Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular Phylogenetics and Evolution 38 (2006) 100–116
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution

Alexandre Hassanin *

Muséum National d'Histoire Naturelle, Département Systématique et Evolution, UMR 5202—Origine, Structure, et Evolution de la Biodiversité, Case postale No. 51, 55, rue Buffon, 75005 Paris, France

Received 17 March 2005; revised 8 August 2005; accepted for publication 6 September 2005
Available online 14 November 2005

Abstract

In this study, mitochondrial sequences were used to investigate the relationships among the major lineages of Arthropoda. The data matrix used for the analyses includes 84 taxa and 3918 nucleotides representing six mitochondrial protein-coding genes (*atp6* and *8*, *cox1–3*, and *nad2*). The analyses of nucleotide composition show that a reverse strand-bias, i.e., characterized by an excess of T relative to A nucleotides and of G relative to C nucleotides, was independently acquired in six different lineages of Arthropoda: (1) the honeybee mite (*Varroa*), (2) Opisthothelae spiders (*Argiope*, *Habronattus*, and *Ornithoctonus*), (3) scorpions (*Euscorpius* and *Mesobuthus*), (4) *Hutchinsoniella* (Cephalocarid), (5) *Tigriopus* (Copepod), and (6) whiteflies (*Aleurodicus* and *Trialeurodes*). Phylogenetic analyses confirm that these convergences in nucleotide composition can be particularly misleading for tree reconstruction, as unrelated taxa with reverse strand-bias tend to group together in MP, ML, and Bayesian analyses. However, the use of a specific model for minimizing effects of the bias, the "Neutral Transition Exclusion" (NTE) model, allows Bayesian analyses to rediscover most of the higher taxa of Arthropoda. Furthermore, the analyses of branch lengths suggest that three main factors explain accelerated rates of substitution: (1) genomic rearrangements, including duplication of the control region and gene translocation, (2) parasitic lifestyle, and (3) small body size. The comparisons of Bayesian Bootstrap percentages show that the support for many nodes increases when taxa with long branches are excluded from the analyses. It is therefore recommended to select taxa and genes of the mitochondrial genome for inferring phylogenetic relationships among arthropod lineages. The phylogenetic analyses support the existence of a major dichotomy within Arthropoda, separating Pancrustacea and Paradoxopoda. Basal relationships between Pancrustacean lineages are not robust, and the question of Hexapod monophyly or polyphyly cannot be answered with the available mitochondrial sequences. Within Paradoxopoda, Chelicerata and Myriapoda are each found to be monophyletic, and *Endeis* (Pycnogonida) is, surprisingly, associated with Acari.

Keywords: mtDNA sequences; Phylogeny; Strand-bias; Long-branch attraction artifact; Parasites; Body size; Arthropoda; Pancrustacea; Paradoxopoda; Chelicerata

1. Introduction

Arthropods are the most abundant and diverse group of animals on Earth, with more than one million described species. Traditionally, the phylum Arthropoda is divided into four extant subphyla: Crustacea (crabs, shrimps, etc.), Hexapoda (insects, diplurans, proturans, and springtails), Myriapoda (centipedes, millipedes, and their kin), and Chelicerata (horseshoe crabs, arachnids, and pycnogonids) (Brusca and Brusca, 2003). The relationships between and within these four major lineages remain one of the most contentious issues in systematics, and many different hypotheses have been proposed in the literature. The traditional morphological hypotheses propose to group Myriapoda either with both Crustacea and Hexapoda into the clade Mandibulata (e.g., Snodgrass, 1938), or with Hexa-

* Fax: +33 1 40 79 30 63.
E-mail address: hassanin@mnhn.fr.

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.09.012

poda alone into the clade Atelocerata (e.g., Cisne, 1974; Kraus, 2001; Snodgrass, 1938).

