Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects

Joël Savard,[1] Diethard Tautz,[1] Stephen Richards,[2] George M. Weinstock,[2] Richard A. Gibbs,[2] John H. Werren,[3] Hervé Tettelin,[4] and Martin J. Lercher[5,6,7]

[1]*Abteilung für Evolutionsgenetik, Institut für Genetik, Universität zu Köln, Köln 50674, Germany;* [2]*Human Genome Sequencing Centre, Baylor College of Medicine, Houston, Texas 77002, USA;* [3]*Department of Biology, University of Rochester, New York 14627, USA;* [4]*The Institute for Genomic Research, Rockville, Maryland 20850, USA;* [5]*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom;* [6]*European Molecular Biology Laboratory, 69012 Heidelberg, Germany*

Comparative studies require knowledge of the evolutionary relationships between taxa. However, neither morphological nor paleontological data have been able to unequivocally resolve the major groups of holometabolous insects so far. Here, we utilize emerging genome projects to assemble and analyze a data set of 185 nuclear genes, resulting in a fully resolved phylogeny of the major insect model species. Contrary to the most widely accepted phylogenetic hypothesis, bees and wasps (Hymenoptera) are basal to the other major holometabolous orders, beetles (Coleoptera), moths (Lepidoptera), and flies (Diptera). We validate our results by meticulous examination of potential confounding factors. Phylogenomic approaches are thus able to resolve long-standing questions about the phylogeny of insects.

[Supplemental material is available online at www.genome.org.]

The four major orders of holometabolous insects (Hymenoptera, Coleoptera, Lepidoptera, and Diptera) encompass over 45% of all known animal species (Hammond 1992). While analyses based on morphological (Kristensen 1999) or individual molecular markers (such as ribosomal RNA; Whiting 2002b) or mitochondrial DNA sequences (Castro and Dowton 2005) have confirmed the monophyly of these orders, they have been unable to elucidate most of the interordinal relationships with sufficient confidence.

A close relationship between Diptera (flies) and Lepidoptera (moths) within the long-recognized Mecopterida assemblage is generally recovered. However, the affinities of Coleoptera (beetles) and more particularly of Hymenoptera (wasps and bees) (Castro and Dowton 2005) remain elusive. In the most widely accepted phylogenetic hypothesis (Kristensen 1999; Whiting 2002b), a preference is given to a sister-group relationship between Hymenoptera and Mecopterida, while Coleoptera are placed at a more basal position as a sister group to the Neuropterida, another long-recognized assemblage.

To resolve the phylogenetic relationships of the major holometabolous orders, we adopt a phylogenomic approach, utilizing a large number of nuclear genes to maximize phylogenetic signal over noise (Eisen and Fraser 2003; Rokas et al. 2003; Delsuc et al. 2005; DeSalle 2005; Philippe et al. 2005a). Such approaches, based on the simultaneous analysis of a large number of nuclear genes, have already been shown to be a promising route to understand deep metazoan relationships (Dopazo and Dopazo 2005; Philippe et al. 2005b). Here, we demonstrate that these methods are also able to resolve long-standing questions about the phylogeny of insects.

Using EST sequences to obtain phylogenomic data sets has proven fruitful, e.g., in the analysis of Eukaryota (Philippe et al. 2004), Amoebae (Bapteste et al. 2002), and Coleoptera relationships (Hughes et al. 2006). The use of EST sequences in phylogenomic studies of insects was suggested earlier (Theodorides et al. 2002), but sufficient data to answer the questions addressed here has only recently become available.

Our analysis focuses on six holometabolous model species, for which large scale sequencing projects are available or in progress. These encompass two dipterans (the fruit fly *Drosophila melanogaster* and the mosquito *Anopheles gambiae*), one lepidopteran (the silk moth *Bombyx mori*), one coleopteran (the flour beetle *Tribolium castaneum*), and two hymenopterans (the honey bee *Apis mellifera*, and the sibling parasitic wasp species *Nasonia vitripennis* and *Nasonia giraulti*). We further include one orthopteran (the grasshopper *Locusta migratoria*) and one hemipteran (the pea aphid *Acyrthosiphon pisum*), both of which are uncontested outgroups to the holometabolous insects based on morphological and molecular markers (Boudreaux 1979; Hennig 1981; Kristensen 1991; Wheeler et al. 2001).

