Current Biology 22, 1309–1313, July 24, 2012 ©2012 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2012.05.018

Report

# Genomic and Morphological Evidence Converge to Resolve the Enigma of Strepsiptera

Oliver Niehuis,[1,3,9,*] Gerrit Hartig,[1,4,9] Sonja Grath,[4,9] Hans Pohl,[5] Jörg Lehmann,[6] Hakim Tafer,[6] Alexander Donath,[1] Veiko Krauss,[6] Carina Eisenhardt,[7] Jana Hertel,[6] Malte Petersen,[1] Christoph Mayer,[1] Karen Meusemann,[1] Ralph S. Peters,[2] Peter F. Stadler,[6] Rolf G. Beutel,[5] Erich Bornberg-Bauer,[4] Duane D. McKenna,[8] and Bernhard Misof[1,*]

[1]Center for Molecular Biodiversity Research
[2]Department Arthropoda
Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany
[3]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
[4]Institute for Evolution and Biodiversity, Evolutionary Bioinformatics Group, University of Muenster, 48149 Muenster, Germany
[5]Institute of Systematic Zoology and Evolutionary Biology with Phyletisches Museum, University of Jena, 07743 Jena, Germany
[6]Institute of Computer Science, University of Leipzig, 04107 Leipzig, Germany
[7]Institute of Biology II, University of Leipzig, 04103 Leipzig, Germany
[8]Department of Biological Sciences, University of Memphis, Memphis, TN 38152, USA

## Summary

**The phylogeny of insects, one of the most spectacular radiations of life on earth, has received considerable attention [1–3]. However, the evolutionary roots of one intriguing group of insects, the twisted-wing parasites (Strepsiptera), remain unclear despite centuries of study and debate [1, 2, 4–11]. Strepsiptera exhibit exceptional larval developmental features, consistent with a predicted step from direct (hemimetabolous) larval development to complete metamorphosis that could have set the stage for the spectacular radiation of metamorphic (holometabolous) insects [1, 12, 13]. Here we report the sequencing of a Strepsiptera genome and show that the analysis of sequence-based genomic data (comprising more than 18 million nucleotides from nearly 4,500 genes obtained from a total of 13 insect genomes), along with genomic metacharacters, clarifies the phylogenetic origin of Strepsiptera and sheds light on the evolution of holometabolous insect development. Our results provide overwhelming support for Strepsiptera as the closest living relatives of beetles (Coleoptera). They demonstrate that the larval developmental features of Strepsiptera, reminiscent of those of hemimetabolous insects, are the result of convergence. Our analyses solve the long-standing enigma of the evolutionary roots of Strepsiptera and reveal that the holometabolous mode of insect development is more malleable than previously thought.**

## Results and Discussion

We sequenced the genome of *Mengenilla moldrzyki* (Figure 1A), a newly discovered species belonging to the early-divergent strepsipteran family Mengenillidae [14]. The draft genome of *M. moldrzyki* was sequenced from genomic DNA using 454-pyrosequencing technology to an estimated coverage of $\geq$14×. De novo assembly of the genome from the obtained reads produced 13,919 scaffolds and 87,021 nonredundant contigs spanning a total of 165 Mb. We inferred 16,772 ab initio models of nuclear-encoded protein-coding (NEPC) genes, of which 13,296 were supported by extrinsic evidence (e.g., transcripts). We also annotated protein domains, DNA methylation-related proteins, noncoding RNAs, and the complete mitochondrial genome (see Tables S1–S7 available online; Figure S1).

The Strepsiptera genome sequence data were exploited to test the following four current competing hypotheses about the phylogenetic origin of Strepsiptera (Figure 1B): (1) Strepsiptera are the sister group of all remaining insects with complete metamorphosis (Holometabola) [15], (2) Strepsiptera are the sister group of beetles (Coleoptera) [8], (3) Strepsiptera are a derived lineage of polyphagan beetles [9], and (4) Strepsiptera are the sister group of Diptera [5, 16, 17]. For this purpose, we assessed orthology among the predicted NEPC genes in the *M. moldrzyki* genome and those of 11 other insect species with sequenced genomes (representing Coleoptera, Diptera, Hymenoptera, Lepidoptera, and Acercaria; [18–28]) using a Markov Cluster algorithm implemented in the software OrthoMCL [29] (Figure 1B). In total, we identified 15,614 groups of orthologous NEPC genes; 4,485 of these groups contained sequences of at least one representative per insect order.