Sequences of the mitochondrial (mt) genome have been widely used to approach this issue. The analyses have suggested unexpected results, which if true would have huge consequences for the interpretation of morphological characters: (i) Crustacea came out as paraphyletic, with Malacostraca being more closely related to Hexapoda than Branchiopoda (Garcia-Machado et al., 1999; Hwang et al., 2001; Nardi et al., 2001, 2003; Wilson et al., 2000); (ii) one study found Hexapoda to be paraphyletic, with Insecta allied with crustaceans rather than with Collembola (Nardi et al., 2003); (iii) Chelicerata and Myriapoda have each come out as para- or polyphyletic (Delsuc et al., 2003; Nardi et al., 2003; Negrisolo et al., 2004); and (iv) the results of Hwang et al. (2001) suggested that Myriapoda group with Chelicerata.

The usefulness of mtDNA as a marker for highly divergent lineages remains controversial (e.g., Curole and Kocher, 1999). There are three main problems with using mt sequences for the phylogeny of Arthropoda. (1) The first arthropods probably arose in ancient Precambrian seas over 600 million years ago (Brusca and Brusca, 2003). As a consequence, mutational saturation due to multiple hits is a major problem in tree reconstruction, and with mt sequences, saturation is all the more important because the mt genome evolves more rapidly than the nuclear genome (e.g., Burger et al., 2003; Li, 1997). (2) The rates of nucleotide substitution have differed among lineages, and taxa evolving faster can group together because of the long-branch attraction (LBA) phenomenon (Felsenstein, 1978). (3) Another problem that can mislead phylogenetic inferences is that the mt genes of some taxa have been affected by important changes in the pattern of substitution, such as reversals of asymmetric mutational constraints (Hassanin et al., 2005).

In this study, phylogenetic relationships among the major groups of Arthropoda were inferred by using a mtDNA fragment including six of the 13 protein-coding genes (*nad2*, *cox1*, *cox2*, *atp6*, *atp8*, and *cox3*), and a taxon sample of 78 arthropods and six outgroup species. The fragment was chosen because the arrangement of these six genes is conserved for most arthropod species. Five species of chelicerates were sequenced for this study to test the monophyly of Chelicerata, as well as relationships within this group. The sequences were examined to identify which species are characterized by a reverse strand-bias that led to extreme evolutionary divergence. It will be shown that phylogenetic inferences based on mtDNA sequences can be strongly biased by divergent sequences causing long-branch attraction (LBA) artifacts of reconstruction. Here, a specific "NTE" model was applied for limiting the impact of such divergent changes on the pattern of substitution. In a second approach of tree reconstruction, taxa with branch lengths significantly longer than others were excluded, in the hope of improving phylogenetic inferences.

2. Materials and methods

2.1. DNA extraction, amplification, and sequencing

Five species of chelicerates were sequenced for this study, including one pycnogonid (*Endeis spinosa*) and four arachnids: *Argiope bruennichi* (Araneae), *Euscorpius flavicaudis* (Scorpiones), *Mastigoproctus giganteus* (Uropygi), and *Phrynus* sp. (Amblypygi). Genomic DNA was isolated by a CTAB (hexadecyltrimethylammonium bromide) procedure (Winnepenninckx et al., 1993). A mtDNA fragment including six protein-coding genes; i.e., *nad2*, *cox1*, *cox2*, *atp8*, *atp6*, and *cox3*, was obtained by amplifying and sequencing nine overlapping fragments using the primer couples listed in Table 1. PCR products were purified using the Jetsorb kit (Genomed) and cloned into a pKS Bluescript (Stratagene) T-hang modified according to the procedure of Holton and Graham (1990). Sequencing reactions were performed with the "Thermosequenase fluorescent-labelled primer cycle sequencing kit with 7-deaza-dGTP" (Amersham, Pharmacia) using a fluorescent primer labelled with CY5. Reaction products were analyzed using an automatic sequencer (Alf Express, Pharmacia). Sequences are deposited in the EMBL/GenBank/DDBJ databases under the accession numbers specified in Table 2.