Results

Candidate orthologous clusters were assembled from known or predicted genes based on a stringent sequence similarity criterion, and were then manually curated to ensure orthology (see Methods). After removing ambiguously aligned regions, we assembled the remaining sequences into a concatenated alignment of 33,809 amino acid positions from 185 nuclear genes. As expected, most genes included here perform housekeeping functions (see Table S1 of the Supplemental information for a list of genes).

[7]**Corresponding author.**
E-mail M.J.Lercher@bath.ac.uk; fax 44-1225-386779.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.5204306. Freely available online through the *Genome Research* Open Access option.

16:1334–1338 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org

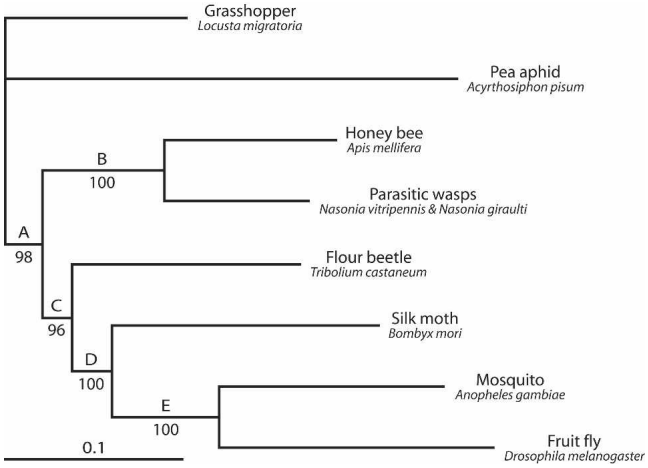

Figure 1. Holometabolous phylogenetic relationships, showing the monophyly of Coleoptera (beetles), Lepidoptera (moths), and Diptera (flies), to the exclusion of Hymenoptera (bees and wasps). Branch lengths are from maximum likelihood. Numbers report maximum likelihood bootstrap support (in percent); Bayesian posterior probabilities and maximum parsimony bootstrap support are ≥99% for each branch (Table 1).

This data set supported the topology in Figure 1 regardless of the phylogenetic methodology (maximum likelihood [Guindon and Gascuel 2003], Bayesian [Yang and Rannala 1997], or maximum parsimony [Felsenstein 2004]), with nearly 100% bootstrap support or 100% posterior probabilities in each case (Table 1). The previously recognized close relationship of Diptera and Lepidoptera is recovered and thus substantiated. However, Hymenoptera and not Coleoptera is the most basal of the four major holometabolous orders. To ensure that our results were not influenced by an unusually evolving subset of sequences (Gadagkar et al. 2005), we utilized a bootstrap strategy based on the resampling of genes (Nei et al. 2001), again resulting in strong support for the topology in Figure 1 (Table 1).

Two common sources of error in phylogenetic reconstructions are compositional biases (Foster and Hickey 1999) and long-branch attraction (Felsenstein 1978). Two of our species, the pea aphid *Acyrthosiphon pisum* and the honey bee *Apis mellifera*, have a strongly AT-biased genome. This is reflected in an overrepresentation of amino acids encoded by AT-rich codons (Table 2), confirmed by statistical analysis (pairwise χ^2-tests, Table 3). The removal of the outlier species *Acyrthosiphon pisum* and *Apis mellifera* from the data set resulted in the same well supported topology (Table 1).

Because long branches in the ingroup are restricted to Diptera and Lepidoptera, whose relative positions are uncontested (Kristensen 1999; Whiting 2002b), long-branch attraction is unlikely to have influenced the topology. However, even though the branches of hymenopterans and *Tribolium* are short, it is still conceivable that genes with particular substitution rates in these species may have biased the phylogeny. Exclusion of such genes on the basis of a relative rate test (Tajima 1993) did not change the tree (Table 1).