After removing ambiguously aligned sites (identified at the amino acid level), we evaluated the aligned amino acid and correspondingly aligned nucleotide sequences of the 4,485 groups of orthologous NEPC genes for their degree of substitutional saturation, relative compositional variance, and for the ratio of potential synapomorphic to potential autapomorphic characters. Compositional heterogeneity among sequences was lowest and the number of potentially informative characters for inferring inter- and intraordinal phylogenetic relationships was highest for RY-recoded (A and G → R; T and C → Y) second codon positions only, as compared to nonrecoded or differently recoded data sets or data subsets (Figure S3). The complete matrix of RY-recoded second codon positions from the 4,485 groups of orthologous NEPC genes consisted of approximately 1.8 million characters—the largest data set ever compiled for inferring the phylogenetic origin of Strepsiptera or any other insect order (Table S8).

We analyzed the RY-recoded second codon positions using maximum likelihood (ML) tree inference. The inferred phylogenetic tree (Figure 1B; Figure S2) was fully resolved and received maximal statistical support for all branches. All intra- and interordinal relationships are fully consistent with the current view of insect phylogenetic relationships [2]

[9]These authors contributed equally to this work
*Correspondence: o.niehuis.zfmk@uni-bonn.de (O.N.), b.misof.zfmk@uni-bonn.de (B.M.)

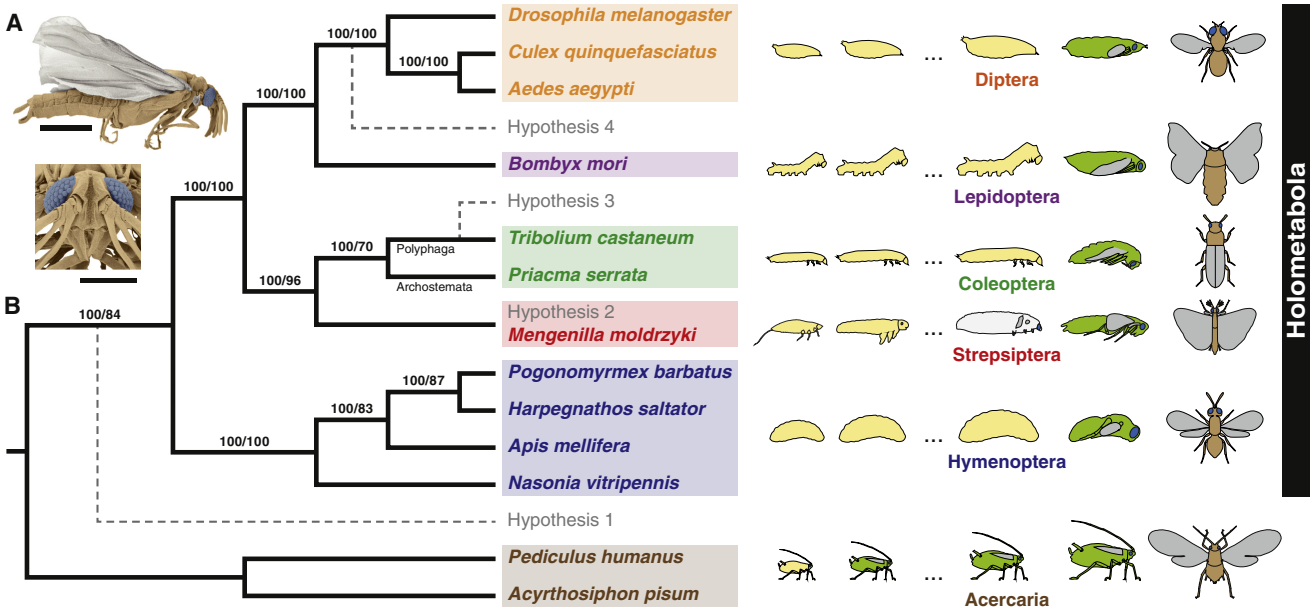


Figure 1. Evolutionary Origin of Twisted-Wing Parasites Inferred from Genomic Evidence

(A) *Mengenilla moldrzyki* male in lateral (top; scale bar represents 1 mm) and frontal (bottom; scale bar represents 500 μm) view (colored SEM micrographs; wings in gray, compound eyes in blue).

