# **Supporting Information**

## **McKenna et al. 10.1073/pnas.0810618106**

## **SI Materials and Methods**

Taxon Sampling. We analyzed up to 8 kb of DNA sequence data from a worldwide sample of 135 weevil genera representing all 7 weevil families, all 26 weevil subfamilies, and 97 genera representing most major tribes in the extraordinarily diverse family Curculionidae [\(supporting information \(SI\) Table S1](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST1) and [Table S3\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST3). Outgroups included 7 subfamilies of basal Chrysomeloidea and *Ericmodes sylvaticus* (Protocucujidae), a member of the closely related superfamily Cucujoidea. Six genes (2 mitochondrial and 4 nuclear) were used in this study: *cytochrome oxidase I*, *18S rDNA*, *28S rDNA*, *16S rDNA*, *Elongation Factor-1a*, and *Arginine Kinase* (*AK*). All *16S rDNA* (1), and select other sequences, were obtained from GenBank. For some genera, chimeras were constructed from sequences for different species to reduce the amount of missing data. All taxa except *Atractuchus* (*18S*), *Brachycerus* (*28S, 18S*), *Bruchela* (*18S*), *Caenominurus* (*28S, 18S*), *Gonipterus* (*18S, 28S*), *Ithycerus* (*18S, EF1a*), *Microcerus* (*18S*), and *Nemonyx* (*18S*) were represented by DNA sequence data from at least 3 of the 6 genes targeted. We used a *16S* rDNA sequence of *Cheloderus* (Oxypeltidae) from Gen-Bank in lieu of a comparable sequence for the closely related outgroup taxon *Oxypeltus* (Oxypeltidae). Overall, our supermatrix contained sequences for  $\approx 70\%$  of the 858 possible taxonby-gene combinations (see [Table S3\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST3). Voucher specimens are deposited at the Harvard University Museum of Comparative Zoology, and nucleotide sequences newly determined here have been deposited in GenBank.

**DNA Isolation and Amplification.** Protocols for DNA isolation and amplification are reported in refs. 2 and 3, with differences as follows: We amplified double-stranded DNA in 8- to  $25-\mu l$ reactions (depending on the gene amplified and other factors) using published or optimized primers [\(Table S5\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST5). All reactions were initially denatured at 94 °C, but the duration of denaturation, and the temperature, duration, and number of cycles of annealing and extension varied by gene. All reactions were performed on MJ Dyad, MJ PTC-200 (MJ Research), and MyCycler (Bio-Rad) thermal cyclers. Amplified *18S* PCR products were cleaned using an exonuclease and alkaline phosphatase protocol (3). Amplified fragments of all other genes were gel purified using a Qiagen QIAquick Gel Purification Kit (Qiagen) before sequencing.

**DNA Sequencing.** Amplified, cleaned PCR products were used in sequencing reactions employing BigDye Terminator chemistry [Applied Biosystems Inc.(ABI)]. Primers used for amplification served as sequencing primers, except when additional internal primers were designed to provide overlapping sequences for large fragments (see [Table S5\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST5). Cycle sequencing reactions were mostly performed in 10- $\mu$ l reactions: 1.5  $\mu$ l ABI Prism BigDye Terminator 3.1, 1.0  $\mu$ l 5× buffer, and 0.33  $\mu$ l each (10  $\mu$ M) primer. The remainder of the mixture was composed of water and template DNA (varied by gene and as needed to adjust DNA concentration). Cycle sequencing reactions consisted of an initial denaturation at 94 °C for 2 min, followed by 25 cycles of 10s at 94 °C denaturation, 5 s at 57 °C annealing and 4 min at 60 °C. Sequencing was performed on ABI 3100 and ABI 3730 DNA sequencers.

**Sequence Alignment.** DNA sequences were edited and preliminarily aligned using the program Sequencher 4.6 (Genecodes). Subsequent alignment was performed with the program ClustalX 1.831 (4) using the default settings. The resulting alignment for each gene was adjusted ''by eye'' in the program MacClade 4.06 (5). Regions of ambiguous alignment in *16S*, *18S*, and *28S*, and introns in  $EF \, 1$ - $\alpha$  and  $AK$  were removed. The individual alignments for each gene were then concatenated in MacClade, and the resulting aligned matrix (6 genes,  $\approx 8$  kb) used in subsequent analyses.