Improper outgroup choice can potentially influence the inferred rooting of the ingroup. While our two outgroup species differ profoundly in evolutionary rate and amino acid composition, the results remained unchanged when using either *Locusta migratoria* or *Acyrthosiphon pisum* individually as the outgroup (Table 1).

Discussion

The interordinal relationships among holometabolous insect orders had previously proven to be notoriously difficult to resolve. However, most researchers assumed a basal split between two super-orders, the Coleoptera–Neuropterida (including the beetles) and the Hymenoptera–Mecopterida (including wasps, flies, and moths) (Kristensen 1999; Whiting 2002b). The tree presented in Figure 1 necessitates a re-evaluation of this consensual view. The Mecopterida, encompassing Diptera (flies) and Lepidoptera (moths), seem now more closely related to Coleoptera (beetles) than to Hymenoptera (wasps and bees).

In the present framework, the position of Neuropterida could not be assessed. Neither previous molecular phylogenies nor morphological characters allow settlement of this issue; in particular, wing structure features have been argued to support a sister-group relationship of Neuropterida with either Coleoptera (Hornschemeyer 2002) or with Mecopterida (Kukalová-Peck and Lawrence 2004). The morphological characters used to support the traditional Holometabola phylogeny should certainly be reanalyzed in the light of the relationships presented here.

Why was the basal position of hymenopterans not discovered in previous molecular phylogenetic studies? A plausible explanation is the lack of resolution power of single molecules when radiations are old or compressed in time (Rokas et al. 2005). Because the phylogenetic split in question occurred at least 275 million years ago (Mya) (Ponomarenko 2002; Rasnitsyn 2002), analyses based on a single molecule (e.g., 18S rRNA) did not provide sufficient resolution (Whiting 2002a). While 60% of the 185 protein alignments analyzed here were better explained by the tree in Figure 1 than by the previously assumed tree (based on likelihood comparisons), only two proteins supported the basal position of Hymenoptera with a bootstrap support >50% when analyzed individually (removing these proteins does not

Table 1. Support values for the nodes in Figure 1

	Data set	A	B	C	D	E
Maximum likelihood bootstrap[a]	Full	98	100	96	100	100
	Gene bootstrap	97	100	98	100	100
	Relative rate test[c]	n/a	n/a	97	100	100
	L. migratoria excluded	n/a	100	100	100	100
	A. pisum excluded	n/a	100	99	100	100
	A. pisum and *A. mellifera* excluded	n/a	n/a	100	100	100
	Single gene consensus[d]	18	65	13	19	53
Bayesian posterior probabilities[b]	Full	100	100	100	100	100
Maximum parsimony bootstrap[a]	Full	100	100	99	100	100

[a]Bootstrap fractions estimated from 1000 maximum likelihood or maximum parsimony calculations (in percent).
[b]Posterior probabilities estimated from all sampled trees in four independent Markov chain Monte Carlo calculations (in percent).
[c]After removal of 71 genes that showed significant rate heterogeneity ($P < 0.2$) between *Nasonia* and *Tribolium castaneum*, and excluding *Apis mellifera* from the alignment.
[d]Majority-rule consensus based on 100 bootstrap replicates per gene; very similar results are obtained when only the best tree contributes for each gene.