2.2. Taxonomic sampling

The taxonomic sampling is composed of 84 species (Table 2). The ingroup is the phylum Arthropoda, represented by 78 species with 31 Insecta, 4 Collembola, 19 Crustacea, 4 Myriapoda, and 15 Chelicerata in addition to the five species of chelicerates sequenced in the present work. The species of arthropods were chosen to include all the diversity available in the nucleotide databases. However, the two highly divergent mtDNA sequences of Hymenoptera (*Apis* NC_001566 and *Melipona* NC_004529) were excluded, to facilitate protein alignments in order to retain more nucleotide sites for the analyses. In addition, five other species were excluded because one or several of the six genes of interest were inverted in their mt genome: *Heterodoxus* due to the inversion of both *atp6* and *atp8* genes; and four insects of the subfamily Aleyrodinae (*Aleurochiton*, *Bemisia*, *Neomaskellia*, and *Tetraleurodes*) due to the inversion of *cox3*. In all outgroup species selected for this study, the six genes of interest have both the same orientation and order as in most arthropod species (order = *nad2–cox1–cox2–atp8–atp6–cox3*). This explains why the outgroup was restricted to six genera belonging to three different Metazoan phyla, i.e., *Bos*, *Myxine*, and *Petromyzon* for Chordata, *Balanoglossus* for Hemichordata, and *Lumbricus* and *Platynereis* for Annelida. Unfortunately, the outgroup selection does not include non-arthropod Ecdysozoans, because the mitochondrial genomes of Nematoda are highly divergent, with many gene rearrangements, and the ones of Onychophora and Tardigrada are not available in the nucleotide databases.

Table 1
Couples of primers used for PCR amplifications

Upper primers	Lower primers (reverse)
U10: 5′-TWAAGCTANTGGGYTCATACCCC-3′	L10: 5′-TTAACTTTGAAGGYTWHTAGTTT-3′ L12: 5′-GATTWGTWGAAWAYAATCATCGC-3′
U4: 5′-GGNHTNCCCCCHTTYHTWGGWTT-3′	L4: 5′-CCRATYATDAYDGGYATWACTAT-3′
U1: 5′-TCWACAAAYCAYAAAGACATTGG-3′ U1′: 5′-TCWACWAAYCAYAARGATATTGG-3′	L1: 5′-ATWGCAAATACWGCTCCTATTGA-3′ L15: 5′-GCTACWACRTARTAWGTRTCATG-3′
U5: 5′-GWGCYCCDGAYATRGCHTTYCC-3′	L11: 5′-GGRGGMANATTRTGADTTCATTC-3′
U2: 5′-GHATAGAYGTWGAYACWCGAGC-3′ U6: 5′-TTYTTYCCNCAACAYTTYYTAGG-3′ U16: 5′-CAATTGACATTRYYCTYCATGA-3′	L2: 5′-CWCCACAAATTTCTGARCATTG-3′ L2′: 5′-ACCACAGATTTCTCTACATTG-3′ L6: 5′-TATTCATADGWYCARTATCATTG-3′
U3: 5′-GAYGTWATYCAYTCWTGAAC-3′ U11: 5′-CADSTHTAGGAKTKAARATAGATGC-3′	L7: 5′-GTGNCCWGCWATYATRTTWGC-3′
U7: 5′-TTTWAATDCCWCAAATWDYHCC-3′	L3: 5′-GMWCCWGTWAARGGTCATGG-3′
U8: 5′-GCWAAYATRATWGCAGGNCAC-3′ U12: 5′-GARTYAGAMATTTAATWCGTCC-3′	L8: 5′-ACRTCDACRAARTGYCARTATCA-3′ L13: 5′-GCRGCTTCAAABCCRAARTGRTG-3′
U9: 5′-ACAGGWTTYCAYGGRMTWCAYGT-3′	L9: 5′-TCWARTRYWCCTTKDTTTCATTC-3′ L14: 5′-RWTCAGGYCGAAACTGAWTGC-3′