(B) Phylogenetic relationships and larval development of holometabolous insects. Numbers above branches are bootstrap support values from analyzing 4,485 protein-coding genes (RY-recoded 2nd codon positions only; ML optimality criterion) and 8,983 near intron pairs (MP optimality criterion). Recent hypotheses on the phylogenetic origin of Strepsiptera are shown in gray. Insect metamorphosis according to Truman and Riddiford [12], with pronymph (yellow) and nymphal stages (green) of insects with direct development (e.g., Acercaria) being equivalent to larval stages (yellow; nymphoid late larval stage of Strepsiptera in white) and pupa (green) of insects with complete metamorphosis (Holometabola); gray, wing buds and wings; blue, compound eyes.

(e.g., Hymenoptera are monophyletic and placed as sister to all remaining Holometabola, Diptera are monophyletic and next to Lepidoptera, and Coleoptera are more closely related to Diptera and Lepidoptera than to Hymenoptera). *M. moldrzyki* is placed as the sister taxon of the flour beetle, *Tribolium castaneum* (Figure 1B; note that the beetle *Priacma serrata* was not included at this step of our investigation). This result implies that Strepsiptera are either the sister group or a highly derived group of Coleoptera.

In addition to the primary sequence-based phylogenetic analyses, we investigated two genomic metacharacter sets as further evidence for the phylogenetic position of Strepsiptera. Specifically, we studied the phylogenetic signal of near intron pairs (NIPs) and that of gene order alignments along the lines with earlier studies that successfully used them to resolve the phylogeny of other holometabolous insects [30] and that of vertebrates [31]. The phylogenetic utility of NIPs is based on the fact that exons smaller than about 50 nucleotides are rare. Hence, introns found in close spatial proximity in orthologous genes of different species are unlikely to have ever coexisted in a single ancestral gene sequence. It is more likely that one intron is lost before the other intron is gained. We identified a total of 8,748 NIPs by studying the gene models of the 4,485 groups of orthologous NEPC genes. Phylogenetic analysis of the NIP characters, of which 1,173 were parsimony informative, under the maximum parsimony (MP) optimality criterion resulted in exactly the same topology as inferred from the primary sequence data (Figure 1B; Figure S2; note that *Priacma serrata* was not included at this step of our investigation).

The second independent approach for phylogenetic reconstruction was based on gene order information. Whereas this approach allows the genome of the species that has to be placed in the tree to be fragmented, all others must be fully assembled at the chromosome level. Accordingly, we used gene orders for *Anopheles gambiae* [32] (replacing *Aedes aegypti* and *Culex quinquefasciatus*), *Apis mellifera*, *Drosophila melanogaster*, and *Tribolium castaneum*, for which at least partial chromosome assemblies exist, and *Nasonia vitripennis*, for which we could exploit linkage map information to map a major fraction of its genome to individual linkage groups. This choice allowed for testing all four aforementioned conflicting phylogenetic scenarios under the assumption that Coleoptera and Diptera are more closely related to each other than to Hymenoptera. Phylogenetic analysis of the spatial arrangement of 791 and 1,433 genes, respectively, depending on whether or not *A. mellifera* with its partial chromosome assembly was part of the analysis, resulted in a topology consistent with those obtained with the previous two methods (Figure 2).

