A transcriptome-based phylogeny of Scarabaeoidea confirms the sister group relationship of dung beetles and phytophagous pleurostict scarabs (Coleoptera)

Article *in* Systematic Entomology · June 2023

DOI: 10.1111/syen.12602

11 authors, including:

Lars Dietz
Ruhr-Universität Bochum
32 PUBLICATIONS **279** CITATIONS
SEE PROFILE

Matthias Seidel
Naturhistorisches Museum Wien
70 PUBLICATIONS **296** CITATIONS
SEE PROFILE

Jonas Eberle
Research Museum Alexander Koenig
77 PUBLICATIONS **711** CITATIONS
SEE PROFILE

Lars Podsiadlowski
University of Bonn
127 PUBLICATIONS **6,930** CITATIONS
SEE PROFILE

Received: 11 March 2023 | Accepted: 16 May 2023

DOI: 10.1111/syen.12602

ORIGINAL ARTICLE

Systematic Entomology

Royal Entomological Society

A transcriptome-based phylogeny of Scarabaeoidea confirms the sister group relationship of dung beetles and phytophagous pleurostict scarabs (Coleoptera)

Lars Dietz[1] | Matthias Seidel[2,3] | Jonas Eberle[4] | Bernhard Misof[1,5] | Thaynara L. Pacheco[1] | Lars Podsiadlowski[1] | Sasanka Ranasinghe[1] | Nicole L. Gunter[6] | Oliver Niehuis[7] | Christoph Mayer[1] | Dirk Ahrens[1]

[1]Museum A. Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany

[2]CeNak Hamburg, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Hamburg, Germany

[3]Naturhistorisches Museum Wien, Austria

[4]Paris-Lodron-University, Salzburg, Austria

[5]Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

[6]The Cleveland Museum of Natural History, Cleveland, Ohio, USA

[7]Institute of Biology I (Zoology), University of Freiburg, Freiburg, Germany

Correspondence
Dirk Ahrens, Museum A. Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany.
Email: ahrens.dirk_col@gmx.de

Funding information
German Research Foundation, Grant/Award Numbers: 6-2, NI 1387/6-1, MI 649/18-1, AH175/6-2, AH 175/6-1; Museum Koenig Bonn; A. Koenig Stiftung Bonn

Abstract

Scarab beetles (Scarabaeidae) are a diverse and ecologically important group of angiosperm-associated insects. As conventionally understood, scarab beetles comprise two major lineages: dung beetles and the phytophagous Pleurosticti. However, previous phylogenetic analyses have not been able to convincingly answer the question whether or not the two lineages form a monophyletic group. Here, we report our results from phylogenetic analyses of more than 4000 genes mined from transcriptomes of more than 50 species of Scarabaeidae and other Scarabaeoidea. Our results provide convincing support for the monophyly of Scarabaeidae, confirming the debated sister group relationship of dung beetles and phytophagous pleurostict scarabs. Supermatrix-based maximum likelihood and multispecies coalescent phylogenetic analyses strongly imply the subfamily Melolonthinae as currently understood being paraphyletic. We consequently suggest various changes in the systematics of Melolonthinae: Sericinae Kirby, 1837 *stat. rest. and sensu n.* to include the tribes Ablaberini, Diphucephalini and Sericini, and Sericoidinae Erichson, 1847 *stat. rest.* and *sensu n.* to include the tribes Automoliini, Heteronychini, Liparetrini, Maechidiini, Phyllotocini, Scitalini, and Sericoidini. Both subfamilies appear to consistently form a monophyletic sister group to all remaining subfamilies so far included within pleurostict scarabs except Orphninae. Our results represent a major step towards understanding the diversification history of one of the largest angiosperm-associated radiations of beetles.

KEYWORDS
phylogeny, reclassification, Scarabaeidae

INTRODUCTION

The evolution of large parts of extant terrestrial biodiversity has been driven by the evolutionary success of angiosperm plants; these radiations have been linked to increased productivity and growth rates of angiosperm vegetation (de Boer et al., 2012), the rise of ectomycorrhiza enhancing chemical weathering of soils (Taylor et al., 2011, 2012), and the promotion of soil nutrient release by angiosperm litter that is easily decomposed (Berendse & Scheffer, 2009). While the diversification of many insects, and especially that of beetles, was directly or

Lars Dietz and Dirk Ahrens contributed equally to this study.

TABLE 1 Overview of results on monophyly of Scarabaeidae retrieved in different previous phylogenetic analyses.

Reference	Scarab dung beetles + Pleurosticti	Data/tree building method
Grebennikov and Scholtz (2004)	No	Larval morphology/Parsimony
Caterino et al. (2005)	No	18S/ML, Parsimony
Smith et al. (2006)	No	18S, 28S/Parsimony
Hunt et al. (2007)	No	28S, 16S, COI/BI
Lawrence et al. (2011)	No	Morphology/Parsimony
Ahrens et al. (2014)	No	18S, 28S, 16S, COI/BI
Bocak et al. (2014)	No	18S, 28S, 16S, COI/ML
Timmermans et al. (2016)	Yes	mt genomes/BI
McKenna et al. (2014)	Yes	CAD + 28S/BI
McKenna et al. (2015)	No	8 nuclear genes/BI
Gunter et al. (2016)	Yes	28S, 16S, COI/BI
Toussaint et al. (2017)	No	8 nuclear genes/BI
Song and Zhang (2018)	Yes	mt genomes/ML
Zhang et al. (2018)	Yes	95 PCG; AA/ML (RAxML)
	Yes	95 PCG; Nucleotides/ ML (RAxML)
	Yes	AA/ML (IQ-TREE)
	No	95 PCG; Nucleotides/ ML (IQ-TREE)
	No	95 PCG; AA/BI
	Yes	Nucleotides/BI
McKenna et al. (2019)	Yes	AA of Transcriptomes (4818 nuclear genes)/ ML
Ayivi et al. (2021)	Yes	mt genomes/BI, ML
Cai et al. (2022)	No	68 single-copy nuclear protein-coding genes/BI (PhyloBayes), ML (IQ-TREE)
Guo et al. (2022)	Yes	mt genomes/BI, ML

Abbreviations: 16S, 16S ribosomal DNA; 18S, 18S ribosomal DNA; 28S, 28S ribosomal DNA; AA, amino acid; BI, Bayesian inference; CAD, carbamoylphosphate synthetase domain of the rudimentary gene; COI, cytochrome oxidase *c* subunit I; ML, maximum likelihood; PCG, protein coding genes.

indirectly fostered by that of angiosperms (Ahrens et al., 2014; Hunt et al., 2007; McKenna et al. 2019), the evolutionary mechanisms and timescales of angiosperm-dependent radiations have remained poorly understood, as the phylogenetic relationships of many lineages remained insufficiently known. This is especially true for scarab beetles (Scarabaeidae), which represent a diverse lineage of beetles feeding predominantly on either angiosperm plants or mammal dung containing about 14 subfamilies and 27,000 species (Scholtz & Grebennikov, 2005). Although traditionally grouped into a single family (Scholtz & Grebennikov, 2005), it is divided into two major lineages: (1) the plant-feeding lineage Pleurosticti, which include, for example, rose chafers, rhinoceros beetles, and Christmas beetles, and (2) a clade of taxa mainly feeding on mammal dung (Aphodiinae + Scarabaeinae).

Molecular phylogenetic analyses of the Scarabaeidae have been controversial, with only 9 out of 21 recently published studies reporting Scarabaeidae being monophyletic (Table 1). Scarabaeidae are part of a wider clade (Scarabaeoidea) which also includes Lucanidae (stag beetles), Geotrupidae (earth-boring dung beetles) and several other families (Scholtz & Grebennikov, 2005), and the monophyly of this superfamily has been confirmed by all major molecular studies (e.g., McKenna et al. 2019; Zhang et al., 2018). The current classification of Scarabaeoidea (Scholtz & Grebennikov, 2005) is founded on morphological evidence (Lawrence & Newton, 1995). Yet, despite extensive research on the morphology of Scarabaeoidea by Browne and Scholtz (1998, 1999), the monophyly of Scarabaeidae has yet to be adequately tested. The analysis of Browne and Scholtz (1998) included only lineages of 'Scarabaeidae', which were rooted with a single outgroup, whereas the analysis of Browne and Scholtz (1999) coded 'Scarabaeidae' as a single terminal taxon. A recent cladistic analysis based on morphology in a wider systematic framework (Lawrence et al., 2011) did not recover Scarabaeidae as a monophyletic group.

A transcriptome-based phylogeny of Coleoptera based on 4818 genes by McKenna et al. (2019) supported the monophyly of Scarabaeidae with high support, although taxon sampling was limited to eight taxa. However, in phylogenies based on some subsets of data such as the first and second nucleotide position, monophyly of Scarabaeidae was not recovered, as scavenger scarab beetles (Hybosoridae) were sister to scarab dung beetles. Another analysis by Zhang et al. (2018), based on a lower number of genes (95) but a somewhat higher number of taxa (12) also found monophyly of Scarabaeidae but with poor support. A re-analysis of a 68-gene subset of that dataset using different methods

(Cai et al., 2022) did not find the family to be monophyletic, as Hybosoridae were sister to pleurosticts.

The uncertainty surrounding the monophyly of Scarabaeidae prompted discussions about the classification of the family (Kohlmann & Morón, 2003). Apart from many unresolved issues of classification within the group, the question arose whether to split scarab dung beetles and pleurostict scarabs (Erichson, 1848) in two or more families (Cherman & Morón, 2014; Morón, 1997). Prior to the era of molecular phylogenies, there has been an impressive array of classification schemes on Scarabaeidae and related families (Balthasar, 1963; Browne & Scholtz, 1995; Crowson, 1955; Endrödi, 1966; Howden, 1982; Iablokoff-Khnzorian, 1977; Janssens, 1949; Kohlmann & Morón, 2003; Lawrence & Newton, 1982, 1995; Medvedev, 1976; Morón, 1984, 1997, 2003; Nikolajev, 1995; Paulian & Baraud, 1982; Ratcliffe & Jameson, 2004; Scholtz, 1990; Scholtz & Grebennikov, 2005; Smith, 2006; Smith et al., 2006). As robust and convincing evidence is lacking for most of these classification schemes, many systematic and taxonomic studies arbitrarily followed one of these classifications according to the author's opinion or geographic provenience (e.g., Cherman & Morón, 2014).