Table 2. Compositional bias, showing the percentage of amino acids encoded by AT-rich and by GC-rich codons, as well as their ratio

Species	% FYMINK[a]	% GARP[b]	FYMINK/GARP
Acyrthosiphon pisum	28.94	23.04	1.26
Anopheles gambiae	27.73	24.72	1.12
Apis mellifera	29.19	23.55	1.24
Bombyx mori	27.63	24.63	1.12
Drosophila melanogaster	27.83	24.41	1.14
Locusta migratoria	27.86	24.56	1.13
Nasonia vitripennis and *Nasonia giraulti*	28.54	24.00	1.19
Tribolium castaneum	28.41	24.15	1.18

[a]AT-rich amino acids: F (Phe), Y (Tyr), M (Met), I (Ile), N (Asn), and K (Lys).
[b]GC-rich amino acids: G (Gly), A (Ala), R (Arg), and P (Pro).

influence the tree topology, data not shown). Accordingly, a consensus tree based on single gene tree reconstructions yielded the same topology as in Figure 1, but without strong bootstrap support (Table 1). The present analysis thus supports the notion that concatenated sequence trees provide more resolution than consensus gene trees (Rokas et al. 2003; Gadagkar et al. 2005): Combined analysis of a large number of sequences was necessary to resolve the deep evolutionary relationships among holometabolous orders.

Previous studies based on the simultaneous analysis of many proteins also failed to recover the topology in Figure 1. Philippe et al. (2004) analyzed 129 proteins in a maximum likelihood framework, resulting in a strongly supported sister-group relationship between *Apis mellifera* and *Bombyx mori*, with *Drosophila melanogaster* located more basal. This topology is inconsistent with the monophyly of Mecopterida, which is well supported by morphological evidence (Kristensen 1999). It is most likely a long-branch attraction artifact, brought about by the insufficient number of holometabolous species and the use of a very divergent lineage (a tick) as the outgroup (Philippe and Laurent 1998). This study highlights the importance of selecting close outgroup species, and the necessity to test the influence of an alternative outgroup choice. A second study (Philippe et al. 2005b) recently analyzed 146 proteins, failing to resolve the relative positions of Coleoptera and Hymenoptera. Here, the potentially large amount of missing data between the insect species considered might be responsible for the lack of resolution.

The interordinal relationships presented in Figure 1 are in fact also suggested by examination of the fossil record (Rohdendorf and Rasnitsyn 1980), and are supported by phylogenetic analysis of intron positions (Krauss et al. 2005) as well as wing characters (Ross 1965; Kukalová-Peck and Lawrence 2004). Furthermore, we note that phylogenies based on 18S rRNA sequences also yielded a basal Hymenoptera among holometabolous insects, although not with a credible level of support (Whiting 2002b). The sum of evidence hence supports the present phylogenomic analysis as a reliable foundation for comparative analyses of the insect model organisms.

Methods

Sequence data

Drosophila melanogaster (Adams et al. 2000), *Anopheles gambiae* (Holt et al. 2002), and *Apis mellifera* peptides were obtained from Ensembl (www.ensembl.org). *Bombyx mori* (Mita et al. 2003), *Locusta migratoria* (Kang et al. 2004), *Tribolium castaneum*, and *Acyrthosiphon pisum* mRNA sequences were downloaded from NCBI (ftp.ncbi.nlm.nih.gov). *Nasonia vitripennis* and *Nasonia giraulti* EST data were generated by authors J.H.W. and H.T. All nucleotide data sets were cleaned of vector, mitochondrial and bacterial contaminations using SeqClean (available from www.tigr.org/tdb/tgi/software/) before being assembled into nonredundant contigs with cap3 using default settings (Huang and Madan 1999). All nucleotide data sets were then searched against all *Drosophila melanogaster* proteins using BLASTx. The reading frame from the best hit was assumed to be the correct reading frame. We then chose the longest run of peptides uninterrupted by a stop codon as the peptide corresponding to each nucleotide contig.

