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim) 2nd edition. In: Kristensen N. P. & Beutel R. G., eds. *Handbook of Zoology, Arthropoda: Insecta*. Berlin/New York: Walter de Gruyter GmbH & Co. (2016).
- Amaral, D. T. *et al.* Transcriptional comparison of the photogenic and non-photogenic tissues of *Phrixothrix hirtus* (Coleoptera: Phengodidae) and non-luminescent *Chauliognathus flavipes* (Coleoptera: Cantharidae) give insights on the origin of lanterns in railroad worms. *Gene Rep.* **7**, 78–86 (2017).
- Toussaint, E. *et al.* The peril of dating beetles. *Syst. Entomol.* **42**, 1–10 (2017).
- Martin, G. J., Branham, M., Whiting, M. F. & Bybee, S. M. Total evidence phylogeny and the evolution of adult bioluminescence in fireflies (Coleoptera: Lampyridae). *Mol. Phyl. Evol.* **107**, 564–575 (2017).
- Bocak, L., Bocakova, M., Hunt, T. & Vogler, A. P. Multiple ancient origins of neoteny in Lycidae (Coleoptera): consequences for ecology and macroevolution. *Proc. R. Soc. B* **275**, 2015–2023 (2008).
- McMahon, D. P. & Hayward, A. Why grow up? A perspective on insect strategies to avoid metamorphosis. *Ecol. Entomol.* **41**, 505–515 (2016).
- Gould, S. J. *Ontogeny and Phylogeny*. Cambridge: Harvard University Press (1977).
- Pollock, D. A. & Normark, B. B. The life cycle of *Micromalthus debilis* LeConte (1878) (Coleoptera: Archostemata: Micromalthidae): historical review and evolutionary perspective. *J. Zool. Syst. Evol. Res.* **40**, 105–112 (2002).
- Jordal, B. H., Beaver, R. A., Normark, B. B. & Farrell, B. D. Extraordinary sex ratios and the evolution of male neoteny in sib-mating *Ozopemon* beetles. *Biol. J. Linn. Soc.* **75**, 353–360 (2002).
- Kiselyova, T. & McHugh, J. V. A phylogenetic study of Dermestidae (Coleoptera) based on larval morphology. *Syst. Entomol.* **31**, 469–507 (2006).