Here, we readdress the question about the controversial sister group relationship of scarab dung beetles and pleurostict scarabs (Ahrens et al., 2014), which is fundamental to gain a more complete understanding of the evolutionary impact of angiosperms on the diversification of Scarabaeidae.

If the monophyly of Scarabaeidae is confirmed, then it raises the question of why angiosperms and their related follow-up radiations (e.g., that of large herbivore mammals; see Ahrens et al., 2014) seemingly had a more significant impact on the radiation of this lineage than on any other scarabaeoid beetle lineage. To this end, we expand the phylogenomic data set compiled by McKenna et al. (2019) within additional taxa of Scarabaeoidea. The expanded taxonomic sampling is used to assess which aspects of the data could result in the inference of incompatible topologies (e.g., Cai et al., 2022; Zhang et al., 2018).

MATERIALS AND METHODS

Taxon sampling and new transcriptome data

We analysed 57 transcriptomes that covered almost all major families of Scarabaeoidea, including 9 of the 14 subfamilies of Scarabaeidae according to Scholtz and Grebennikov (2005), and two outgroup taxa (Table S1). The outgroups were chosen as members of the closest relatives to Scarabaeoidea, and for their relatively short branches in the phylogeny of McKenna et al. (2019). Fifteen of these transcriptomes had been published by McKenna et al. (2019). Two transcriptomes were sequenced in context of the 1KITE project but had not been analysed and published before. For details relating to the extraction of total mRNA and fragmentation, construction of complementary deoxyribonucleic acid (cDNA) libraries, and tagging of these two datasets see Peters et al. (2017). Forty additional transcriptomes were specifically generated in context of this study (Table S1).

Extraction of RNA, cDNA library construction, library normalisation, and Illumina sequencing were carried out by a commercial sequencing company (Starseq, Mainz, Germany). In brief, tissues preserved in RNAlater (Qiagen, Hilden, Germany) were lysed and homogenised with a Precellys Evolution tissue lyser and a corresponding Lysis Kit, CKmix (Bertin Technologies SAS, Montigny-le-Bretonneux, France), in 2 mL volumes, containing a mix of 1.4 and 2.8 mm ceramic beads. RNA isolation was done with the Quick RNA miniprep kit (Zymo Research, Irvine, California). Library preparation was done with NEBNext Ultra II Directional RNA library preparation kit (NEB, Ipswich, Massachusetts). All mRNA libraries were sequenced with Illumina HiSeq 2000 sequencers (Illumina, San Diego, California), using paired-end 150-bp read length.

All raw nucleotide sequences are deposited at the National Center for Biotechnology Information (NCBI), Sequence Read Archive (see Table S1 for accession numbers).

The raw data of previously published transcriptomes of 17 Scarabaeoidea (see above) and of two outgroup taxa (*Ocypus brunnipes* (Fabricius) [Staphylinidae] and *Helophorus nanus* (Sturm) [Helophoridae]; McKenna et al. 2019) were downloaded from NCBI. All raw nucleotide reads from both newly sequenced and published transcriptomes were trimmed with TrimGalore 0.6.6 (Krueger et al., 2021) and assembled with Trinity 2.11.0 (Grabherr et al., 2011) using the software's default settings. Transcriptome assemblies (see Table S1 for accession numbers) are deposited at the Transcriptome Shotgun Assembly Database, NCBI Bioproject ID PRJNA906571 for newly sequenced transcriptomes or PRJNA936991 for re-assemblies of published transcriptomes (http://www.ncbi.nlm.nih.gov/bioproject). The assemblies were filtered with a custom Perl script (trinity_longest_d.pl, see File S1) to retain only the longest isoform (as identified by Trinity) per locus, as loci with multiple isoforms could otherwise be falsely discarded as paralogs in the gene orthology assessment step.

For quality assessment, we searched these filtered assemblies with BUSCO 4.0.6 (Manni et al., 2021) for Endopterygota single-copy orthologues from the endopterygota_odb10 dataset using transcriptome mode.

Data extraction and alignment

Tab-delimited files were downloaded from the OrthoDB10 database (v10.orthodb.org, last accessed 22 April 2023; Kriventseva et al., 2019) for all groups of orthologous genes (= orthologue groups) at the hierarchical level Coleoptera that were present in at least eight of the nine coleopteran genomes in the database and single copy in all of them. This included a total of 4296 genes. This is similar to the principle of universal single-copy orthologues (USCOs) as used by the programme BUSCO (Simão et al., 2015). However, USCOs have to be present and single copy in at least 90% of all known genomes of a taxonomic group. In this case, this would mean that the genes would have to be present in all nine annotated coleopteran genomes available at the time. To avoid excluding genes that may be absent simply due to the incompleteness

of one of the nine genome assemblies, we decided to include genes present in only eight of the nine coleopteran genomes.

Furthermore, we downloaded the official gene sets (OGS) for all nine available coleopteran genomes from OrthoDB10. Tab-delimited files were modified for use in Orthograph 0.7.1 (Petersen et al., 2017) and used together with the OGS, to create a SQLite database of the genetic information with that programme. Hidden Markov models (HMMs) were created with Orthograph from the available amino acid (AA) sequences of each orthologue group. HMMs were used then to extract the target genes from the filtered Trinity contigs of each specimen with Orthograph, using the software's default settings.

USCO nucleotide and corresponding AA sequences that were identified and inferred with Orthograph were aligned in two different ways. We first aligned the inferred AA sequences against the HMMs from Orthograph with hmmalign (part of the HMMER 3.3 package; Eddy, 2011; http://eddylab.org/software/hmmer/hmmer.org). The AA alignment was then used as a blueprint to align the corresponding nucleotide sequences with pal2nal 14.1 (Suyama et al., 2006). Alignment regions not covered by the HMMs were removed with a custom Perl script (hmmalign_cut2_d.pl; File S2). We additionally aligned the AA sequences with MAFFT 7.305b (Katoh & Standley, 2013) using the L-INS-i algorithm. The corresponding nucleotide sequence alignments were inferred using again the software pal2nal. Poorly aligned regions in the AA sequence alignments were identified with ALISCORE 2.0 (Kück et al., 2010; Misof & Misof, 2009) and removed from the AA sequence alignments and corresponding nucleotide sequence alignments with the software ALICUT 2.31 (available from: https://github.com/PatrickKueck/AliCUT). Outlier sequences were identified and removed with the software OliInSeq 0.9.3 (https://github.com/cmayer/OliInSeq) using the software's default parameters. As the third codon position is typically hyper-variable and often exhibits inhomogeneous nucleotide frequencies, we also used a custom Perl script (extract_codpos_d.pl; File S3) to generate nucleotide sequence alignments in which the third codon position was removed.

To test the monophyly of Scarabaeidae with data from another dataset, we downloaded the OrthoDB10 Endopterygota data from BUSCO 4.0.6 (Manni et al., 2021; Simão et al., 2015), including HMMs and information files, from the BUSCO website (busco.ezlab.org). This set comprises 2124 genes that are present in single copy in at least 90% of all known genomes of Endopterygota (hereafter referred to as Endopterygota USCOs). Searching for the orthologues of *Onthophagus taurus* (Schreber) in this and the Coleoptera-specific set showed that 1496 genes are shared by both datasets. The OGS for all 56 species in that dataset were downloaded from OrthoDB and used to create an SQLite database with Orthograph. Together with the HMMs from BUSCO, the information was used to extract the Endopterygota USCO genes of each specimen with Orthograph as described above.

Phylogenetic tree inference

We inferred phylogenetic trees from the Coleoptera and Endopterygota data sets, analysing the AA sequence alignments generated with hmmalign and MAFFT and the corresponding nucleotide sequence alignments (NT) inferred with pal2nal. We then considered two sets of nucleotide sequence alignments: those that included all three codon positions (NT123) and those that include only first and second codon positions (NT12). The multiple sequence alignments were analysed using coalescent-based and concatenation-based tree inference methods. For conducting the concatenation-based phylogenetic analyses, the multiple sequence alignments of a given type (i.e., AA, nucleotide) of all genes were concatenated with a custom Perl script (concat_eogs_part_d.pl; File S4). The resulting super-alignments were then analysed under maximum likelihood (ML) with IQ-TREE 2.1.2 (Minh et al. 2020). The datasets were partitioned by gene. The best-fitting model and partitioning scheme were inferred with ModelFinder, using the IQ-TREE option -m MFP + MERGE (Chernomor et al. 2016; Kalyaanamoorthy et al., 2017). Branch support was assessed with approximate likelihood ratio tests and via ultrafast bootstrapping (Hoang et al., 2018) applying 1000 replicates and nearest neighbour interchange as tree rearrangement method. For conducting coalescent-based phylogenetic analyses, we first calculated phylogenetic trees of each gene with IQ-TREE, determining the best-fitting model with ModelFinder. The resulting gene trees were used for analyses with ASTRAL 5.6.1 (Zhang et al., 2018) with default parameters that we used to conduct the coalescent-based phylogenetic analysis. All trees were rooted with *Helophorus* Fabricius (Helophoridae) and *Ocypus* Leach (Staphylinidae) as outgroups.

To test the effect of missing data, we generated reduced versions of all nucleotide and amino acid datasets in which positions with a taxon coverage of <70% were removed with a custom Perl script (removegaps_d.pl; File S5). The previously described phylogenetic analyses were repeated using these reduced datasets.

Furthermore, we examined the effect of varying substitution rates between different genes. For this test, we divided datasets containing at least 70% complete positions into sets of fast and of slowly evolving genes. Using a custom script (pairwise_id2.pl; File S6), we calculated the pairwise sequence identity within each gene alignment according to Sharma et al. (2014). We then ordered both the hmmalign- and MAFFT-inferred multiple sequence alignments of individual genes by pairwise identity and divided them into two sets with high (slow evolving) and with low (fast evolving) pairwise sequence identity, each including 50% of the genes. We then conducted concatenation-based and coalescent-based analyses on these sets using the same methods as described above.

Alignments of individual and concatenated loci including resulting trees (Supplementary Files S7–S12) are deposited in Dryad (doi: 10.5061/dryad.d51c5b07h).

Topology tests

To assess support for the monophyly of Scarabaeidae in our datasets, we conducted a number of topology tests on the six different datasets (i.e., NT12, NT123, and AA, inferred using hmmalign or MAFFT).