2.3. DNA alignments

DNA alignments were performed using amino acid sequences with Sequence Alignment Editor Version 2.0 alpha 11 (Andrew Rambaut, software available at http://evolve.zoo.ox.ac.uk/). All regions in the alignments involving ambiguity for the position of the gaps were excluded from the analyses to avoid erroneous hypotheses of primary homology. The gap placement was considered unambiguous when only one local sequence alignment was possible due to the conservation of both gap length and amino acid motifs adjacent to the N- and C-boundaries of the gap. The reduced alignment of mt sequences consists of 3918 nucleotides (nt). It is available upon request.

2.4. Analyses of the nucleotide composition for mtDNA sequences

For each of the 84 mt sequences, the nucleotide percentages were calculated at the synonymous third positions for three groups of codons: (1) the NNN group includes all fourfold degenerate codons at third position; (2) the NNR group includes all twofold degenerate codons with a purine (A or G) at third position; and (3) the NNY group includes all twofold degenerate codons with a pyrimidine (C or T) at third position. Because of variations in the mt genetic code of the Metazoa (Knight et al., 2001; Yokobori et al., 2001), the composition of NNN, NNR, and NNY groups varies between Hemichordata, Chordata, and the two other phyla of Metazoa (Annelida and Arthropoda). The NNN group consists of the nine codons for amino acids A, G, L2, P, R, S1, S2, T, and V, except in Chordata because of exclusion of S2; the NNR group comprises the six codons for amino acids E, K, L1, M, Q, and W, except in Hemichordata because of exclusion of K and M; and the NNY group includes the eight codons for amino acids C, D, F, H, I, N, S2, and Y, except in Annelida and Arthropoda because of exclusion of S2, and in Hemichordata because of exclusion of I, N, and S2.

All six protein-coding genes here examined (i.e., *atp6* + *8*, *cox1–3*, and *nad2*) are located on the same strand for all taxa included in this study. The strand-bias in nucleotide composition was analyzed at fourfold degenerate sites by comparing the frequencies of complementary nucleotides at third positions of NNN codons; i.e., A (%) vs. T (%), and C (%) vs. G (%). It was also analyzed at twofold degenerate sites by comparing the frequency of complementary nucleotides at third positions of NNR and NNY codons. The strand-bias was described by skewness (Lobry, 1995; Perna and Kocher, 1995), which measures on one strand the relative number of A's to T's (AT skew = [A − T]/[A + T]) and C's to G's (CG skew = [C − G]/[C + G]). S-AT_4 and S-CG_4 are the skews calculated at fourfold degenerate sites, while S-AT_2 and S-CG_2 are the skews calculated at twofold degenerate sites.

2.5. Phylogenetic analyses

Phylogenetic analyses were performed using maximum parsimony (MP), maximum likelihood (ML), and Bayesian methods. MP analyses were carried out with PAUP 4.0b10 (Swofford, 2003). Bootstrap proportions (BP_{MP}) were obtained from 1000 replicates by using 10 replicates of random stepwise-addition of taxa. Bootstrap proportions (100 replicates) were also obtained under the fast-ML method (BP_{ML}) using PHYML version 2.4.3 (Guindon and Gascuel, 2003). MODELTEST 3.06 (Posada and Crandall, 1998) was used for choosing the model of DNA substitution that best fits the data, by using the Akaike Information Criterion (AIC), rather than the "hierarchical likelihood ratio tests," as Posada and Buckley (2004) have recently shown that this approach presents several impor-