25. Bocak, L., Grebennikov, V. V. & Masek, M. A new species of *Dexoris* (Coleoptera: Lycidae) and parallel evolution of brachyptery in the soft-bodied elateroid beetles. *Zootaxa* **3721**, 495–500 (2013).
26. Naoki, T., Bocak, L. & Ghani, I. A. Discovery of a new species of the brachelytrous net-winged beetle genus *Alyculus* (Coleoptera: Lycidae) from Peninsular Malaysia. *Zootaxa* **4144**, 145–150 (2015).
27. Wong, A. T. C. A new species of neotenus beetle, *Duliticolia hoiseni* (Insecta: Coleoptera: Cantharoidea: Lycidae) from Peninsular Malaysia and Singapore. *Raffl. Bull. Zool.* **44**, 173–187 (1996).
28. Cicero, J. M. Ontophylogenetics of cantharoid larviforms (Coleoptera: Cantharoidea). *Coleopt. Bull.* **42**, 105–151 (1988).
29. Miller, R. S. A revision of the *Leptolycini* (Coleoptera: Lycidae) with a discussion of paedomorphosis. PhD Thesis. Columbus: The Ohio State University (1991).
30. Jeng, M. L. *Comprehensive phylogenetics, systematics, and evolution of neoteny of Lampyridae* (Insecta: Coleoptera). PhD thesis, Lawrence: University of Kansas (2008).
31. Bocak, L. & Brlík, M. Revision of the family Omalidae (Coleoptera, Elateroidea). *Ins. Syst. Evol.* **39**, 189–212 (2008).
32. Masek, M., Ivie, M. A., Palata, V. & Bocak, L. Molecular phylogeny and classification of Lyropaeini (Coleoptera: Lycidae) with description of larvae and new species of *Lyropaeus*. *Raffl. Bull. Zool.* **62**, 136–145 (2014).
33. Masek, M., Palata, V., Bray, T. C. & Bocak, L. Molecular phylogeny reveals high diversity, geographic structure and limited ranges in neotenic net-winged beetles *Platerodrilus* (Coleoptera: Lycidae). *PLoS One* **10**, e0123855 (2015).
34. Bocek, M., Fancello, L., Motyka, M., Bocakova, M. & Bocak, L. The molecular phylogeny of Omalidae (Coleoptera) defines the family limits and demonstrates low dispersal propensity and the ancient vicariance patterns. *Syst. Entomol.* **43**, 250–261 (2018).
35. Bourgeois, J. Monographie des Lycides de l'ancien-monde. *L'Abeille* **20**, 1–120 (1882).
36. Bocak, L., Kundrata, R., Andújar-Fernández, C. & Vogler, A. P. The discovery of Iberobaeniidae (Coleoptera: Elateroidea): a new family of beetles from Spain, with immatures detected by environmental DNA sequencing. *Proc. R. Soc. B.* **283**, 20152350 (2016).
37. Bocak, L. *et al.* Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Syst. Entomol.* **39**, 97–110 (2014).
38. Reddy, S. *et al.* Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst. Biol.* **66**, 857–879 (2017).
39. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
40. Prum, R. O. *et al.* A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573 (2015).
41. Bray, T. C. & Bocak, L. Slowly dispersing neotenic beetles can speciate on a penny coin and generate space-limited diversity in the tropical mountains. *Sci. Rep.* **6**, 33579 (2016).
42. Kundrata, R. & Bocak, L. Taxonomic review of Drilini (Elateridae: Agrypninae) in Cameroon reveals high morphological diversity, including the discovery of five new genera. *Ins. Syst. Evol.* **48**, 441–492 (2017).
43. Ponomarenko, A. G. *The geological history of beetles*. Pp. 155–171. In *Biology, Phylogeny, and Classification of Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson* (eds by Pakaluk, J. and Slipinski, S. A.). Warszawa: Muzeum i Instytut Zoologii PAN (1995).
44. Doludenko, M. P., Ponomarenko, A. G. & Sakulina, G. V. *La géologie du gisement unique de la faune et de la flore du jurassique supérieur d'Aulié (Karatau, Kazakhstan du Sud)*. Moscow: Academie des Sciences de l'URSS, Inst. Géologique (1990).
45. Moore, B. P. & Brown, W. V. Identification of warning odour components, bitter principles and antifeedants in an aposematic beetle—*Metriorrhynchus rhipidium* (Coleoptera: Lycidae). *Ins. Biochem.* **15**, 493–499 (1981).
46. Eisner, T. *et al.* Defensive chemistry of lycid beetles and of mimetic cerambycid beetles that feed on them. *Chemoeology* **18**, 109–119 (2008).
47. Motyka, M., Kampova, L. & Bocak, L. Phylogeny and evolution of Müllerian mimicry in aposematic *Dilophotes*: evidence for advergence and size-constraints in evolution of mimetic sexual dimorphism. *Sci. Rep.* **8**, 3744 (2018).
48. Riddiford, L. M. How does juvenile hormone control insect metamorphosis and reproduction? *Gen. Comp. Endocrin.* **179**, 477–484 (2012).
49. Jindra, M., Palli, S. R. & Riddiford, L. M. The Juvenile Hormone Signaling Pathway in Insect Development. *Ann. Rev. Entomol.* **58**, 181–204 (2013).
50. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: Norton (1989).
51. Beatty, J. Replaying Life's Tape. *J. Philos.* **103**, 336–362 (2006).
52. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor, <https://www.biorxiv.org/content/biorxiv/early/2018/04/09/274100.full.pdf> (accessed on April 30th, 2018) (2018).
53. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
54. Vurtture, G. W. *et al.* GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
55. Li, D., Liu, C., Luo, R., Sadakane, K. & Lam, T. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
56. Li, D. *et al.* MEGAHITv1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
57. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
58. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl. 2), 215–225 (2003).
59. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2017).
60. Petersen, M. *et al.* Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics* **18**, 111 (2017).
61. Ye, B., Zhang, Y., Shu, J., Wu, H. & Wang, H. RNA-sequencing analysis of fungi-induced transcripts from the bamboo wireworm *Melanotus cribricollis* (Coleoptera: Elateridae) larvae. *PLoS One* **13**, e019118 (2018).
62. Wang, K., Hong, W., Jiao, H. & Zhao, H. Transcriptome sequencing and phylogenetic analysis of four species of luminescent beetles. *Sci. Rep.* **7**, 1814 (2017).
63. Amaral, D. T., Mitani, Y., Ohmiya, Y. & Viviani, V. R. Organization and comparative analysis of the mitochondrial genomes of bioluminescent Elateroidea (Coleoptera: Polyphaga). *Gene* **586**, 254–262 (2016).
64. Fallon, T. R. *et al.* 2017 Firefly genomes illuminate parallel origins of bioluminescence in beetles, <https://www.biorxiv.org/content/biorxiv/early/2018/02/25/237586.full.pdf>. Accessed on June 22nd, 2018.
65. Fallon, T. R., Li, F., Vicent, M. A. & Weng, J. Sulfoluciferin is Biosynthesized by a Specialized Luciferin Sulfoltransferase in Fireflies. *Biochemistry* **55**, 3341–3344 (2016).
66. Zdobnov, E. M. *et al.* OrthoDBv9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2016).
67. Poelchau, M. *et al.* The i5k Workspace@NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res.* **43**(Database issue), D714–719 (2015).