Identification of orthologs

We performed BLASTp searches of all proteome pairs. Orthologs were selected based on reciprocal best BLAST hits (Tatusov et al. 1997) using an E-value cut-off of 10^{-25}. A group of sequences with exactly one member in each species (including either one or both *Nasoni*a species) was accepted as a candidate orthologous family if each sequence had each of the other family sequences as the best BLASTp hit in the respective proteome. This requirement of all-against-all reciprocal best hits is very stringent, and thus gives good confidence in the inferred orthology. Multiple sequence alignments were performed with MUSCLE (Edgar 2004) using default settings. Resulting alignments were then manually curated to ensure completeness and consistency. Poorly conserved families or clusters potentially containing paralogous sequences were discarded. For the sibling *Nasonia* species, when orthologous sequences were available for both species, the longer one was chosen. Alignments were then purged from unreliably aligned positions as well as gaps with Gblocks (Castresana 2000) using highly stringent settings, where all sequences flanking a

Table 3. χ^2 (upper triangle) and Bonferroni-corrected *P*-values (lower triangle) showing that *Acyrthosiphon pisum* and *Apis mellifera* have a significant compositional bias compared to most other species

	Anopheles gambiae	*Apis Mellifera*	*Acyrthosiphon pisum*	*Bombyx mori*	*Drosophila melanogaster*	*Locusta migratoria*	*Nasonia vitripennis* and *N. giraulti*	*Tribolium castaneum*
A. gambiae	—	43.84	39.35	2.23	9.23	5.52	16.97	15.13
A. pisum	0.000042	—	10.42	34.85	36.12	30.55	13.97	20.12
A. mellifera	0.00027	1	—	32.76	29.87	28.71	10.88	12.42
B. mori	1	0.0018	0.0041	—	5.61	2.24	11.46	9.59
D. melanogaster	1	0.0011	0.013	1	—	9.08	9.95	8.49
L. migratoria	1	0.0099	0.020	1	1	—	9.14	7.32
N. vitripennis and *N. giraulti*	1	1	1	1	1	1	—	2.02
T. castaneum	1	0.48	1	1	1	1	1	—

block had to be conserved and where blocks smaller than 20 amino acids were discarded. We concatenated the final set of 185 nuclear sequences for phylogenetic analysis, resulting in an eight-species alignment of 33,809 amino acid positions. The list of genes included in our analysis is available as Supplemental-Table S1.

Phylogenetic reconstructions

We first analyzed the data in a maximum likelihood framework, using phyML (Guindon and Gascuel 2003) under an empirical model of amino acid substitutions (Jones et al. 1992), allowing for substitution rate variation among sites with a gamma distribution (four rate categories). Branch support values in Table 1 are from analysis of 1000 bootstrap replicates. Using alternative models of amino acid evolution (WAG, Whelan and Goldman 2001; and VT, Bapteste et al. 2002) led to the same well supported tree.

Additionally, we estimated the tree in a Bayesian framework, using MrBayes (Huelsenbeck and Ronquist 2001) and employing the same model of sequence evolution as above. We ran four independent searches, each starting from a random tree and sampling every tenth tree over 100,000 generations. Each run had equilibrated after less than 1000 generations; thus, the first 100 trees were disregarded as burn-in. The independent runs consistently resulted in the same topology; posterior probabilities (Table 1) were calculated from all sampled trees across independent runs. Finally, we also analyzed 1000 bootstrap replicates under a maximum parsimony criterion, using PROTPARS from the PHYLIP package (Felsenstein 2004).

Gene bootstrap resampling

To test if the obtained tree was dominated by one or a few disparate genes, we performed maximum likelihood analyses of 1000 bootstrap data sets obtained from the resampling of complete genes (Nei et al. 2001). For each replicate, we drew 185 gene alignments from the full curated data set described above. Because genes were not removed from the pool after being chosen, each bootstrap data set contained some gene alignments more than once, while others were missing altogether; this is analogous to the widely used bootstrap strategy based on individual amino acid sites. For each replicate, alignments were then concatenated and analyzed with phyML as above.

Relative rate tests

To determine if the relative position of Coleoptera and Hymenoptera was caused by rate variation among these orders, we performed a relative rate test (Tajima 1993) between *Nasonia* and *Tribolium castaneum*. We restricted this analysis and the following tree reconstruction to sites that were identical between *Locusta migratoria* and *Acyrthosiphon pisum*, and where the ancestral state could thus be inferred reliably. To be conservative, we removed all genes for which $\chi^2 > 1.64$, i.e., $P < 0.2$ (1 degree of freedom) (Kumar and Hedges 1998). Excluding *Apis mellifera*, we concatenated the remaining 114 genes into an alignment of 16,495 amino acid positions. Maximum likelihood bootstrap analysis was performed as above.