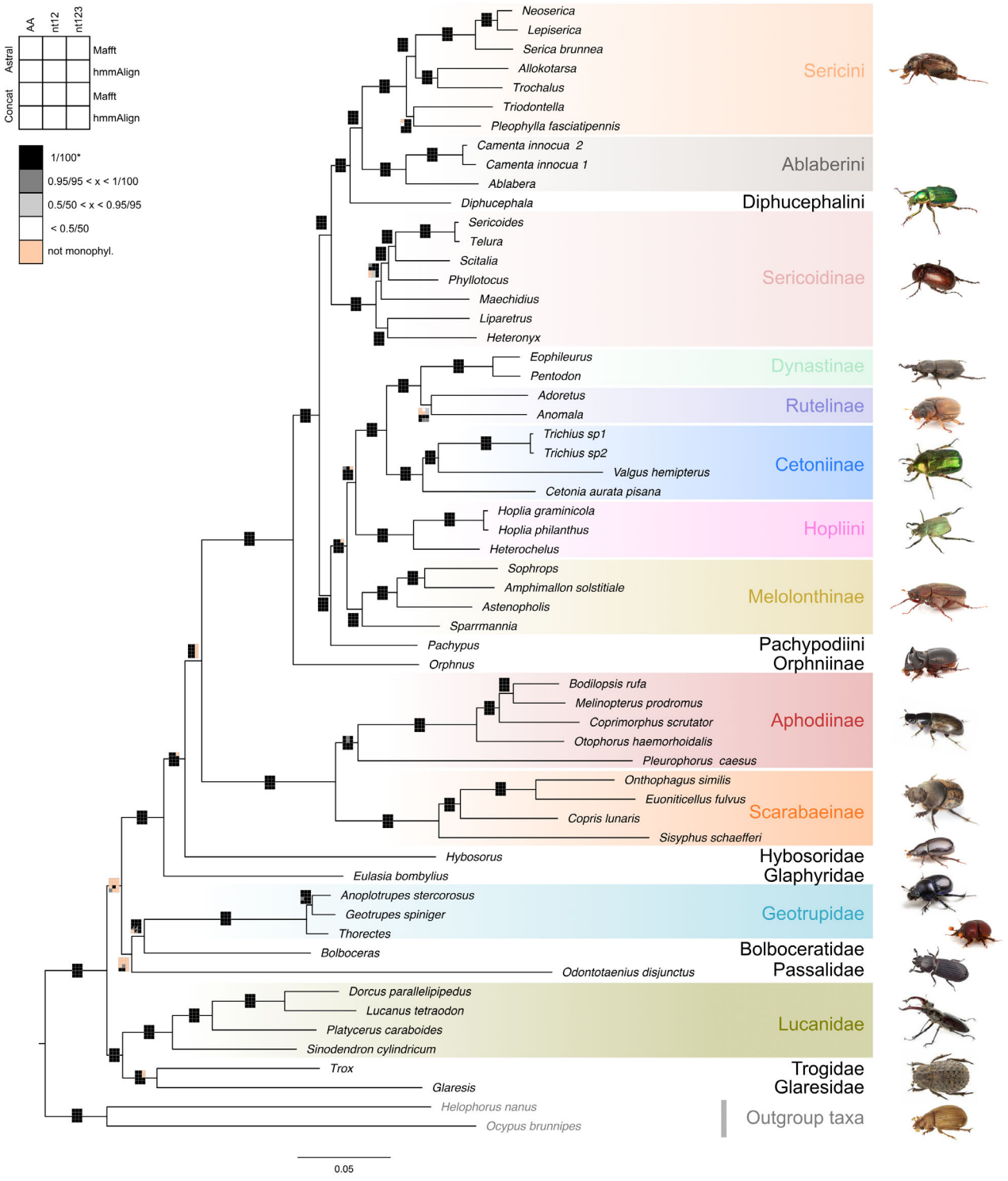

FIGURE 1 Phylogenetic tree from the concatenated MAFFT alignment of the nucleotide data (Coleoptera single copy orthologues, full dataset) containing only the first and second base pair (nt12). Branch support (approximate likelihood ratio tests/ultrafast bootstrap for maximum likelihood trees, local posterior probability for ASTRAL trees) and recovery of clades with other alignments and tree reconstruction approaches as well as the current systematic assignment of lineages are mapped onto the branches.

First, we computed the likelihood scores of ML trees constrained to support different topologies for each dataset with IQ-TREE. The constrained trees covered all possible phylogenetic relationships of the following four taxa: Glaphyridae, Hybosoridae, Scarabaeinae + Aphodiinae, Pleurosticti. We compared the likelihood of the constraint trees using a variety of resampling tests in IQ-TREE using RELL approximation

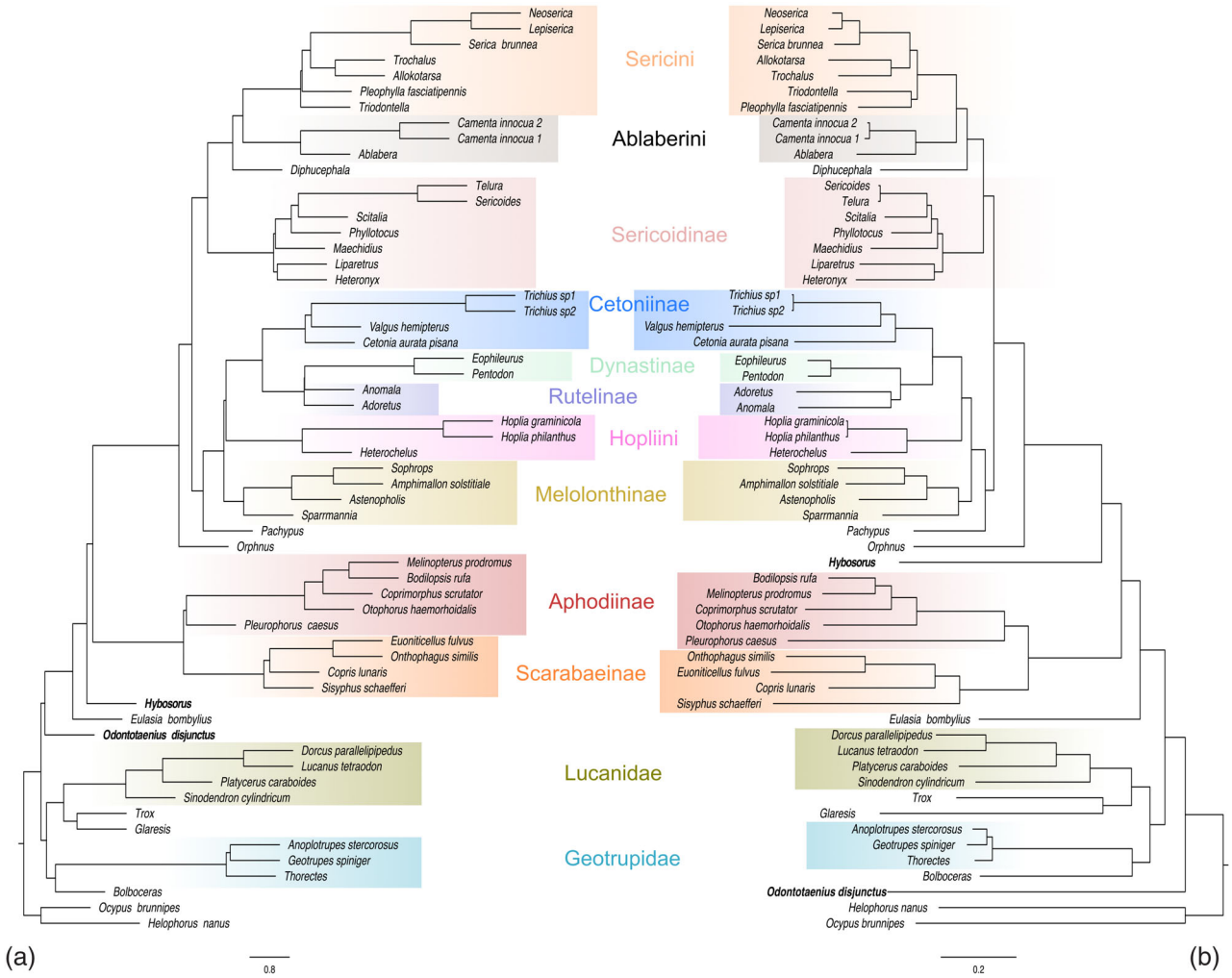

FIGURE 2 Contrasting tree topologies based on the Coleoptera single copy orthologues obtained with coalescent-based tree search with ASTRAL (amino acid sequences; left side) and with maximum-likelihood analysis of concatenated data using all nucleotides (right side); major lineages are highlighted; single taxa changing phylogenetic position are marked in bold.

(Kishino et al., 1990) with 10,000 iterations, including bootstrap proportion (BP), Kishino–Hasegawa (KH) test (Kishino & Hasegawa, 1989), Shimodaira–Hasegawa (SH) test (Shimodaira & Hasegawa, 1999), expected likelihood weights (ELW; Strimmer & Rambaut, 2002), and the approximately unbiased (AU) test (Shimodaira, 2002).

We used the same strategy to assess support for all possible phylogenetic relationships between Dynastinae and the genera *Anomala* Samouelle and *Adoretus* Laporte (both part of the potentially paraphyletic Rutelinae) as well as between Melolonthinae s. str., Hopliini, and Cetoniinae + Rutelinae + Dynastinae (hereafter referred to as CRD).

We further tested the monophyly of Scarabaeidae using four-cluster likelihood mapping (Strimmer & von Haeseler, 1997) in IQ-TREE. For this purpose, we divided the taxon set into Hybosoridae, Scarabaeinae + Aphodiinae, Pleurosticti, and all others. All 4410 unique quartets containing one taxon of each quartet were tested for their support for the three possible four-taxon trees.

RESULTS

Data completeness

The percentage of single-copy genes with full sequences found by BUSCO ranged between 60.8% and 87.1% (mean 77.5%), except for the *Camenta innocua* (Boheman) transcriptome from McKenna et al. (2019) for which it was only 38.9% (Table S1).

Nucleotide sequences from all but one of 4296 Coleoptera-specific genes and from all 2120 Endopterygota USCO genes were successfully recovered. In total, the dataset of the Coleoptera-specific genes aligned with hmmalign comprised 6,960,192 nucleotide positions, of which 3,670,744 were parsimony informative. The overall alignment completeness was 62.7%. The corresponding MAFFT-aligned dataset comprised 3,929,520 nucleotides, of which 2,186,562 were parsimony informative. The alignment completeness was 79.6%. The dataset of Endopterygota USCO genes comprised 2,337,876

nucleotides aligned with hmmalign, of which 1,126,272 were parsimony informative. The alignment completeness was 62.8%. The corresponding dataset aligned with MAFFT was 1,649,496 nucleotide positions long, of which 893,640 were parsimony informative. The alignment completeness was 82.2%.