Table 2
Taxonomic sampling

			mtDNA taxa	Accession No.	mt gene order type
Arthropoda					
Chelicerata	Arachnida	Acari	*Amblyomma triguttatum*	NC_005963	*Rhipicephalus*
			Carios capensis	NC_005291	*Limulus*
			Haemaphysalis flava	NC_005292	*Rhipicephalus*
			Ixodes hexagonus	NC_002010	*Ixodes hexagonus*
			Ixodes holocyclus	NC_005293	*Limulus*
			Ixodes persulcatus	NC_004370	*Limulus*
			Ornithodoros moubata	NC_004357	*Limulus*
			Ornithodoros porcinus	NC_005820	*Limulus*
			Rhipicephalus sanguineus	NC_002074	*Rhipicephalus*
			Varroa destructor	NC_004454	*Varroa*
		Amblypygi	*Phrynus sp.*	AY731172[a]	partial, *Limulus*
		Araneae	*Argiope bruennichi*	AY731171[a]	partial, *Habronattus*
			Habronattus oregonensis	AY571145	*Habronattus*
			Ornithoctonus huwena	AY309259	*Ornithoctonus*
			Heptathela hangzhouensis	AY309258	*Limulus*
		Scorpiones	*Euscorpius flavicaudis*	AY731175[a]	partial, *Limulus*
			Mesobuthus gibbosus	NC_006515	*Mesobuthus*
		Uropygi	*Mastigoproctus giganteus*	AY731174[a]	partial, *Limulus*
	Xiphosura		*Limulus polyphemus*	NC_003057	*Limulus*
	Pycnogonida		*Endeis spinosa*	AY731173[a]	partial, *Limulus*
Myriapoda	Chilopoda		*Lithobius forficatus*	NC_002629	*Lithobius*
			Scutigera coleoptrata	NC_005870	*Scutigera*
	Diplopoda		*Narceus annularus*	NC_003343	*Narceus*
			Thyropygus sp.	NC_003344	*Narceus*
Crustacea	Cephalocarida		*Hutchinsoniella macracantha*	NC_005937	*Hutchinsoniella*
	Malacostraca	Dendrobranchiata	*Penaeus monodon*	NC_002184	*Drosophila*
		Hoplocarida	*Squilla mantis*	NC_006081	*Drosophila*
		Pleocyemata	*Pagurus longicarpus*	NC_003058	*Pagurus*
			Cherax destructor	NC_011243	*Cherax*
			Callinectes sapidus	NC_006281	*Callinectes*
			Portunus trituberculatus	NC_005037	*Callinectes*
			Panulirus japonicus	NC_004251	*Drosophila*
	Branchiopoda		*Daphnia pulex*	NC_000844	*Drosophila*
			Triops cancriformis	NC_004465	*Drosophila*
			Artemia franciscana	NC_001620	*Artemia*
	Cirripedia		*Pollicipes polymerus*	NC_005936	*Pollicipes*
			Tetraclita japonica	NC_008974	*Tetraclita*
			Megabalanus volcano	NC_006293	*Megabalanus*
	Copepoda		*Tigriopus japonicus*	NC_003979	*Tigriopus*
	Branchiura		*Argulus americanus*	NC_005935	*Argulus*
	Ostracoda		*Vargula hilgendorfii*	NC_005306	*Vargula*
	Pentastomida		*Armillifer armillatus*	NC_005934	*Armillifer*
	Remipedia		*Speleonectes tulumensis*	NC_005938	*Speleonectes*
Insecta	Coleoptera	Cucujiformia	*Crioceris duodecimpunctata*	NC_003372	*Drosophila*
			Tribolium castaneum	NC_003081	*Drosophila*
		Elateriformia	*Pyrocoelia rufa*	NC_003970	*Drosophila*
	Diptera	Brachycera	*Bactrocera oleae*	NC_005333	*Drosophila*
			Ceratitis capitata	NC_000857	*Drosophila*
			Cochliomyia hominivorax	NC_002660	*Drosophila*
			Chrysomya putoria	NC_002697	*Chrysomya*
			Drosophila melanogaster	NC_001709	*Drosophila*
			Drosophila yakuba	NC_001322	*Drosophila*
		Nematocera	*Anopheles gambiae*	NC_002084	*Anopheles*
			Anopheles quadrimaculatus	NC_000875	*Anopheles*
	Hemiptera	Euhemiptera	*Philaenus spumarius*	NC_005944	*Drosophila*
			Triatoma dimidiata	NC_002609	*Drosophila*
		Sternorrhyncha	*Aleurodicus dugesii*	NC_005939	*Aleurodicus*
			Trialeurodes vaporariorum	NC_006280	*Trialeurodes*
			Schizaphis graminum	NC_006158	*Schizaphis*
			Pachypsylla venusta	NC_006157	*Drosophila*
	Lepidoptera	Bombycoidea	*Antheraea pernyi*	NC_004622	*Antheraea*