Given the overwhelming support for a close phylogenetic relationship of twisted-wing parasites and beetles, which is also reflected by (1) their high similarity in the protein domain content, (2) the primary sequence information of noncoding RNAs, and (3) the results of other phylogenetic analyses, including those of amino acid sequences (Figures S2 and S4), we next addressed the remaining question of whether or not Strepsiptera are highly derived beetles. For this purpose, we screened contig sequences from an early draft genome of *Priacma serrata* (Archostemata), a representative of an early-divergent lineage of beetles that is the sister group of all remaining extant Coleoptera [33, 34]. We identified the sequences of 3,018 of the 4,485 studied orthologous genes in the *P. serrata* draft genome and aligned them to the

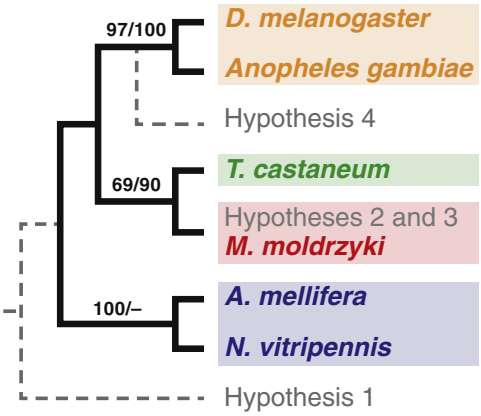


Figure 2. Phylogenetic Relationships of Strepsiptera to other Holometabolous Insects Inferred from Gene Order Distances

Numbers above branches are bootstrap support values from estimating distances with and without *Apis mellifera* (791/1,433 genes). Recent hypotheses on the phylogenetic origin of Strepsiptera are shown in gray. Abbreviations: *A*, *Apis*; *D*, *Drosophila*; *N*, *Nasonia*; *M*, *Mengenilla*; *T*, *Tribolium*.

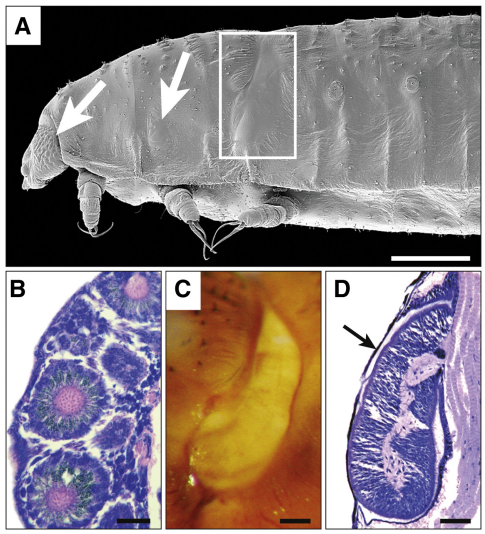


Figure 3. Peculiar Larval Developmental Features of Strepsiptera Reminiscent of Those of Hemimetabolous Insects

(A) SEM micrograph of late larval male of *Mengenilla chobauti*. Left arrow points to left compound eye; right arrow points to bud of left forewing on mesothorax; white rectangle defines sector with bud of left hindwing on metathorax; scale bar represents 500 μm.

(B) Cross-section through compound eye of late larval male of *Eoxenos laboulbenei* (Strepsiptera: Mengenillidae). Tissue was stained with basic fuchsin and methylene blue; scale bar represents 20 μm.

(C) Light microscopic image of left wing bud on metathorax of late larval male of *M. chobauti*; scale bar represents 100 μm.

(D) Cross-section through left wing bud on metathorax of late larval male of *E. laboulbenei*. Tissue stained with basic fuchsin and methylene blue; arrow points to the cuticula from a preceding larval stage; scale bar represents 100 μm.

sequences of the corresponding orthologs from the aforementioned insect species. The new matrix of RY-recoded second codon positions consisted of approximately 1.7 million characters (Table S9). We then repeated the sequence-based phylogenetic analysis, this time including the data from *P. serrata*. The inferred relationships of holometabolous insects were identical with the previously obtained ones, and all branches of the phylogenetic tree again received maximal statistical support (Figure 1B; Figure S2). The sequence data overwhelmingly support a sister group relationship between the archostematan beetle (*P. serrata*) and the polyphagan beetle (*T. castaneum*), indicating that Coleoptera represent a monophyletic group that does not include Strepsiptera. The analysis of NIPs provided additional and independent support for Strepsiptera being the sister group of beetles (Figure 1B; Figure S2).