68. McKenna, D. D. *et al.* Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biol.* **17**, 227 (2017).
69. Keeling, C. I. *et al.* Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biol.* **14**, R27 (2013).
70. Richards, S. *et al.* Tribolium genome sequencing consortium). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
71. Shelton, J. M. *et al.* Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. *BMC Genomics* **16**, 734 (2015).
72. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
73. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763 (2014).
74. Peters, R. S. *et al.* Evolutionary History of the Hymenoptera. *Curr. Biol.* **27**, 1013–1018 (2017).
75. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
76. Misof, B. & Misof, K. A. Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* **58**, 21–34 (2009).
77. Kück, P. *et al.* Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* **7**, 10 (2010).
78. Kück, P. & Longo, G. C. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* **11**, 81 (2014).
79. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
80. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
81. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
82. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).
83. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
84. Strimmer, K. & von Haeseler, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci.* **94**, 6815–6819 (1997).
85. Grunewald, S., Spillner, A., Bastkowski, S., Bogershausen, A. & Moulton, V. SuperQ: computing Supernetworks from quartets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 151–160 (2013).
86. Bastkowski, S. *et al.* SPECTRE. A suite of PhylogEnetiC tools for reticulate evolution. *Bioinformatics* **34**, 1056–1057 (2017).
87. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2005).
88. Sayyari, E. & Mirarab, S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol. Biol. Evol.* **33**, 1654–1668 (2016).
89. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
90. Simon, S., Blanke, A. & Meusemann, K. Reanalyzing the Palaeoptera problem - The origin of insect flight remains obscure. *Arthropod Struct. Dev.* **47**, 328–338 (2018).
91. Shen, X. X., Hittinger, C. T. & Rokas, A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecol. Evol.* **1**, 126 (2017).

Acknowledgements

The authors are obliged to V. Kuban (Brno) for information on *P. angulosus*, R. Kundrata and M. Baena for the specimen of *D. mauritanicus*. The study was funded by the IGA and GACR projects (PrF-2018, 18-14942S to L.B., M.Ma., M.Mo., and D.K.), Leverhulme Trust (F/00696/P to A.P.V. and L.B.) and the NHM Biodiversity Initiative.

Author Contributions

D.K. analyzed genomic data and carried out, sequence alignments and phylogenetic analyses, D.K., M.M. and M.B. participated in other data analyses, D.K. and M.M. carried out the statistical analyses; all coauthors helped draft the manuscript; L.B. provided specimens, L.B. and A.P.V. conceived, designed and coordinated the study and drafted the manuscript. All authors gave final approval for publication.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35328-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018