(*continued on next page*)

Table 2 (*continued*)

			mtDNA taxa	Accession No.	mt gene order type
			Bombyx mandarina	NC_003395	*Antheraea*
			Bombyx mori	NC_002355	*Antheraea*
		Pyraloidea	*Ostrinia nubilalis*	NC_003367	*Antheraea*
			Ostrinia furnacalis	NC_003368	*Antheraea*
	Odonata		*Orthetrum triangulare*	AB126005	partial, *Drosophila*
	Orthoptera		*Locusta migratoria*	NC_001712	*Locusta*
			Periplaneta fuliginosa	NC_006076	*Drosophila*
	Plecoptera		*Pteronarcys princeps*	NC_006133	*Drosophila*
	Psocoptera		*Caecilius quercus*	AF335996	partial, *Caecilius*
			Lepidopsocid RS2001	NC_004816	*Lepidopsocid*
	Thysanoptera		*Thrips imaginis*	NC_004371	*Thrips*
	Zygentoma		*Thermobia domestica*	NC_006080	*Drosophila*
			Tricholepidion gertschi	NC_005437	*Drosophila*
Collembola	Hypogastruridae		*Gomphiocephalus hodgsoni*	NC_005438	*Drosophila*
	Onychiuridae		*Onychiurus orientalis*	NC_006074	partial, *Tetrodontophora*
	Onychiuridae		*Tetrodontophora bielanensis*	NC_002735	*Tetrodontophora*
	Poduridae		*Podura aquatica*	NC_006075	partial, *Podura*
Outgroup					
	Annelida		*Lumbricus terrestris*	NC_001673	*Lumbricus*
			Platynereis dumerilii	NC_000931	*Platynereis*
	Chordata		*Bos taurus*	NC_001567	*Bos*
			Petromyzon marinus	NC_001626	*Petromyzon*
			Myxine glutinosa	NC_002639	*Bos*
	Hemichordata		*Balanoglossus carnosus*	NC_001887	*Balanoglossus*

[a] Present study.

tant advantages. The selected likelihood model was the General Time Reversible model (Yang, 1994) with among-site substitution-rate heterogeneity described by a gamma distribution and a fraction of sites constrained to be invariable (GTR + I + Γ).