The first sequenced genome of a twisted-wing parasite allowed the critical evaluation of current hypotheses on the phylogenetic origin of the enigmatic insect order Strepsiptera and provided strong support for Strepsiptera as the closest living relatives of beetles. Although our taxon sampling did not include Neuropterida (alderflies, dobsonflies, snakeflies, ant lions, and relatives), a close phylogenetic relationship between Neuropterida and Strepsiptera appears unlikely from a morphological point of view and would, among other unlikely events, require the independent evolution of posteromotorism, flight with the hindwings only, and a pupa with immobile mandibles (pupa adectica) in Coleoptera and Strepsiptera [2]. A sister group relationship of Strepsiptera and Coleoptera, which is in accordance with morphological evidence [2] and results of some molecular analyses [8, 10, 35], implies that the appearance of compound eyes and the presence of wing buds in late larval Strepsiptera are due to convergence instead of representing ancestral hemimetabolous developmental traits (Figure 1B). This shows that the sequence of holometabolous development, with late instar larvae exhibiting wing imaginal discs and only the pupal stage featuring visible wing buds, is not immutable. The striking similarity of the wing buds and complex eyes of the Strepsiptera late instar larvae (Figure 3) to those of hemimetabolous insect nymphs suggests the reuse of a pre-existing developmental program (homoiology), possibly triggered by a simple change of developmental timing (heterochrony). Our analyses demonstrate that the development of wing imaginal discs and the absence of compound eyes in larval stages are ground plan features of the extremely successful Holometabola and that Strepsiptera are consequently not the "missing link" between hemi- and holometabolous insects.

## Experimental Procedures

### Genome Sequencing and Assembly

The genome of *Mengenilla moldrzyki* was sequenced using a GS XLR 70 (Titanium) sequencer (Roche, Indianapolis, IN, USA) and tissue samples collected at the type locality (Tunisia, Parc Nationale du Jebil, N 32°58′40″/E 009°02′33″). Five PicoTiterPlates were dedicated to an unpaired shotgun (fragmented) library with genomic DNA from a single male. Two PicoTiterPlates were dedicated to a 3 kb mate-pair library with genomic DNA from 14 males. A normalized complementary DNA library from seven adult males was sequenced on an additional PicoTiterPlate. The genome and the complementary transcriptome data were assembled with Newbler 2.3 (Roche). The coverage of the sequenced *M. moldrzyki* genome was estimated with the *l*-mer approach implemented in the software GSP 1.06 (http://gsizepred.sourceforge.net). Genome sequences of *Priacma serrata* were obtained using an Illumina Hi-Seq 2000 sequencer (San Diego, CA, USA) to sequence two paired-end fragment libraries with 500 bp inserts using DNA from two adult males collected in Montana (Gallatin National Forest, N 45°35′27″/W 111°01′30″). The obtained sequence reads were assembled with CLCbio's Genomics Workbench 4.7.1 (Cambridge, MA, USA). Sequence data of the genome shotgun projects have been deposited in the Dryad data repository (http://datadryad.org/;doi:10.5061/dryad.ts058) and at DDBJ/EMBL/GenBank under the accession numbers AGDA00000000 and AGRH00000000.