Phylogenetic analyses

The choice of the alignment software (hmmalign vs. MAFFT) had little impact on the results of the phylogenetic analyses (Figures 1, 2, and S1–S12), and neither did the presence of incomplete alignment sites. However, the choice of the alignment software had an impact on the inferred position of Passalidae (e.g., Figures S7–S12). Both sets of genes yielded very similar phylogenies, except that in trees based on the smaller Endopterygota USCO dataset (Figures S37–S60), the single species of passalid included in our analyses was consistently placed as sister to Glaphyridae + Hybosoridae + Scarabaeidae, while in the larger Coleoptera-specific set (Figures S1–S36), its position differed in the various phylogenetic analyses. Including only fast or slowly evolving genes (Figures S25–S36) did not lead to noteworthy consistent differences in topology, although support values were generally lower than with the complete gene set. The most significant topological differences were found between trees inferred from datasets that included all three codon positions (NT123; e.g., Figures S9 and S12) and between trees inferred from datasets which included only nucleotides of the first and second codon position (e.g., Figures S8 and S11) and those that were analysed on the AA level (e.g., Figures S7 and S10). Whether a concatenation-based (e.g., Figures S7–S12) or a coalescent-based approach (e.g., Figures S1–S6) was used to analyse the data had little impact on the unrooted topology within Scarabaeoidea, although the rooting of the tree (i.e., the placement of the outgroup) was affected. All datasets yield well-resolved trees, in which most splits received maximal support.

Scarabaeidae are strongly supported as being monophyletic by all datasets except the nucleotide datasets that include the third codon position (NT123; Figure 1). Analysis of the latter suggested Hybosoridae as the sister group of pleurostict Scarabaeidae.

TABLE 2 Likelihood tests regarding the monophyly of Scarabaeidae using the coleopteran-specific dataset, with constraint trees using a variety of resampling tests in IQ-TREE using the RELL approximation including bootstrap proportion (BP), Kishino–Hasegawa test (KH), Shimodaira–Hasegawa test (SH), expected likelihood weights (ELW), and the approximately unbiased (AU) test.

Tree	logL	deltaL	BP-RELL	p-KH	p-SH	c-ELW	p-AU
hmmalign NT12							
1: Dung scarabs + Pleurosticti	−48,463,502	0	1	1	1	1	1
2: Hybosoridae + Pleurosticti	−48,465,916	2414	0	0	0	0	2.45E-05
3: Hybosoridae + Dung scarabs	−48,465,122	1620.3	0	0	0	0	1.68E-59
hmmalign NT123							
1: Dung scarabs + Pleurosticti	−124,696,860	1522.3	0	0	0	0	1.08E-05
2: Hybosoridae + Pleurosticti	−124,695,338	0	1	1	1	1	1
3: Hybosoridae + Dung scarabs	−124,697,335	1997.4	0	0	0	0	1E-07
hmmalign AA							
1: Dung scarabs + Pleurosticti	−44,349,511	0	1	1	1	1	1
2: Hybosoridae + Pleurosticti	−44,353,242	3730.9	0	0	0	0	1.48E-08
3: Hybosoridae + Dung scarabs	−44,351,488	1976.8	0	0	0	0	7.45E-09
MAFFT NT12							
1: Dung scarabs + Pleurosticti	−31,506,595	0	1	1	1	1	1
2: Hybosoridae + Pleurosticti	−31,508,450	1855	0	0	0	0	0.000172
3: Hybosoridae + Dung scarabs	−31,507,967	1372	0	0	0	0	3.29E-69
MAFFT NT123							
1: Dung scarabs + Pleurosticti	−87,783,308	2052.7	0	0	0	0	4.4E-82
2: Hybosoridae + Pleurosticti	−87,781,255	0	1	1	1	1	1
3: Hybosoridae + Dung scarabs	−87,783,596	2341.4	0	0	0	0	1.48E-48
MAFFT AA							
1: Dung scarabs + Pleurosticti	−26,130,488	0	1	1	1	1	1
2: Hybosoridae + Pleurosticti	−26,133,387	2898.7	0	0	0	0	6.66E-06
3: Hybosoridae + Dung scarabs	−26,132,268	1780	0	0	0	0	4.07E-06

Note: The confirmed most likely topology is highlighted in bold.
Abbreviations: AA, amino acid; NT, nucleotide.

TABLE 3 Likelihood tests regarding the monophyly of Rutelinae using the coleopteran-specific dataset, with constraint trees using a variety of resampling tests in IQ-TREE using the RELL approximation including bootstrap proportion (BP), Kishino–Hasegawa test (KH), Shimodaira–Hasegawa test (SH), expected likelihood weights (ELW), and the approximately unbiased (AU) test.

Tree	logL	deltaL	Bp-RELL	p-KH	p-SH	c-ELW	p-AU
hmmalign NT12							
1: *Adoretus* + *Anomala*	−48,463,550	0	0.986	0.983	1	0.986	0.991
2: *Adoretus* + Dynastinae	−48,463,962	412.1	0.0006	0.0012	0.0012	0.000607	0.0013
3: *Anomala* + Dynastinae	−48,463,842	292.12	0.0134	0.0168	0.0277	0.0135	0.0129
hmmalign NT123							
1: *Adoretus* + *Anomala*	−124,695,262	0	0.97	0.976	1	0.97	0.991
2: *Adoretus* + Dynastinae	−124,695,650	387.47	0.0053	0.0063	0.0129	0.00534	0.00767
3: *Anomala* + Dynastinae	−124,695,590	327.26	0.0244	0.0236	0.0407	0.0245	0.0198
hmmalign AA							
1: *Adoretus* + *Anomala*	−44,349,512	0	1	1	1	1	1
2: *Adoretus* + Dynastinae	−44,349,994	482.06	0.0002	0.0001	0.0002	0.000195	0.000127
3: *Anomala* + Dynastinae	−44,350,021	508.52	0.0002	0	0.0002	0.00019	0.000164
MAFFT NT12							
1: *Adoretus* + *Anomala*	−31,506,594	0	1	1	1	1	1
2: *Adoretus* + Dynastinae	−31,507,065	470.61	0.0001	0.0002	0.0003	0.0001	7.03E-05
3: *Anomala* + Dynastinae	−31,507,098	503.3	0	0	0	5.19E-05	0.000247
MAFFT NT123							
1: *Adoretus* + *Anomala*	−87,781,254	0	0.998	0.997	1	0.998	0.998
2: *Adoretus* + Dynastinae	−87,781,673	419.51	0.0021	0.0031	0.0048	0.00214	0.0029
3: *Anomala* + Dynastinae	−87,781,779	524.79	0.0001	0.0002	0.0005	0.00013	0.000376
MAFFT AA							
1: *Adoretus* + *Anomala*	−26,130,489	0	1	1	1	1	1
2: *Adoretus* + Dynastinae	−26,130,951	462.27	0.0003	0.0001	0.0005	0.000317	0.000697
3: *Anomala* + Dynastinae	−26,130,995	506.31	0.0002	0.0001	0.0002	0.000164	0.000818

Note: The confirmed most likely topology is highlighted in bold.
Abbreviations: AA, amino acid; NT, nucleotide.

All remaining scarabaeoid families that were represented by more than one species in our datasets were consistently supported as being monophyletic.

Phylogenetic analysis of the NT12 and AA supermatrices suggested Lucanidae is the sister group of Glaresidae + Trogidae. They furthermore suggested Geotrupidae + Bolboceratidae is the sister group of a clade that comprised Glaphyridae, Hybosoridae, and Scarabaeidae. Passalidae was placed in various positions within that group depending on the specific analysis. Hybosoridae was found as the sister group of Scarabaeidae. The monophyly of Scarabaeidae was always maximally supported by our data (Figures 1, S1, and S2). Coalescence-based summary trees of all gene trees differed from the supermatrix-based trees in suggesting Bolboceratidae +Geotrupidae being the sister group of Glaresidae, Lucanidae, and Trogidae. However, some the coalescence-based summary trees inferred from exclusively slowly evolving genes showed the same topology as the concatenation-based trees.

The two subgroups of Scarabaeidae, Aphodiinae + Scarabaeinae, and Pleurosticti were always strongly supported as being monophyletic, as were the two subfamilies, Aphodiinae and Scarabaeinae. Within Pleurosticti, *Orphnus* MacLeay (Orphninae) was consistently found as sister group of the remaining pleurostict scarabaeid lineages, that were divided into two major clades: the first clade comprised (1) Sericoidinae—Australasian and Neotropical taxa referred to by some authors as Southern World Melolonthinae (Ahrens et al., 2011, 2014; Ahrens & Vogler, 2008; Šípek et al., 2016) or Liparetrinae (Eberle et al., 2019; Lacroix, 2007, 2014; Pacheco et al., 2022)—and (2) Ablaberini + Sericini + the Australian genus *Diphucephala* (i.e., Diphucephalini), which represented the sister group of the two former tribes. The second clade comprised *Pachypus* Dejean (Pachypodini) and three major lineages, in which *Pachypus* was found to be sister group of the latter. The phylogenetic relationships among the latter three lineages differed among the inferred trees. These three lineages were: (1) Hopliini, (2) Melolonthini, including the genus *Sparrmannia* Laporte (currently placed within the probably polyphyletic Tanyproctini; Eberle et al., 2019), and (3) Cetoniinae + (Dynastinae + Rutelinae). Supermatrix-based phylogenetic analyses consistently provided support for Melolonthini being the sister group of the remaining two lineages. Coalescence-based analyses were inconsistent in regard of the phylogenetic arrangement of

TABLE 4 Likelihood tests regarding the monophyly sister group relationship of Hopliini and Melolonthinae s. str. (including Melolonthini and *Sparrmannia*, excluding *Pachypus* and Sericini) using the coleopteran-specific dataset, with constraint trees using a variety of resampling tests in IQ-TREE using the RELL approximation including bootstrap proportion (BP), Kishino–Hasegawa test (KH), Shimodaira–Hasegawa test (SH), expected likelihood weights (ELW), and the approximately unbiased (AU) test.