Phylogenetic analysis of individual proteins

We also analyzed each individual protein alignment with the maximum likelihood method as described above. To analyze each protein's support for the tree in Figure 1 compared to the previously assumed tree (with the positions of hymenopterans and *Tribolium* exchanged), we compared the likelihoods calculated under each topology; statistical support was estimated using the method of Kishino and Hasegawa (1989) as implemented in the PAML package (Yang 1997).

Compositional heterogeneity among species pairs was assessed with a χ^2-test, where $\chi_{mn}^2 = \Sigma_i (f_{mi} - f_{ni})^2/(f_{mi} + f_{ni})$, with f_{mi} the total number of amino acids of type i in the concatenated sequence for species m. The values in Table 3 are based on those amino acids that are biased in the GC content of their codons (FYMINK/GARP, see also Table 2; 9 degrees of freedom). Qualitatively very similar results are obtained when using all amino acids (data not shown).

All phylogenomic analyses and tests were implemented in Perl scripts, which are available in the Supplemental material.

Acknowledgments

J.H.W. thanks Wayne Hunter (USDA, ARS) and Phat Dang (USDA, ARS) for construction of the Nasonia EST libraries. We wish to thank Shannon K. McWeeney and Csaba Pal for helpful discussions. J.S. and D.T. acknowledge support through grants from the HFSPO and the DFG. M.J.L. acknowledges financial support from the Royal Society and the DFG. The Nasonia EST project of J.H.W. and H.T. was supported by the 21st Century Research & Technology Fund.

References

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.

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., et al. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci.* **99:** 1414–1419.

Boudreaux, H.B. 1979. *Arthropod phylogeny with special reference to insects*. Wiley, New York.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17:** 540–552.

Castro, L.R. and Dowton, M. 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Mol. Phylogenet. Evol.* **34:** 469–479.

Delsuc, F., Brinkmann, H., and Philippe, H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6:** 361–375.

DeSalle, R. 2005. Animal phylogenomics: Multiple interspecific genome comparisons. *Methods Enzymol.* **395:** 104–133.

Dopazo, H. and Dopazo, J. 2005. Genome-scale evidence of the nematode-arthropod clade. *Genome Biol.* **6:** R41.

Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:** 1792–1797.

Eisen, J.A. and Fraser, C.M. 2003. Phylogenomics: Intersection of evolution and genomics. *Science* **300:** 1706–1707.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27:** 401–410.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA.

Foster, P.G. and Hickey, D.A. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48:** 284–290.

Gadagkar, S.R., Rosenberg, M.S., and Kumar, S. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *J. Exp. Zool. (B Mol. Dev. Evol.)* **304:** 64–74.

Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52:** 696–704.

Hammond, P. 1992. Species inventory. In *Global biodiversity* (ed. B. Groombridge), pp. 17–39. Chapman & Hall, London.

Hennig, W. 1981. *Insect phylogeny*. Wiley, New York.