Bayesian analyses were conducted on MrBayes v3.0b4 (Huelsenbeck and Ronquist, 2001). The Bayesian approach combines the advantages of defining an explicit model of molecular evolution and of obtaining a rapid approximation of posterior probabilities of trees by use of Markov Chains Monte Carlo (MCMC) (Huelsenbeck et al., 2001). Two different models of evolution were used for the analyses: (i) a single GTR + I + Γ model for all sites; and (ii) the NTE method (''Neutral Transitions Excluded'') of Hassanin et al. (2005) for minimizing the effects of a reverse strand-bias in the data. In the NTE method, the original nucleotide matrix was recoded by excluding all neutral and quasi-neutral transitions, i.e., the nucleotide-substitution types most likely to be affected by the bias. Neutral transitions are all the synonymous transitions, i.e., all transitions at third codon-positions, and transitions at first positions of leucine codons (TTR and CTN). Quasi-neutral transitions are non-synonymous transitions involving easily interchangeable amino acid residues (Hassanin et al., 1998; Naylor and Brown, 1997), i.e., ACN $\Longleftrightarrow$ GCN (T $\Longleftrightarrow$ A), ATN $\Longleftrightarrow$ GTN (I/M $\Longleftrightarrow$ V), CTN $\Longleftrightarrow$ TTY (L2 $\Longleftrightarrow$ F), ACN $\Longleftrightarrow$ ATN (T $\Longleftrightarrow$ I/M), and GCN $\Longleftrightarrow$ GTN (A $\Longleftrightarrow$ V). In practice, purines are coded by R and pyrimidines by Y at all third codon-positions, at first positions of the codons CTN (L2) and TTN (L1 and F), and at first and second positions of the codons ACN (T), ATN (I and M), GCN (A), and GTN (V). In the NTE approach, a GTR + I + Γ model was used for first and second codon-positions, and a two-state substitution model + I + Γ was used for third codon-positions.

All Bayesian analyses were done with four independent Markov chains run for 1,000,000 Metropolis-coupled MCMC generations, with tree sampling every 100 generations and a burn-in of 1000 trees. The analyses were run twice using different random starting trees to evaluate the congruence of the likelihood values and posterior clade probabilities (Huelsenbeck et al., 2002). Since several studies have shown that posterior probabilities put overconfidence on a given phylogenetic hypothesis (e.g., Douady et al., 2003), I also applied Bootstrap resampling procedures to the Bayesian approach. Bayesian Bootstrap proportions (BP_B) were calculated as follows: (1) 100 bootstrapped data sets were generated with the program SEQBOOT under the PHYLIP package Version 3.6b (Felsenstein, 2004); then, each data set was analyzed under MrBayes v3.0b4, with four independent Markov chains run for 250,000 Metropolis-coupled MCMC generations, with tree sampling every 100 generations and a burn-in of 1000 trees. Finally, the 100 Bayesian majority rule consensus trees were used for constructing the BP_B consensus tree on PAUP.

2.6. Bayesian relative rate tests

Bayesian relative rate tests were used to test the hypothesis that taxa have been diverging from their common ancestor at an equal rate (Wilcox et al., 2004). The distribution of lengths for all branches was obtained by saving branch lengths for each of 100 trees sampled during the Bayesian tree search using the NTE model. For each sampled tree, the distance from the most recent common ancestor (MRCA) of the ingroup to each of the terminal taxa was then estimated by summing branch lengths using Cadence (v1.08b, T.P. Wilcox, available at http://www.biosci.utexas.edu/antisense/). The ingroup MRCA is identified by outgroup comparison, and is simply the ancestral node shared by all Protostome taxa (Arthropoda and Annelida), to the exclusion of the Deuterostome taxa (Chordata and Hemichordata; Fig. 2A).