**Phylogenetic Analyses.** Phylogenetic analyses were conducted on the 8-kb molecular supermatrix using Bayesian and ML inference. A partitioned ML BS analysis (1,000 inferences, 12 partitions, CAT substitution model, individual per partition branchlength optimization) was implemented in the program RAxML version 7.04 (6) using the CIPRES cluster at the San Diego Supercomputing Center. Partitions were: *28S*, *18S*, *16S*, *COI* (separate partitions for first, second, and third positions), *EF1-a* (separate partitions for first, second, and third positions), and *AK* (separate partitions for first, second, and third positions). Partitioned BI analyses (12 partitions,  $GTR+I+\Gamma$ , estimated base frequencies, four  $\Gamma$  categories) were implemented in the program BEAST 1.4.7 (7). Analyses employing an unweightedpair group method with arithmatic mean or random starting tree failed to execute (returning the "initial model is invalid" error), so we obtained a more optimal starting tree by executing a preliminary run of 10<sup>6</sup> generations with monophyly constraints on the ingroup, outgroup, all weevil families, and select subfamilies of Curculionidae (Dryophthorinae, Platypodinae, and Scolytinae). We used the last tree (with branch lengths) obtained from this analysis as a starting tree for subsequent more thorough analyses. We ran 2 separate BEAST analyses on the maximum- and minimum-age constrained data sets, each with a constraint on the monophyly of the ingroup (but no other monophyly constraints). We ran two BEAST analyses on the maximum- and minimum-age data sets (65–75 million generations, preburnin 10<sup>6</sup> generations, sampling every 1,000 generations), for a total of 4 separate analyses. All trees were rooted with *Ericmodes sylvaticus* based on refs. 8 and 9. Graphical and statistical analyses implemented in the program Tracer 1.4 (10) were used to assess convergence and otherwise check performance and accuracy of the BEAST analyses. Specifically with regard to convergence, a trace plot of log-likelihoods from the BEAST output (ultimately, the last 5 million generations from each run) showed no obvious trends or large-scale fluctuations. This suggested that the Markov Chain Monte Carlo had converged and that mixing was adequate. We also used Tracer to assess effective sample size and to analyze/evaluate the marginal posterior probability distribution of select parameters (e.g., mutation rate and tree height) from the BEAST analyses. Based on these analyses, we imposed a very conservative burn-in on each tree file, then combined the last 5,000 trees from each of the paired minimum- and maximum-age analyses (for a total of 10,000 trees), and used these to estimate PPs, to obtain maximum clade-credibility trees, and to estimate divergence times (see below) and associated 95% confidence intervals for the minimum- and maximum-age analyses [using the programs LogCombiner 1.4.7, PAUP\* 4.03b10 (11), and TreeAnnotator 1.4.7]. BS values  $\geq 90\%$  (under ML) or posterior probability values  $\geq 0.95$ (under BI) were considered to constitute strong internodal support, while BS values  $\geq 75\%$  (and <90%) or posterior probability values  $\geq 0.80$  (and  $\leq 0.95$ ) were considered to constitute moderate internodal support.

**Testing Alternative Phylogenetic Hypotheses.** We investigated the degree to which select alternative phylogenetic hypotheses were supported by our data by estimating the posterior probabilities of alternative topologies (under BI), and by comparing the ML trees obtained with and without monophyly constraints on each group of interest using the KH test (12), as implemented in PAUP. The KH test is in principle not appropriate in this situation; that is, it is insufficiently conservative because the individual topologies compared were chosen in advance. However, because we recovered no significant *P*-values, more appropriate tests, such as the Shimodaira-Hasegawa test (13) and the Approximately Unbiased test (14), which are more conservative, also will not recover significant *P*-values. For the KH test, constraint trees were prepared in MacClade, and a thorough ML search was performed on each in RAxML using the CIPRES cluster (12 partitions,  $GTR+I+G$  substitution model, individual per partition branch-length optimization). The significance of differences between trees was determined with a BS test (1,000 replicates, resampling extimated log likelihoods approximation) imposing the parameter estimates and base frequencies estimated from the ML tree without partitions.

**Divergence-Time Estimates.** Divergence times were coestimated with phylogeny using the Bayesian relaxed molecular clock method (15) in the program BEAST (7). We assumed the uncorrelated lognormal prior model of rate change, a Yule prior process to model speciation, and used automatic tuning of operators. We conservatively selected and applied fossil age