Tree	logL	deltaL	Bp-RELL	p-KH	p-SH	c-ELW	p-AU
hmmalign NT12							
1: **Hopliini + Dynastinae, etc.**	−48,463,522	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−48,466,907	3384.2	0	0	0	0	9.98E-06
3: Melolonthinae + Hopliini	−48,466,288	2765.9	0	0	0	0	9.98E-06
hmmalign NT123							
1: **Hopliini + Dynastinae, etc.**	−124,695,344	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−124,703,111	7766.7	0	0	0	0	1.01E-08
3: Melolonthinae + Hopliini	−124,701,830	6486	0	0	0	0	2.33E-43
hmmalign AA							
1: **Hopliini + Dynastinae, etc.**	−44,349,512	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−44,352,720	3208.2	0	0	0	0	1.65E-52
3: Melolonthinae + Hopliini	−44,352,216	2704.6	0	0	0	0	1.22E-57
MAFFT NT12							
1: **Hopliini + Dynastinae, etc.**	−31,506,595	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−31,509,134	2539.6	0	0	0	0	2.63E-44
3: Melolonthinae + Hopliini	−31,508,736	2141.3	0	0	0	0	1.64E-67
MAFFT NT123							
1: **Hopliini + Dynastinae, etc.**	−87,781,254	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−87,787,291	6037.7	0	0	0	0	4.16E-50
3: Melolonthinae + Hopliini	−87,786,383	5129.7	0	0	0	0	4.65E-07
MAFFT AA							
1: **Hopliini + Dynastinae, etc.**	−26,130,487	0	1	1	1	1	1
2: Melolonthinae + Dynastinae, etc.	−26,132,818	2331	0	0	0	0	1.09E-45
3: Melolonthinae + Hopliini	−26,132,612	2124.8	0	0	0	0	6.7E-09

Note: The confirmed most likely topology is highlighted in bold.
Abbreviations: AA, amino acid; NT, nucleotide.

the three lineages. While the monophyly of Dynastinae + Rutelinae was consistently well supported, a monophyly of Rutelinae was strongly supported only in concatenation-based trees. In coalescence-based trees, the grouping was not consistently found, and if so, it typically received only low support—a pattern also confirmed by results of previous studies in which monophyly of Rutelinae did not result (e.g., Ahrens et al., 2014; Ahrens & Vogler, 2008; Neita-Moreno et al., 2019; Šípek et al., 2016).

With Cetoniinae + (Dynastinae + Rutelinae) consistently found nested within the lineages so far classified as Melolonthinae, our study confirmed the paraphyly of Melolonthinae (Ahrens et al., 2014; Ahrens & Vogler, 2008; Gunter et al., 2016; McKenna et al. 2019; Neita-Moreno et al., 2019; Šípek et al., 2016).

Topology tests

In tests assessing the support for monophyly of Scarabaeidae, the originally inferred topology (i.e., monophyly of Scarabaeidae) consistently received the highest support when analysing the NT12 and AA supermatrices. Likewise, the monophyly of Hybosoridae + Pleurosticti consistently received the highest support when analysing the NT123 datasets. Alternative trees were rejected with $p < 0.001$ (Table 2). The topology tests supported the sister group relationship of Hopliini and the CRD clade, rejecting alternative topologies with $p < 0.001$. Likelihood tests testing the monophyly of Rutelinae supported the latter irrespective of what dataset we analysed, but often only with p between 0.01 and 0.05 (Tables 3 and 4).

Four-cluster likelihood mapping revealed support for a monophyly of Scarabaeidae when analysing the NT12 and AA datasets (Figure 3). When analysing the NT12 datasets, a monophyly of Scarabaeidae was supported by 55%–60% of the quartets. When analysing the AA datasets, the support was >80%. However, it should be noted that in all analyses almost all quartets that include the most remotely related outgroups *Helophorus* and *Ocypus* supported Hybosoridae + Pleurosticti, while among those containing the closest outgroup *Eulasia* Truqui (Glaphyridae), more than 90% supported monophyly of

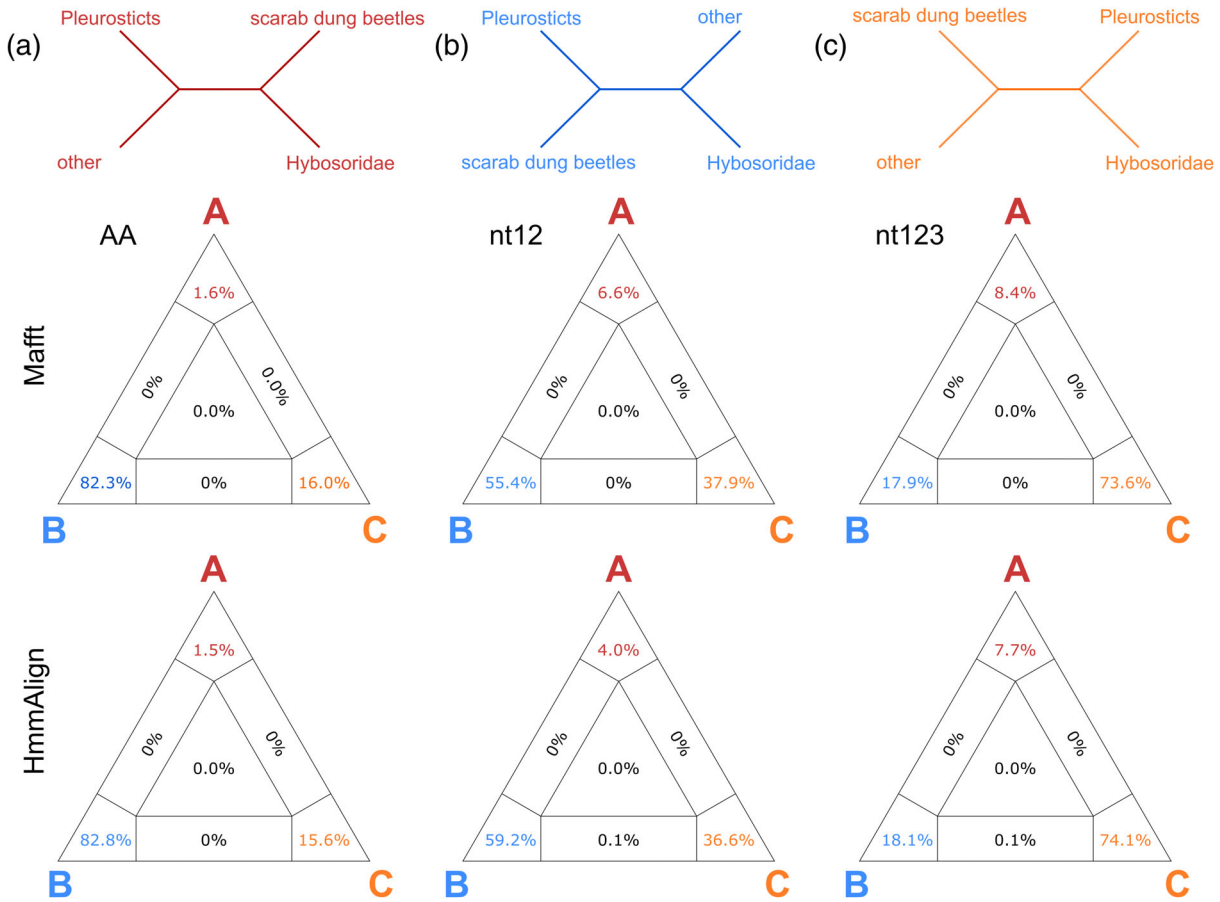

FIGURE 3 Four-cluster likelihood mapping based on alignments of the amino acid (AA), nt12, and nt123 datasets of Coleoptera single-copy orthologues with hmmalign and MAFFT.

Scarabaeidae, even when analysing the NT12 datasets. When analysing the NT123 datasets, more than 70% of quartets supported Hybosoridae + Pleurosticti.

DISCUSSION

The interfamilial results found in our study agree well with those of other multi-gene phylogenetic studies with less taxon and/or gene sampling (i.e., Cai et al., 2022; McKenna et al. 2019; Zhang et al., 2018). One exception is the position of Passalidae, which other studies found to be phylogenetically closely related to Geotrupidae + Bolboceratidae (Beza-Beza et al., 2020; McKenna et al. 2019). We found this phylogenetic position only in a few of our trees. In the majority of the inferred trees, we found Passalidae to be closely related to Scarabaeidae. As Passalidae had been represented by only few taxa on relatively long branches in all phylogenetic studies (including ours) conducted so far, a definitive statement on the phylogenetic position of Passalidae must await broader taxonomic sampling of the family itself and of related clades.

Our results confirm the early divergence of Orphninae within Pleurosticti, which had remained unresolved in some earlier molecular studies (e.g., Ahrens et al., 2014). However, an early divergence had been suspected based on morphological evidence (e.g., Ahrens, 2006; Browne & Scholtz, 1998). For example, Orphninae lack the derived conformation of spiracles after which Pleurosticti is named (Erichson, 1848).

The results of our study confirmed the sister group relationship of dung feeding scarabs and phytophagous pleurostict scarabs (i.e., monophyly of Scarabaeidae), which was not always found in previous studies with limited gene sampling (see Table 1). We found monophyly of Scarabaeidae with both supermatrix and coalescent-based tree reconstruction approaches. However, tree reconstruction results based on the nucleotide sequence data heavily depended on whether or not the third codon position was included—a phenomenon frequently observed in phylogenomic studies (Li et al., 2014).

The results of the likelihood mapping analyses strongly suggested that the non-monophyly of Scarabaeidae in the analyses of the NT123 data is an artefact. As this result is suggested primarily by quartets that include distantly related outgroup taxa, it can possibly be explained by long-branch attraction between the outgroup and the relatively long-branched dung-feeding scarabs. Likewise, long-branch attraction between the CRD clade and outgroups may have led to an artificial grouping of Melolonthinae and Hopliini.

The inferred monophyly of Scarabaeidae supports the initial hypothesis that the radiation of angiosperms had primarily affected

species diversity, diversity of feeding habits, and morphological disparity of only a single lineage of Scarabaeoidea. This sheds new light on possible causes for their successful diversification, which might be highly lineage related, possibly also in regard to genome-driven events (e.g., McKenna et al. 2019). The successful diversification of other major herbivorous beetle lineages, such as the Phytophaga (Chrysomeloidea + Curculionoidea), was attributed to the genomic presence of plant cell wall-degrading enzymes (PCWDEs) obtained from bacteria and fungi (McKenna et al. 2019). However, the presence of PCWDEs in Scarabaeoidea was only limited to glycosine hydrolase 1 and 9 that were expected to occur in most beetle species (McKenna et al. 2019). Cellulase, hemicellulase, pectinase, xylanase, and other polysaccharide-degrading enzymes have been documented in the hind/midgut of several scarab lineages, some of which were attributed to endosymbiotic bacteria (Bauchop & Clarke, 1975; Huang et al., 2010; Wada et al., 2014).The presence of these gut endosymbionts has likely promoted the diversification in association with angiosperm plants and could have been a trigger for the successful radiation of Scarabaeidae.