### Gene Annotation and Orthology

We used MAKER 2.02 with the ab initio gene prediction programs Augustus 2.4, GeneMark-ES 2.3a, and SNAP 2010-07-28 to infer models of NEPC genes [36–40]. We provided MAKER transcript sequences of *M. moldrzyki* and those of other Strepsiptera species downloaded from GenBank (release 179.0; October 5, 2010) and amino acid sequences downloaded from the UniprotKB and TrEMBL protein databases (October 5, 2010) as extrinsic evidence. Mitochondrial genes were annotated with MITOS (http://mitos.bioinf.uni-leipzig.de). Protein domains of NEPC genes were annotated with Pfam_scan.pl 1.3 and HMMER 3.0 and domains from the Pfam database version 24 [41, 42]. DNA methylation-related proteins were searched for and annotated with BLAST 2.2.24+ using amino acid sequences of corresponding proteins in *Apis mellifera* from RefSeq version 48 as query [43]. Noncoding RNAs were annotated with transfer RNA (tRNA)scan-SE 1.21 (tRNA genes and tRNA pseudogenes), RNAmmer 1.2 (18/28S and 5S ribosomal RNA), BLAST 2.2.8 (ncRNAs in general), rfam_scan.pl 1.0 and Infernal 1.02 (snoRNAs), and GotohScan 2.0 (microRNAs), using sequence data from the Rfam database 10.0, GenBank, and miRBase 16.0 [44–49]. Orthology of NEPC genes among species with annotated genome was assessed with OrthoMCL 2.0 [29]. Orthologous NEPC genes in the early draft genome of *Priacma serrata* were identified by reciprocal search using BLAST 2.2.20 and amino acid sequences of NEPC genes from *Tribolium castaneum* and *M. moldrzyki* as queries [46]. The annotated mitochondrial genome of *M. moldrzyki* has been deposited at DDBJ/EMBL/GenBank under the accession number JQ398619. All other annotations are available from the Dryad data repository (http://datadryad.org/;doi:10.5061/dryad.ts058).

### Phylogenetic Analyses

Orthologous amino acid sequences were aligned with MAFFT 6.833b (L-INS-i algorithm), and the resulting alignments were refined with MUSCLE 3.7 [50, 51]. The amino acid alignments were used as blueprints to align the corresponding coding sequences using PAL2NAL 13 [52]. To improve the signal-to-noise ratio in the amino acid alignments of orthologous genes, we used ALISCORE 2.0 to identify and subsequently remove regions in the alignment, whose amino acid pattern-matches did not differ from a random pattern-match [53]. Substitutional saturation was assessed by calculating the observed distances between sequences and comparing them with corrected distances calculated with MEGA5 (Tamura-Nei substitution model) and using a guide tree inferred under the ML optimality criterion when analyzing the amino acid supermatrix of all 4,485 orthologous NEPC genes with RAxML 7.2.8-ALPHA (LG substitution matrix, empirically estimated amino acid frequencies [+F]; rate heterogeneity among sites modeled with gamma distribution [+$\Gamma$]) [54, 55]. The relative compositional variance (RCV) among sequences was calculated with the formula given by Phillips and Penny [56] and excluding constant sites. The signal-to-noise ratio in the data was assessed by calculating the proportion of internal branch lengths to all branch lengths using the minimum evolution (ME) optimality criterion and measuring the branch lengths in the above guide tree with PHYLIP 3.69 [57]. Partition schemes and substitution model parameters were evaluated with ModelGenerator 0.85 [58]. Matrices of RY-recoded second codon positions were analyzed with RAxML using the GTRGAMMA model and specifying 14 partitions, each uniting genes with a similar purine (R) frequency. The concatenated nucleotide sequence alignment of 13 noncoding RNAs (*bantam*, *mir-124*, *mir-133*, *mir-184*, *mir-190*, *mir-263*, *mir-275*, *mir-277*, *mir-305*, *mir-7*, *mir-9*, *U2*, and *U6atac*) was also analyzed with RAxML 7.2.8-ALPHA under the maximum likelihood (ML) optimality criterion, using a mixed RNA-DNA substitution model (S7D model and GTRGAMMA model for paired and unpaired nucleotides, respectively). Near intron pair (NIP) characters were analyzed under the MP optimality criterion using PAUP* 4.0b10 (heuristic tree search: random stepwise addition of taxa [1,000 replicates] and TBR branch-swapping) [59]. Gene order alignments were studied with the program TIBA using the double cut-and-join (DCJ) model for distance correction [31]. Statistical bootstrap support values were estimated from 1,000 (sequence-based and gene order analyses) and 10,000 (NIP character analysis) replicates. The primary sequence-based data matrices, the NIP character matrices, and the gene order alignments have been deposited in the Dryad data repository (http://datadryad.org/;doi:10.5061/dryad.ts058).