- 1. Wink M, Mikes Z, Rheinheimer J (1997) Phylogenetic relationships in weevils (Coleoptera: Curculinoidea) inferred from nucleotide sequences of mitochondrial 16S rDNA. *Naturwissenschaften* 84:318–321.
- 2. McKenna DD, Farrell BD (2005) Molecular phylogenetics and evolution of host plant use in the Neotropical rolled leaf 'hispine' beetle genus *Cephaloleia* (Chevrolat) (Chrysomelidae: Cassidinae). *Mol Phylogenet Evol* 37:117–131.
- 3. McKenna DD, Farrell BD (2006) Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. *Proc Natl Acad Sci USA* 103:10947–10951.
- 4. Thompson TJ, Gibson F, Plewniak F, Jeanmougin D, Higgins G (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
- 5. Maddison WP, Maddison DR (2002) MacClade: Analysis of phylogeny and character evolution v. 4.05 (Sinauer, Sunderland MA).
- 6. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web-Servers. *Syst Biol* 75:758–771.
- 7. Drummond AJ, Rambaut (2007) A BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- 8. Hunt T, et al. (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318:1913–1916.
- 9. McKenna DD, Farrell BD (2009) in *Timetrees of Life,* eds Hedges B, Kumar S (Oxford Univ Press, Oxford), pp. 278–289.

constraints from 2 recent reviews (16, 17), using only the oldest fossils that could be unequivocally assigned (based on character evidence) to extant weevil subfamilies or families [\(Table S4\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST4). Consequently, several fossils were excluded from use as constraints because their age or identity was uncertain, or their placement was rendered equivocal by paraphyly or polyphyly in preliminary analyses. The stage boundaries and terminology we used follow ref. 18. Prior estimates for the divergence dates for selected nodes were specified using uniform distributions, except for the ingroup root node, for which we had sufficient information to specify a transformed lognormal distribution with a ''hard'' minimum bound based on the minimum age of the oldest unequivocal fossil weevil (150.8 Ma; with 0% probability of the divergence being younger than this date) (see [Table S4\)](http://www.pnas.org/cgi/data/0810618106/DCSupplemental/Supplemental_PDF#nameddest=ST4), a mean estimate of 171.5 Ma based on a mean estimate for the age of Curculionoidea from ref. 8, and a ''soft'' maximum bound based on the maximum age reported in ref. 8 for the series Cucujiformia (236.2 Ma, with 5% probability of the divergence being older than this date). The minimum bounds placed on uniform distributions were based on the minimum age of the oldest unequivocal fossil for a given taxon. The maximum bounds represent the oldest age plausible for a given taxon based on palaeontological or other evidence. When the stratigraphic position of a fossil was not well resolved, or the age was reported with stage-level (or similar) resolution, or was otherwise uncertain, we used the accepted absolute age of the upper and lower boundary of the reported formation or stage interval as constraints in separate minimum- and maximum-age analyses.

- 10. Rambaut A, Drummond AJ (2007) Tracer v. 1.4, Available from http://beastbioedacuk/ Tracer.
- 11. Swofford DL (2002) Phylogenetic Analysis Using Parsimony (\*and other Methods) v. 4.03b10 (Sinauer, Sunderland MA).
- 12. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimation of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol* 29:170–179.
- 13. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116.
- 14. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
- 15. Drummond AJ, Ho SYW, Philipps MJ, Rambaut (2006) A Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
- 16. Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa* 1668:491–520.
- 17. Gratshev VG, Zherikhin VV (2003) in *Proceedings of the 2nd Congress on Palaeoentomology,* eds Krzeminska E, Krzeminski W (Krakow, Poland), *Acta Zool Cracov* 46 *Suppl* pp 129–138.
- 18. Ogg JG, Ogg G, Gradstein FM (2008) *The Concise Geologic Time Scale* (Cambridge Univ Press, Cambridge).

**Table S1. Extant families and subfamilies of Curculionoidea recognized in this study, following Oberprieler, Marvaldi, and Anderson (1)**



PNAS PNAS

1. Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils every-where. *Zootaxa* 1668:491–520.

**Table S2. Molecular divergence dates (point estimates based on the minimum-age and maximum-age maximum-clade credibility trees) and 95% confidence intervals (CI) calculated for the ages of family-level nodes and monophyletic (or near) subfamilies of Curculionidae**



**Table S3. Weevil genera sampled and the geographic distribution of each as reported in Alonso-Zarazaga and Lyal, except for Platypodinae and Scolytinae which are not included in this reference, and for which we have listed the geographic region from which the specimen was collected (not the entire known distribution)**







Abbreviations: Afrotropical (AT), Australasian (AU), Eastern Palearctic (EP), Holarctic (HA), Nearctic (NA), Neotropical (NT), Oriental (OL), Palearctic (PA), and Western Palearctic (WP). Detailed collection data is available from D.D.M.

1The placement of Bagoini is tentative, based on morphology, but it has formerly been assigned to Curculioninae, Molytinae, and so forth. [Alonso-Zarazaga MA, Lyal CHC (1999) *A World Catalogue of Families and Genera of Curculionoidea (Insecta: Coleoptera) (excepting Scolytidae & Platypodidae*) (Entomopraxis SCP, Barcelona)].