As one key innovation in this regard can be seen the development of female accessory glands at the end of the digestive and genital duct, which are present in Aphodiinae and pleurostict scarabs (Ahrens, 2006). Accessory glands are known to have an important function in the transmission of endosymbiont bacteria (see above) that are known, beyond their active part in cellulose digestion (Martin, 1983; Martin et al., 1991), to assist the production of pheromones (Hoyt et al., 1971). The latter could have had an important impact on the improvement of chemical communication in these groups (Pacheco et al., 2022). It should be noted that accessory glands are absent in Scarabaeinae, but other mechanisms for transmission of endosymbiotic bacteria have been documented (Estes et al., 2013). Nevertheless, the confirmed monophyly of Scarabaeidae allows us to interpret the reduction of accessory glands in Scarabaeinae dung beetles as a true loss and not as a parallelism between pleurostict scarabs and Aphodiinae.

The diverse Mesozoic fossil record of Hybosoridae (Lu et al., 2022) still provides important information for dating the onset of scarab divergences, given the poor fossil record of the Mesozoic Scarabaeidae (Krell, 2007), particularly in amber which often allows a more accurate classification and a more robust systematic placement of its fossils. While we confirm the sister-group relationship between Scarabaeidae and Hybosoridae, the limited taxonomic sampling of this study restricts our ability to accurately place most of the known fossils into the inferred phylogenetic tree. Therefore, we explicitly refrained from inferring divergence time estimates.

Implications on the classification of Scarabaeidae

Our phylogenetic analyses on Scarabaeidae revealed that there is—from a phylogenetic point of view—no necessity to split Scarabaeidae into two families. Given the monophyly of dung beetles and phytophagous pleurostict scarabs, such a splitting (Cherman & Morón, 2014) would be rather arbitrary and in contrary to the aim of maintaining a stable nomenclature and classification, which are important backbones of all biodiversity-related databases (e.g., NCBI, GBIF). It should be noted, though, that our study did not include some taxonomic lineages (e.g., Ochodaeidae, Eremazinae) that earlier studies found (although with poor support) within a clade which included Glaphyridae, Hybosoridae, scarab dung beetles, and pleurostict Scarabaeidae (Ahrens et al., 2014; Neita-Moreno et al., 2019). Interpretations for classification and evolution of the Scarabaeidae will be thus more robust when the position of these groups is better known.

The phylogenetic tree hypothesis supported by our study (Figure 1) shows some examples of non-homogenous lineage classifications, in which sister taxa, according to the current classification, are either classified as tribes or as subfamilies (e.g., Pachypodini vs. Melolonthinae vs. Rutelinae/Dynastinae/Cetoniinae vs. Hopliini). Given the non-monophyly of Melolonthinae as currently understood (Bouchard et al., 2011; Smith, 2006), the question arises to which clade the name 'Melolonthinae' should be referred, with respective modifications to the current classification. Many unanswered questions remain due to the limited sampling here and contradictory tree topologies compared with and between previous studies (e.g., Ahrens et al., 2014; Eberle et al., 2019).

The clear phylogenetic separation of the monophyletic 'Southern World' Melolonthinae and the clade containing Sericini, Ablaberini, and Diphucephalini from the remaining pleurostict scarabs (i.e., Cetoniinae, Dynastinae, Melolonthinae, Rutelinae, etc.) makes it reasonable to treat both lineages as separate subfamilies. Both lineages are currently classified as a series of tribes within Melolonthinae (Smith, 2006). A revised classification that elevates these clades to subfamilies would alleviate the problem, at least in part, of rendering Melolonthinae polyphyletic. It would also help to focus on the problem of whether Melolonthinae are monophyletic under inclusion of Hopliini and Macrodactylini (the latter not included in taxonomic sampling of this study). In regard to the sister group relationship Hopliini + Melolonthini, based on the current sampling there is good support that they do not form a monophyletic group (Table 4) although due to the limited sampling our results should not be regarded as fully conclusive.

Currently, the clade Melolonthini, which we refer here to a restricted interpretation of the subfamily 'Melolonthinae', includes several lineages currently circumscribed as subtribes, such as Enariina, Leucopholina, Melolonthina, Pyglina, Rhizotrogina, and Schizonychina. It also includes several other minor lineages (Eberle et al., 2019) and could at least contain the genus *Sparrmannia*. The latter is so far assigned to Tanyproctini but the position of other genera of the polyphyletic Tanyproctini remains yet uncertain (Eberle et al., 2019). A restricted Melolonthinae (see Figure 1) would be a starting point for a re-classification that would allow retaining well-established subfamily names (e.g., Cetoniinae, Dynastinae, Rutelinae). However, the exact extent of this clade Melolonthinae is yet to be identified, particularly with reference to the other lineages so far classified as 'Melolonthinae' (e.g., Diplotaxini, Hopliini, Macrodactylini; e.g., Ahrens et al., 2014). To further address this topic, the taxonomic sampling needs to be extended, also to allow more robust statistical topology testing (Tables 2–4; Figure 3). The same applies to

Dynastinae + Rutelinae: the monophyly of Rutelinae with respect to Dynastinae recovered by our results was often not supported in other studies (e.g., Ahrens et al., 2014; Ahrens & Vogler, 2008; Neita-Moreno et al., 2019; Šípek et al., 2016), although Guo et al. (2022) also found a clade of Adoretini + Anomalini to the exclusion of Dynastinae using mitochondrial genomes, similar to our results. However, as our datasets contained only one representative each of two of the seven currently recognised tribes of Rutelinae, a decision on the classification of these two subfamilies should await further studies.

The following formal classification changes are proposed. Sericinae Kirby, 1837 *stat. rest.* and *sensu n.* is re-elevated to subfamily and revised to include the tribes Ablaberini, Diphucephalini and Sericini. Sericoidinae Erichson, 1847 *stat. rest.* and *sensu n.* is re-elevated to subfamily and revised to include the tribes Heteronychini, Liparetrini, Maechidiini, Phyllotocini, Scitalini and Sericoidini. Sericoidinae Erichson, 1847 has formally priority over the younger name Liparetrinae Burmeister, 1855 (and other tribal names within the lineage) and also includes the tribe Automoliini not included in this analysis but being confirmed in previous molecular phylogenies to be part of the same lineage (Ahrens et al., 2014; Ahrens & Vogler, 2008). Based on morphological characters, other candidate members of this subfamily are Colymbomorphini, Comophorinini, Pachytrichini, and Phyllotocidiini, however, their phylogenetic placement has not been confirmed yet with molecular data.

The major subdivision of pleurostict scarabs between Rutelinae + Cetoniinae + Dynastinae + Melolonthinae on the one hand and Sericinae/Sericoidinae on the other was also retrieved by Ahrens (2006) based on a morphology-based phylogeny, although the position of Hopliinae was uncertain. In that study, Sericoidinae was recovered as paraphyletic in respect to Sericinae. The lineage Rutelinae + Cetoniinae + Dynastinae + Melolonthinae (Figure 1) is well characterised by placoid, round antennal sensilla (Bohacz et al., 2020; Pacheco et al., 2022) as well as by adjacent metatibial spines (Ahrens, 2006). While the monophyly of Sericinae is well supported by unique elongate, placoid antennal sensilla (Pacheco et al., 2022), Sericoidinae share in many respects a mix of ancestral and derived characters, and show highly plastic character transformations, for example in sensilla (from trichoid to scale like and placoid sensilla; Pacheco et al., 2022) or mouth parts (e.g., towards pollen feeding in Phyllotociini). All this made it so far difficult to characterise this diverse group based on morphology. Conversely, these results bring some generally accepted apomorphies into question. The metatibial spines separated by base of tarsomere 1, which has been for a long time considered to be a key character to Sericinae, has to be considered a plesiomorphy given the sister group relationship of Pleurosticti and Orphninae (which also share the same character). As a result, all Mesozoic fossils so far assigned to Sericinae based on this trait (see Krell, 2007) are very likely in need of a reclassification.

Apart of some of our robust results regarding the monophyly of Scarabaeidae, we consider this work also as a primer and starting point for further and more detailed phylogenomic research in Scarabaeoidea. In this, the generated transcriptomic data will serve as backbone for other approaches such as DNA target enrichment approaches, which would possibly allow to considerably extend the taxon sampling[1]. Because also dry museum specimens could be analysed this way, we expect that even yet entirely obscure lineages (e.g., Belohinidae, Dynamopodinae, Phaenomeridinae), which never have been considered in any phylogenetic analysis, will find their place in a phylogenetic tree.

AUTHOR CONTRIBUTIONS

Lars Dietz: Conceptualization; methodology; software; data curation; formal analysis; validation; visualization; investigation; writing—original draft; writing—review and editing. **Matthias Seidel:** Validation; writing—review and editing; investigation; resources. **Jonas Eberle:** Data curation; investigation; writing—review and editing; resources. **Bernhard Misof:** Writing—review and editing; resources; funding acquisition. **Thaynara L. Pacheco:** Resources; writing—review and editing; investigation. **Lars Podsiadlowski:** Methodology; resources; supervision; writing—review and editing; funding acquisition. **Sasanka Ranasinghe:** Resources; writing—review and editing; investigation. **Nicole L. Gunter:** Writing—review and editing; investigation. **Oliver Niehuis:** Conceptualization; methodology; supervision; writing—review and editing; writing—original draft; funding acquisition; investigation; project administration; resources. **Christoph Mayer:** Conceptualization; methodology; data curation; supervision; resources; writing—review and editing; writing—original draft; funding acquisition; validation; software. **Dirk Ahrens:** Conceptualization; investigation; methodology; data curation; supervision; resources; project administration; visualization; validation; funding acquisition; writing—original draft; writing—review and editing.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

The study was funded by the German Research Foundation grants (AH 175/6-1, AH175/6-2 to Dirk Ahrens, MI 649/18-1 to Bernhard Misof, NI 1387/6-1 and 6-2 to Oliver Niehuis) as well as by institutional funding of the Museum Koenig Bonn and the A. Koenig Stiftung Bonn.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All Supplementary Files including alignments of individual and concatenated loci and resulting trees (Supplement Files 7-12) are available in Dryad (doi:10.5061/dryad.d51c5b07h).