### Accession Numbers

Sequence data of the genome shotgun projects have been deposited in the Dryad data repository (http://datadryad.org/;doi:10.5061/dryad.ts058) and at DDBJ/EMBL/GenBank under the accession numbers AGDA00000000 and AGRH00000000. The annotated mitochondrial genome of *M. moldrzyki* has been deposited at DDBJ/EMBL/GenBank under the accession number JQ398619. All other annotations are available from the Dryad data repository (http://datadryad.org/;doi:10.5061/dryad.ts058).

### Supplemental Information

Supplemental Information includes four figures, nine tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2012.05.018.

### Acknowledgments

We are grateful to Robert C. Edgar, Carson Holt, and Mario Stanke for help installing or using their software. We further acknowledge Berthold Fartmann, Jürgen Gadau, Gerald Nyakatura, Christopher D. Smith, and Xin-Xing Tan for technical advice and Thomas P. Niesel for IT support. We thank J. Wolfgang Wägele for helpful comments on an early draft of the manuscript. Michael Ivie kindly assisted with locating field sites for collecting *Priacma serrata*. Lars Podsiadlowski helped with submission of the mitochondrial data to GenBank. We are grateful for allocated computer time from the Cologne High Efficiency Operating Platform for Sciences (CHEOPS) at the University of Cologne. Sequencing and assembly of *P. serrata* was supported by a grant to D.D.M. from the University of Memphis FedEx Institute of Technology. O.N. acknowledges Judith Korb for providing her research facilities in an early phase of this project.

Received: April 3, 2012
Revised: May 4, 2012
Accepted: May 4, 2012
Published online: June 14, 2012

### References

1. Kristensen, N.P. (1999). Phylogeny of endopterygote insects, the most successful lineage of living organisms. Eur. J. Entomol. *96*, 237–253.
2. Beutel, R.G., Friedrich, F., Hörnschemeyer, T., Pohl, H., Hünefeld, F., Beckman, F., Meier, R., Misof, B., Whiting, M.F., and Vilhemsen, L. (2011). Morphological and molecular evidence converge upon a robust phylogeny of the megadiverse Holometabola. Cladistics *27*, 341–355.
3. Trautwein, M.D., Wiegmann, B.M., Beutel, R., Kjer, K.M., and Yeates, D.K. (2012). Advances in insect phylogeny at the dawn of the postgenomic era. Annu. Rev. Entomol. *57*, 449–468.
4. Pierce, W.D. (1909). A monographic revision of the twisted winged insects comprising the order Strepsiptera Kirby. Bull. U. S. Nat. Mus. *66*, 1–232.
5. Whiting, M.F., and Wheeler, W.C. (1994). Insect homeotic transformation. Nature *368*, 696.
6. Carmean, D., and Crespi, B.J. (1995). Do long branches attract flies? Nature *373*, 666.
7. Proffitt, F. (2005). Parasitology. Twisted parasites from "outer space" perplex biologists. Science *307*, 343.
8. Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M.A., Winterton, S.L., and Yeates, D.K. (2009). Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. BMC Biol. *7*, 34.
9. McKenna, D.D., and Farrell, B.D. (2010). 9-genes reinforce the phylogeny of holometabola and yield alternate views on the phylogenetic placement of Strepsiptera. PLoS ONE *5*, e11887.
10. Longhorn, S.J., Pohl, H.W., and Vogler, A.P. (2010). Ribosomal protein genes of holometabolan insects reject the Halteria, instead revealing a close affinity of Strepsiptera with Coleoptera. Mol. Phylogenet. Evol. *55*, 846–859.
11. McMahon, D.P., Hayward, A., and Kathirithamby, J. (2011). Strepsiptera. Curr. Biol. *21*, R272.
12. Truman, J.W., and Riddiford, L.M. (1999). The origins of insect metamorphosis. Nature *401*, 447–452.
13. Pohl, H., and Beutel, R.G. (2008). The evolution of Strepsiptera (Hexapoda). Zoology (Jena) *111*, 318–338.
14. Pohl, H., Niehuis, O., Gloyna, K., Misof, B., and Beutel, R.G. (2012). A new species of *Mengenilla* (Insecta, Strepsiptera) from Tunisia. ZooKeys. *198*, 79–102.
15. Kristensen, N.P. (1981). Phylogeny of insect orders. Annu. Rev. Entomol. *26*, 135–157.

16. Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., and Wheeler, W.C. (1997). The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. *46*, 1–68.
17. Wheeler, W.C., Whiting, M., Wheeler, Q.D., and Carpenter, J.M. (2001). The phylogeny of the extant hexapod orders. Cladistics *17*, 113–169.
18. Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.
19. Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., et al.; Biology Analysis Group. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science *306*, 1937–1940.
20. Honeybee Genome Sequencing Consortium. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature *443*, 931–949.
21. Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., et al. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science *316*, 1718–1723.
22. Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Beeman, R.W., Brown, S.J., Bucher, G., et al.; Tribolium Genome Sequencing Consortium. (2008). The genome of the model beetle and pest *Tribolium castaneum*. Nature *452*, 949–955.
23. Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., et al. (2010). Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science *330*, 86–88.
24. Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N., Donahue, G., Yang, P., Li, Q., Li, C., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science *329*, 1068–1071.
25. International Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. *8*, e1000313.
26. Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. USA *107*, 12168–12173.
27. Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J., et al.; Nasonia Genome Working Group. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science *327*, 343–348.
28. Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M., Zimin, A., Yandell, M., Holt, C., Hu, H., Abouheif, E., Benton, R., et al. (2011). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proc. Natl. Acad. Sci. USA *108*, 5667–5672.
29. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.
30. Krauss, V., Thümmler, C., Georgi, F., Lehmann, J., Stadler, P.F., and Eisenhardt, C. (2008). Near intron positions are reliable phylogenetic markers: an application to holometabolous insects. Mol. Biol. Evol. *25*, 821–830.
31. Lin, Y., Rajan, V., and Moret, B.M.E. (2011). Bootstrapping phylogenies inferred from rearrangement data. In Proceedings of the 11th Workshop on Algorithms in Bioinformatics WABI'11, Lecture Notes in Computer Science, Vol. 6833, Przytycka, T.M., and Sagot, M.-F., eds. (Berlin, Heidelberg: Springer), pp. 175–187.
32. Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. Science *298*, 129–149.
33. Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O.S., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., et al. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science *318*, 1913–1916.
34. Lawrence, J.F., Ślipiński, A., Seago, A.E., Thayer, M.K., Newton, A.F., and Marvaldi, A.E. (2011). Phylogeny of the Coleoptera based on morphological characters of adults and larvae. Ann. Zool. *61*, 1–217.
35. Ishiwata, K., Sasaki, G., Ogawa, J., Miyata, T., and Su, Z.-H. (2011). Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. Mol. Phylogenet. Evol. *58*, 169–180.
36. Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. *18*, 188–196.
37. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19*, ii215–ii225.
38. Lukashin, A.V., and Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. *26*, 1107–1115.
39. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. *33*, 6494–6506.
40. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59.
41. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763.
42. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40* (Database issue), D290–D301.
43. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
44. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.
45. Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.
46. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
47. Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., and Bateman, A. (2011). Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res. *39*, D141–D145.
48. Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335–1337.
49. Hertel, J., de Jong, D., Marz, M., Rose, D., Tafer, H., Tanzer, A., Schierwater, B., and Stadler, P.F. (2009). Non-coding RNA annotation of the genome of *Trichoplax adhaerens*. Nucleic Acids Res. *37*, 1602–1615.
50. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.
51. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.
52. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612.
53. Misof, B., and Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst. Biol. *58*, 21–34.
54. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.
55. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.
56. Phillips, M.J., and Penny, D. (2003). The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. *28*, 171–185.
57. Felsenstein, J. (1989). PHYLIP – phylogeny inference package (version 3.2). Cladistics *5*, 164–166.
58. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and Mclnerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. BMC Evol. Biol. *6*, 29.
59. Swofford, D.L. (2003). PAUP*. Phylogenetic Analysis using Parsimony (* and other methods) (Sunderland, MA: Sinauer).