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.
Hornschemeyer, T. 2002. Phylogenetic significance of the wing-base of the Holometabola (Insecta). *Zool. Scr.* **31:** 17–29.
Huang, X. and Madan, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* **9:** 868–877.
Huelsenbeck, J.P. and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.
Hughes, J., Longhorn, S.J., Papadopoulou, A., Theodorides, K., de Riva, A., Mejia-Chang, M., Foster, P.G., and Vogler, A.P. 2006. Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (beetles). *Mol. Biol. Evol.* **23:** 268–278.
Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8:** 275–282.
Kang, L., Chen, X., Zhou, Y., Liu, B., Zheng, W., Li, R., Wang, J., and Yu, J. 2004. The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. *Proc. Natl. Acad. Sci.* **101:** 17611–17615.
Kishino, H. and Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* **29:** 170–179.
Krauss, V., Pecyna, M., Kurz, K., and Sass, H. 2005. Phylogenetic mapping of intron positions: A case study of translation initiation factor eIF2γ. *Mol. Biol. Evol.* **22:** 74–84.
Kristensen, N.P. 1991. The phylogeny of hexapods. In *The insects of Australia: A Textbook for students and research workers* (eds. I.D. Naumann et al.), pp. 125–140. Melbourne University Press, Carlton, Australia.
Kristensen, N.P. 1999. Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur. J. Entomol.* **96:** 237–253.
Kukalová-Peck, J. and Lawrence, J.F. 2004. Relationships among coleopteran suborders and major endoneopteran lineages: Evidence from hind wing characters. *Eur. J. Entomol.* **101:** 95–144.
Kumar, S. and Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. *Nature* **392:** 917–920.
Mita, K., Morimyo, M., Okano, K., Koike, Y., Nohata, J., Kawasaki, H., Kadono-Okuda, K., Yamamoto, K., Suzuki, M.G., Shimada, T., et al. 2003. The construction of an EST database for *Bombyx mori* and its application. *Proc. Natl. Acad. Sci.* **100:** 14121–14126.
Nei, M., Xu, P., and Glazko, G. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci.* **98:** 2497–2502.
Philippe, H. and Laurent, J. 1998. How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* **8:** 616–623.
Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Mol. Biol. Evol.* **21:** 1740–1752.
Philippe, H., Delsuc, F., Brinkmann, H., and Lartillot, N. 2005a. Phylogenomics. *Annu. Rev. Ecol. Evol. Syst.* **36:** 541–562.
Philippe, H., Lartillot, N., and Brinkmann, H. 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22:** 1246–1253.
Ponomarenko, A.G. 2002. Superorder Scarabaeidea Laicharting, 1781. Order Coleoptera Linné, 1758. The Beetles. In *History of insects* (eds. A.P. Rasnitsyn and D.L.J. Quicke), pp. 164–176. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Rasnitsyn, A.P. 2002. Cohors Scarabaeiformes Laicharting, 1781. The Holometabolans. In *History of insects* (eds. A.P. Rasnitsyn and D.L.J. Quicke), pp. 157–159. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Rohdendorf, B.B. and Rasnitsyn, A.P. 1980. *Historical development of the class insecta*. Nauka Press, Moscow.
Rokas, A., Williams, B.L., King, N., and Carroll, S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425:** 798–804.
Rokas, A., Kruger, D., and Carroll, S.B. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310:** 1933–1938.
Ross, H.H. 1965. *A textbook of entomology*. Wiley, New York.
Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135:** 599–607.
Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278:** 631–637.
Theodorides, K., De Riva, A., Gomez-Zurita, J., Foster, P.G., and Vogler, A.P. 2002. Comparison of EST libraries from seven beetle species: Towards a framework for phylogenomics of the Coleoptera. *Insect Mol. Biol.* **11:** 467–475.
Wheeler, W.C., Whiting, M., Wheeler, Q.D., and Carpenter, J.M. 2001. The phylogeny of the extant hexapod orders. *Cladistics* **17:** 113–169.
Whelan, S. and Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18:** 691–699.
Whiting, M.F. 2002a. Phylogeny of the holometabolous insect orders based on 18S ribosomal data: When bad things happen to good data. In *Molecular systematics and evolution:Theory and practice* (ed. G.G.R. DeSalle, and W.C. Wheeler), pp. 69–83. Birkhäuser Verlag AG, Basel, Switzerland.
Whiting, M.F. 2002b. Phylogeny of the holometabolous insect orders: Molecular evidence. *Zool. Scr.* **31:** 3–15.
Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13:** 555–556.
Yang, Z.H. and Rannala, B. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo method. *Mol. Biol. Evol.* **14:** 717–724.

Received November 3, 2005; accepted in revised form April 3, 2006.