ORCID

Lars Dietz https://orcid.org/0000-0001-6469-381X
Matthias Seidel https://orcid.org/0000-0002-4913-8778
Jonas Eberle https://orcid.org/0000-0003-2519-0640

[1]Also to address the open systematic position of very recently established tribes (Evans & Smith, 2020; Smith & Evans, 2018).

Bernhard Misof https://orcid.org/0000-0003-4175-6798
Thaynara L. Pacheco https://orcid.org/0000-0001-9503-7751
Lars Podsiadlowski https://orcid.org/0000-0001-7786-8930
Sasanka Ranasinghe https://orcid.org/0000-0001-6455-8079
Nicole L. Gunter https://orcid.org/0000-0002-8355-0862
Oliver Niehuis https://orcid.org/0000-0003-4253-1849
Christoph Mayer https://orcid.org/0000-0001-5104-6621
Dirk Ahrens https://orcid.org/0000-0003-3524-7153

REFERENCES

Ahrens, D. (2006) The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). *Systematic Entomology*, 31, 113–144.

Ahrens, D., Schwarzer, J. & Vogler, A.P. (2014) The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B*, 281, 20141470.

Ahrens, D., Scott, M. & Vogler, A.P. (2011) A phylogeny of monkey beetles based on mitochondrial and ribosomal DNA (Coleoptera: Scarabaeidae: Hopliini). *Molecular Phylogenetics and Evolution*, 60, 408–415.

Ahrens, D. & Vogler, A.P. (2008) Towards the phylogeny of chafers (Sericini): analysis of alignment-variable sequences and the evolution of segment numbers in the antennal club. *Molecular Phylogenetics and Evolution*, 47, 783–798.

Ayivi, S.P.G., Tong, Y., Storey, K.B., Yu, D.N. & Zhang, J.Y. (2021) The mitochondrial genomes of 18 new Pleurosticti (Coleoptera: Scarabaeidae) exhibit a novel trnQ-NCR-trnI-trnM gene rearrangement and clarify phylogenetic relationships of subfamilies within Scarabaeidae. *Insects*, 12(11), 1025.

Balthasar, V. (1963) *Monographie der Scarabaeidae und Aphodiidae der palaearktischen und orientalischen Region (Coleoptera: Lamellicornia)*, Vol. 1–3. Prague: Tschechoslowak Akademie der Wissenschaften, p. 287.

Bauchop, T. & Clarke, R.T.J. (1975) Gut microbiology and carbohydrate digestion in the larva of *Costelytra zealandica* (Coleoptera: Scarabaeidae). *New Zealand Journal of Zoology*, 2, 237–243.

Berendse, F. & Scheffer, M. (2009) The angiosperm radiation revisited, an ecological explanation for Darwin's 'abominable mystery'. *Ecology Letters*, 12, 865–872.

Beza-Beza, C.F., Jiménez-Ferbans, L. & McKenna, D.D. (2020) Phylogeny and systematics of wood-degrading Neotropical bess beetles (Coleoptera: Passalidae: Passalinae). *Arthropod Systematics and Phylogeny*, 78(2), 287–308.

Bocak, L., Barton, C., Crampton-Platt, A., Chesters, D., Ahrens, D. & Vogler, A.P. (2014) Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Systematic Entomology*, 39, 97–110.

Bohacz, C., Harrison, J. & Ahrens, D. (2020) Comparative morphology of antennal surface in Pleurostict scarab beetles (Coleoptera). *Zoomorphology*, 139, 327–346.

Bouchard, P., Bousquet, Y., Davies, A., Alonso-Zarazaga, M., Lawrence, J., Lyal, C. et al. (2011) Family-group names in Coleoptera (Insecta). *ZooKeys*, 88, 1–972.

Browne, D.J. & Scholtz, C.H. (1995) Phylogeny of the families of Scarabaeoidea (Coleoptera) based on the characters of the hindwing articulation, hindwing base and wing venation. *Systematic Entomology*, 20, 145–173.

Browne, J. & Scholtz, C.H. (1998) Evolution of the scarab hindwing articulation and wing base: a contribution toward the phylogeny of the Scarabaeidae (Scarabaeoidea: Coleoptera). *Systematic Entomology*, 23, 307–326.

Browne, J. & Scholtz, C.H. (1999) A phylogeny of the families of Scarabaeoidea (Coleoptera). *Systematic Entomology*, 24, 51–84.

Cai, C., Tihelka, E., Giacomelli, M., Lawrence, J.F., Ślipiński, A., Kundrata, R. et al. (2022) Integrated phylogenomics and fossil data illuminate the evolution of beetles. *Royal Society Open Science*, 9, 211771.

Caterino, M.S., Hunt, T. & Vogler, A.P. (2005) On the constitution and phylogeny of Staphyliniformia (Insecta: Coleoptera). *Molecular Phylogenetics and Evolution*, 34(3), 655–672.

Cherman, M.A. & Morón, M.A. (2014) Validación de la familia Melolonthidae Leach, 1819 (Coleoptera: Scarabaeoidea). *Acta Zoológica Mexicana*, 30(1), 201–220.

Chernomor, O., von Haeseler, A. & Minh, B.Q. (2016) Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65(6), 997–1008.

Crowson, R.A. (1955) *The natural classification of the families of Coleoptera*. London: Nathaniel Lloyd & Co., Ltd, p. 187.

de Boer, H.J., Eppinga, M.B., Wassen, M.J. & Dekker, S.C. (2012) A critical transition in leaf evolution facilitated the cretaceous angiosperm revolution. *Nature Communications*, 3, 1221.

Eberle, J., Sabatinelli, G., Cillo, D., Bazzato, E., Šípek, P., Sehnal, R. et al. (2019) A molecular phylogeny of Melolonthinae chafer beetles revisits the polyphyly of Tanyproctini (Coleoptera, Scarabaeidae). *Zoologica Scripta*, 48, 349–358.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Computational Biology*, 7, e1002195.

Endrödi, S. (1966) Monographie der Dynastinae (Coleoptera, Lamellicornia). Teil 1. *Abhandlungen Staatliches Museum für Tierkunde*, 33, 1–460.

Erichson, W.F. (1848) Naturgeschichte der Insecten Deutschlands. Abt. I. *Coleoptera*, 3, 801–968.

Estes, A.M., Hearn, D.J., Snell-Rood, E.C., Feindler, M., Feeser, K., Abebe, T. et al. (2013) Brood ball-mediated transmission of microbiome members in the dung beetle, *Onthophagus taurus* (Coleoptera: Scarabaeidae). *PLoS One*, 8(11), e79061.

Evans, A.V. & Smith, A. (2020) On the tribal classification of the Nearctic Melolonthinae (Coleoptera: Scarabaeidae), with descriptions of new species of Acoma Casey, 1889. *Zootaxa*, 4748(1), 51–77.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652.

Grebennikov, V.V. & Scholtz, C.H. (2004) The basal phylogeny of Scarabaeoidea (Insecta: Coleoptera) inferred from larval morphology. *Invertebrate Systematics*, 18, 321–348.

Gunter, N.L., Weir, T.A., Slipinksi, A., Bocak, L. & Cameron, S.L. (2016) If dung beetles (Scarabaeidae: Scarabaeinae) arose in association with Dinosaurs, did they also suffer a mass co-extinction at the K-Pg boundary? *PLoS One*, 11(5), e0153570.

Guo, S., Lin, X. & Song, N. (2022) Mitochondrial phylogenomics reveals deep relationships of scarab beetles (Coleoptera, Scarabaeidae). *PLoS One*, 17(12), e0278820.

Hoang, D.T., Chernomor, O., von Haeseler, A. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522.

Howden, H.F. (1982) Larval and adult characters of *Frickius* Germain, its relationships to the Geotrupini, and a phylogeny of some major taxa in the Scarabaeoidea (Insecta: Coleoptera). *Canadian Journal of Zoology*, 60, 2713–2724.

Hoyt, C.P., Osborne, G.O. & Mulcock, A.P. (1971) Production of an insect sex attractant by symbiotic bacteria. *Nature*, 230, 472–473.

Huang, S.W., Zhang, H.Y., Marshall, S. & Jackson, T.A. (2010) The scarab gut: a potential bioreactor for bio-fuel production. *Insect Science*, 17(3), 175–183.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R. et al. (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, 318, 1913–1916.

Iablokoff-Khnzorian, S.M. (1977) Über die Phylogenie der Lamellicornia. *Entomologische Abhandlungen des Staatlichen Museum für Tierkunde*, 41, 135–199.

Janssens, A. (1949) Table synoptique et essai de classification pratique des Coléoptères Scarabaeidae. *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique*, 25, 1–30.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780.

Kishino, H. & Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution*, 29, 170–179.

Kishino, H., Miyata, T. & Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31, 151–160.

Kohlmann, B. & Morón, M.A. (2003) Análisis histórico de la clasificación de los Coleoptera Scarabaeoidea o Lamellicornia. *Acta Zoológica Mexicana*, 90, 175–280.

Krell, F.T. (2007) Catalogue of fossil Scarabaeoidea (Coleoptera: Polyphaga) of the Mesozoic and tertiary – version 2007. DMNS Technical Report, 2007-8, 1–79.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A. et al. (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47(D1), D807–D811.

Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. (2021) FelixKrueger/TrimGalore: v0.6.6 – DOI via Zenodo (TrimGalore v.0.6.7). Zenodo. https://doi.org/10.5281/zenodo.5127899

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W. et al. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, 7, 10.

Lacroix, M. (2007) Pachydeminae du Monde (Scarabaeoidea, Melolonthidae). Genera et Catalogue commenté. M. Lacroix/Collection Hannetons, 450 pp.

Lacroix, M. (2014) Melolonthidae Malgaches (Coleoptera, Scarabaeoidea). Catalogue commenté. M. Lacroix/Collection Hannetons, 260 pp.

Lawrence, J.F. & Newton, A.F. (1995) Families and subfamilies of Coleoptera (with selected genera, notes and references, and data on family-group names). In: Pakaluk, J. & Slipinski, S.A. (Eds.) *Biology, phylogeny, and classification of Coleoptera: papers celebrating the 80th birthday of Roy A. Crowson*. Warsaw: Muzeum I Instytut Zoologii PAN, pp. 779–1006, 277.