#### **Table S4. Primers used for amplification and sequencing**

NAS PNAS



1. Halaynch K, Lutz RA, Vrijenhoek RC (1998) Evolutionary origins and age of vestimentiferan tube-worms. *Cah Biol Mar* 39:355–358.

2. Whiting MF (2002) Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zool Script* 31(1):93–104.

3. Sequeira AS, Normark BB, Farrell BD (2000) Evolutionary assembly of the conifer fauna: Distinguishing ancient from recent associations in bark beetles. *Proc R Soc Lond B Biol Sci* 267:2359–2366.

4. Van der Auwera G, Chapelle S, De Wachter R (1994) Structure of the large ribosomal subunit RNA of *Phytophthora megasperma*, and phylogeny of the Oomycetes. *FEBS Lett* 338(2):133–136.

5. Mallatt J, Sullivan J (1998) 28S and 18S ribosomal DNA sequences support the monophyly of lampreys and hagfishes. *Mol Biol Evol* 15:1706–1718.<br>6. Danforth BN, Ji S (1998) Elongation factor-1alpha occurs as two copies i

7. Normark BB, Jordal BH, Farrell BD (1999) Origin of a haplodiploid beetle lineage. *Proc R Soc Lond, B, Biol Sci* 266:2253–2259.

8. McKenna DD, Farrell BD (2005) Molecular phylogenetics and evolution of host plant use in the Neotropical rolled leaf &lquote;hispine' beetle genus *Cephaloleia* (Chevrolat) (Chrysomelidae: Cassidinae). *Mol Phylogenet Evol* 37:117–131.

9. Danforth BN, Lin C, Fang J (2005) How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Syst Entomol* 30:549–562.

10. Jordal BH (2007) Reconstructing the phylogeny of Scolytinae and close allies: Major obstacles and prospects for a solution. *U S Dep Agric For Serv Res Paper* 45:3–8.

11. Simon C, et al. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* 87:651–701.

12. Folmer O, Black MB, Hoch W, Lutz RA, Vrijehock RC (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Bio Biotechnol* 3:294–299.

### **Table S5. Fossil ages applied as constraints**

PNAS

JAC<br>J



Asterisks indicate that analyses were done with both minimum and maximum ages to account for uncertainty. Character evidence in support of the placements of these fossils in established groups can be found in the references cited.

1. Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa* 1668:491–520.

2. Gratshev VG, Zherikhin VV (2003) in *Proceedings of the 2nd Congress on Palaeoentomology,* eds Krzeminska E, Krzeminski W (Krakow, Poland), *Acta Zool Cracov* 46 *Suppl* pp 129–138.

3. Gratshev VG, Zherikhen VV (2000) in *Studies on Fossils in Amber, with Particular Reference to the Cretaceous of New Jersey,* ed Grimaldi D. (Backhuys, Leiden), pp 241–254.

4. Zherikhin VV, Gratshev VG (2004) Fossil curculionoid beetles (Coleoptera, Curculionoidea) from the Lower Cretaceous of Northeastern Brazil. *Paleontol J* 38:528–537. 5. Grimaldi DA, Maisey JG (1990) in *Insects from the Santana Formation, Lower Cretaceous, of Brazil,* ed Grimaldi DA, *Bull American Mus Nat Hist* 195:5–14.

6. Kuschel G (1983) Past and present of the relict family Nemonychidae (Coleoptera: Curculionoidea). *GeoJournal* 7:499–504.

7. Scudder SH (1893) *Tertiary Rhynchophorous Coleoptera of the United States. U.S. Geological Survey Monograph* 21 (Government Printing Office, Washington D.C.).

8. Skalski AW, Veggiani A (1990) Fossil resin in Sicily and the Northern Apennines: Geology and organic content. *Prace Muzeum Ziemi* 41:37–49.

9. Kohring R, Schlüter T (1989) Historische und paläontologische Bestandsaufnahme des Simetits, eines fossilen Harzes mutmaßlich mio/pliozänen Alters aus Sizilien. *Doc Nat* 56:33–58.

10. Britton EB (1960) Beetles from the London Clay (Eocene) of Bognor Regis, Sussex. *Bull Br Mus Nat Hist (Geol*) 4(2):27–50.

11. Kuschel G, Oberprieler RG, Rayner RJ (1994) Cretaceous weevils from southern Africa, with description of a new genus and species (Coleoptera : Curculionoidea). *Entomol Scand* 25:137–149.

12. Arnoldi LV (1977) Rhynchophora (in part), Mesozoic Beetles. *Tr Paleontol Instl Akad Nauk SSSR* 161:144–176 [in Russian].