Lawrence, J.F. & Newton, A.F., Jr. (1982) Evolution and classification of beetles. *Annual Review of Ecology, Evolution, and Systematics*, 13, 261–290.

Lawrence, J.F., Ślipiński, A., Seago, A.E., Thayer, M.K., Newton, A.F. & Marvaldi, A.E. (2011) Phylogeny of the Coleoptera based on morphological characters of adults and larvae. *Annales Zoologici*, 61, 1–217.

Li, B., Lopes, J.S., Foster, P.G., Embley, T.M. & Cox, C.J. (2014) Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Molecular Biology and Evolution*, 31, 1697–1709.

Lu, Y., Ballerio, A., Wang, S., Zou, Z., Gorb, S.N., Wang, T. et al. (2022) The evolution of conglobation in Ceratocanthinae. *Communications Biology*, 5, 777.

Manni, M., Berkeley, M.R., Seppey, M. & Zdobnov, E.M. (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654.

Martin, M.M. (1983) Cellulose digestion in insects. *Comparative Biochemistry and Physiology*, 75A(3), 313–324.

Martin, M.M., Jones, C.G. & Bernays, E.A. (1991) The evolution of cellulose digestion in insects [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 333(1267), 281–288.

McKenna, D.D., Farrell, B.D., Caterino, M.S., Farnum, C.W., Hawks, D.C., Maddison, D.R. et al. (2014) Phylogeny and evolution of Staphyliniformia and Scarabaeiformia: Forest litter as a stepping stone for diversification of nonphytophagous beetles. *Systematic Entomology*, 40, 35–60.

McKenna, D.D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D.J. et al. (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences of the U S A*, 116(49), 24729–24737.

McKenna, D.D., Wild, A.L., Kanda, K., Bellamy, C.L., Beutel, R.G., Caterino, M.S. et al. (2015) The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the cretaceous terrestrial revolution. *Systematic Entomology*, 40, 835–880.

Medvedev, S.I. (1976) Sistematike i filogenii plastinchatouyx zhukov Palearktiki. *Entomologischeskoe Obozrenje*, 55, 400–409.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534.

Misof, B. & Misof, K. (2009) A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, 58, 21–34.

Morón, M.A. (1984) *Escarabajos, 200 millones de años de evolución. Publicación 14*. México: Instituto de Ecología, p. 131.

Morón, M.A. (1997) Antecedentes. In: Morón, M.A., Ratcliffe, B.C. & Deloya, C. (Eds.) *Atlas de los escarabajos de México. Coleoptera Lamellicornia. Vol. I. Familia Melolonthidae*. México: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad y Sociedad Mexicana de Entomología, pp. 1–8, 280.

Morón, M.A. (2003) *Atlas de los escarabajos de México. Coleoptera: Lamellicornia. Vol. II Familias Scarabaeidae, Trogidae, Passalidae y Lucanidae*. Barcelona: Argania editio, p. 227.

Neita-Moreno, J.C., Agrain, F.A., Eberle, J., Ahrens, D. & Pereyra, V. (2019) On the phylogenetic position and systematics of extant and fossil Aclopinae (Coleoptera: Scarabaeidae). *Systematic Entomology*, 44, 709–727.

Nikolajev, G.V. (1995) New data on the systematics of the subfamily Ochodaeinae (Coleoptera, Scarabaeidae). *Zoologicheskiy Zhurnal*, 74, 72–82.

Pacheco, T.L., Bohacz, C., Ballerio, A., Schoolmeesters, P. & Ahrens, D. (2022) Revisiting trends of morphology of antennal sensilla in Scarabaeoid beetles. *Zoomorphology*, 141, 315–326.

Paulian, R. & Baraud, J. (1982) *Faune des Coléoptères de France, II. Lucanoidea et Scarabaeoidea*. Paris: Éditions Lechevalier, p. 477.

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K. et al. (2017) Evolutionary history of the hymenoptera. *Current Biology*, 27(7), 1013–1018.

Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R.S. et al. (2017) Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, 18, 111.

Ratcliffe, B.C. & Jameson, M.L. (2004) The revised classification for Scarabaeoidea: what the hell is going on? *Scarabs*, 15, 3–10.

Scholtz, C.H. (1990) Phylogenetic trends in the Scarabaeoidea (Coleoptera). *Journal of Natural History*, 24, 1027–1066.

Scholtz, C.H. & Grebennikov, V.V. (2005) Scarabaeoidea Latreille, 1802. In: Beutel, R.G. & Leschen, R.A.B. (Eds.) *Handbook of zoology, Coleoptera, beetles. Volume 1. Morphology and systematics (Archostemata, Adephaga, Myxophaga), Polyphaga partim*. Berlin, New York: Walter de Gruyter, pp. 368–425, 567.

Sharma, P.P., Kaluziak, S.T., Pérez-Porro, A.R., González, V.L., Hormiga, G., Wheeler, W.C. et al. (2014) Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, 31, 2963–2984.

Shimodaira, H. (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, 51, 492–508.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114–1116.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212.
Šípek, P., Fabrizi, S., Eberle, J. & Ahrens, D. (2016) A molecular phylogeny of rose chafers (Coleoptera: Scarabaeidae: Cetoniinae) reveals a complex and concerted morphological evolution related to their flight mode. *Molecular Phylogenetics and Evolution*, 101, 163–175.
Smith, A. & Evans, A.V. (2018) Taxonomic review of Athliini (Coleoptera: Scarabaeidae: Melolonthinae), a new tribe of scarab beetles endemic to South America. *Zootaxa*, 4471(2), 279–308.
Smith, A.B.T. (2006) A review of the family-group names for the superfamily Scarabaeoidea (Coleoptera) with corrections to nomenclature and a current classification. *Coleopterists Society Monograph*, 5, 144–204.
Smith, A.B.T., Hawks, D.C. & Heraty, J.M. (2006) An overview of the classification and evolution of the major scarab beetle clades (Coleoptera: Scarabaeoidea) based on preliminary molecular analyses. *Coleopterists Society Monograph*, 5, 35–46.
Song, N. & Zhang, H. (2018) The mitochondrial genomes of phytophagous scarab beetles and systematic implications. *Journal of Insect Science*, 18, 11.
Strimmer, K. & Rambaut, A. (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society B*, 269, 137–142.
Strimmer, K. & von Haeseler, A. (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 6815–6819.
Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34, W609–W612.
Taylor, L., Banwart, S., Leake, J. & Beerling, D.J. (2011) Modeling the evolutionary rise of ectomycorrhiza on sub-surface weathering environments and the geochemical carbon cycle. *American Journal of Science*, 311, 369–403.
Taylor, L.L., Banwart, S.A., Valdes, P.J., Leake, J.R. & Beerling, D.J. (2012) Evaluating the effects of terrestrial ecosystems, climate and carbon dioxide on weathering over geological time: a global-scale process-based approach. *Philosophical Transactions of the Royal Society B*, 367, 565–582.
Timmermans, M.J.T.N., Barton, C., Haran, J., Ahrens, D., Culverwell, C.L., Ollikainen, A. et al. (2016) Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Genome Biology and Evolution*, 8, 161–175.
Toussaint, E.F.A., Seidel, M., Arriaga-Varela, E., Hájek, J., Král, D., Sekerka, L. et al. (2017) The peril of dating beetles. *Systematic Entomology*, 42, 1–10.
Wada, N., Sunairi, M., Anzai, H., Iwata, R., Yamane, A. & Nakajima, M. (2014) Glycolytic activities in the larval digestive tract of *Trypoxylus dichotomus* (Coleoptera: Scarabaeidae). *Insects*, 5(2), 351–363.
Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 153.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Files S1–S12. See Dryad link: https://doi.org/10.5061/dryad.d51c5b07h.

File S1. Perl script to filter Trinity contigs for the longest variant. (trinity_longest_d.pl).

File S2. Perl script to remove unaligned regions from hmmalign output. (hmmalign_cut2_d.pl).

File S3. Perl script to remove third codon positions from an alignment. (extract_codpos_d.pl).

File S4. Perl script to concatenate gene alignments into a partitioned superalignment (concat_eogs_part_d.pl).

File S5. Perl script to remove alignment positions present in less than a given number of taxa. (removegaps_d.pl).

File S6. Perl script to calculate average pairwise identities between sequences within alignments. (pairwise_id2.pl).

File S7. Coalescent-based phylogenetic trees from ASTRAL analysis (astral_trees.zip; see astral-trees-descriptions.txt for descriptions).

File S8. Concatenated sequence alignments of coding sequences of all loci in a dataset in FASTA format (concat_alignments.zip; see concat-alignments-descriptions.txt for descriptions).

File S9. Maximum-likelihood phylogenetic trees created with IQ-TREE based on concatenated alignments of coding sequences of USCO loci (concat_trees.zip; see concat-trees-descriptions.txt for descriptions).

File S10. Sequence alignments of coding sequences of individual loci in FASTA format (gene_alignments.zip; see gene-alignments-descriptions.txt for descriptions).

File S11. Maximum-likelihood phylogenetic trees created with IQ-TREE for individual loci (gene_trees.zip; see gene-trees-descriptions.txt for descriptions).

File S12. Partition files for concatenated gene alignments in NEXUS format (partition_files.zip; see partition-files-descriptions.txt for descriptions).

Table S1. Overview on samples included in this study, referring to taxon name, current systematic placement, collection site, voucher number, accession numbers of sequence data repositories. The table also includes information on the number of sequences in each transcriptome, BUSCO results, number of recovered genes (Endopterygota USCOs vs. Coleoptera-specific genes), alignment completeness [in %] (Endopterygota USCOs, hmmalign vs. MAFFT vs. Coleoptera-specific genes, hmmalign vs. MAFFT).

Table S2. Likelihood tests regarding the monophyly of different groupings using the Endopterygota USCO dataset, with constraint trees using a variety of resampling tests in IQ-TREE using the RELL approximation including bootstrap proportion (BP), Kishino–Hasegawa test (KH), Shimodaira–Hasegawa test (SH), expected likelihood weights (ELW), and the approximately unbiased (AU) test.

How to cite this article: Dietz, L., Seidel, M., Eberle, J., Misof, B., Pacheco, T.L., Podsiadlowski, L. et al. (2023) A transcriptome-based phylogeny of Scarabaeoidea confirms the sister group relationship of dung beetles and phytophagous pleurostict scarabs (Coleoptera). *Systematic Entomology*, 1–15. Available from: https://doi.org/10.1111/syen.12602