3. Results

3.1. Evidence for reversals of strand-compositional bias

For each of the 84 taxa examined, the strand-bias in nucleotide composition was studied on synonymous sites at third codon-positions. AT and CG skews were calculated at twofold degenerate third codon-positions (S-AT_2 and S-CG_2) and fourfold degenerate third codon-positions (S-AT_4 and S-CG_4). The results show that AT and CG skews are positive for most taxa (data not shown), confirming that the mt genome of most animals is affected by a strand-compositional bias characterized by an excess of A relative to T nucleotides and of C relative to G nucleotides. However, for 10 of the genera all skews were found negative, indicating that their mt genome presents a reverse strand-compositional bias, i.e., characterized by an excess of T relative to A nucleotides and of G relative to C nucleotides. These 10 genera belong to six different lineages: (1) the order Scorpiones with *Euscorpius* (S-$AT_4 = -0.44$; S-$CG_4 = -0.39$; S-$AT_2 = -0.17$; S-$CG_2 = -0.61$) and *Mesobuthus* (S-$AT_4 = -0.43$; S-$CG_4 = -0.76$; S-$AT_2 = -0.39$; S-$CG_2 = -0.86$); (2) the spider suborder Opisthothelae with *Argiope* (S-$AT_4 = -0.07$; S-$CG_4 = -0.56$; S-$AT_2 = -0.09$; S-$CG_2 = -0.56$), *Habronattus* (S-$AT_4 = -0.19$; S-$CG_4 = -0.53$; S-$AT_2 = -0.12$; S-$CG_2 = -0.79$), and *Ornithoctonus* (S-$AT_4 = -0.16$; S-$CG_4 = -0.68$; S-$AT_2 = -0.20$; S-$CG_2 = -0.70$); (3) the mite suborder Mesostigmata only represented by *Varroa* (S-$AT_4 = -0.23$; S-$CG_4 = -0.47$; S-$AT_2 = -0.05$; S-$CG_2 = -0.57$); (4) the order Cephalocarida with *Hutchinsoniella* (S-$AT_4 = -0.69$; S-$CG_4 = -0.73$; S-$AT_2 = -0.24$; S-$CG_2 = -0.79$); (5) the order Copepoda with *Tigriopus* (S-$AT_4 = -0.25$; S-$CG_4 = -0.45$; S-$AT_2 = -0.21$; S-$CG_2 = -0.31$); and (6) the family Aleyrodidae (whiteflies) with *Aleurodicus* (S-$AT_4 = -0.34$; S-$CG_4 = -0.74$; S-$AT_2 = -0.04$; S-$CG_2 = -0.66$) and *Trialeurodes* (S-$AT_4 = -0.54$; S-$CG_4 = -0.34$; S-$AT_2 = -0.25$; S-$CG_2 = -0.50$). As these six lineages are not closely related, it can be assumed that asymmetric mutational patterns have been reversed six times independently.

3.2. Artifacts of reconstruction due to independent reversals of asymmetric mutational constraints

The mtDNA data matrix including 3918 nt characters and 84 taxa was first analyzed by MP and ML methods. Bootstrap values obtained with both of these methods are indicated in the ML tree of the Fig. 1. Taking into account the background knowledge in arthropod classification and phylogeny, many taxa present odd positions. In particular, the 10 genera with reverse strand-compositional bias are grouped together in spite of their known distant relationships ($BP_{ML/MP} = 50/78$; see the gray rectangle in Fig. 1). Consequently, many higher taxa are found para- or polyphyletic, including Paradoxopoda (i.e., Chelicerata + Myriapoda), Myriapoda, Chelicerata, Araneae, Arachnida, Acari, Pancrustacea, Crustacea, Hexapoda, and Insecta.

By contrast, the Bayesian tree reconstructed by using the GTR + I + Γ model is more in agreement with what is known about arthropod phylogeny (data not shown). Three taxa, which were found para- or polyphyletic in the ML and MP analyses, are now monophyletic: (1) Araneae (Bayesian posterior probability: $PP_B = 1$), as *Heptathela* is grouped with Opisthothelae spiders (*Argiope*, *Habronattus*, and *Ornithoctonus*); (2) Acari ($PP_B = 1$), as *Varroa* is sister-group of Ixodida mites; and (3) Sternorrhyncha ($PP_B = 0.90$), as whiteflies (*Aleurodicus* and *Trialeurodes*) are allied with *Pachypsylla* and *Schizaphis*. However, many higher taxa remain para- or polyphyletic, such as Pancrustacea, Crustacea, Hexapoda, Insecta, Paradoxopoda, Chelicerata, Myriapoda, and Arachnida (data not shown). In particular, some taxa with reverse strand-bias remain associated in spite of their distant relationships: *Hutchinoniella* (Cephalocarida) and *Tigriopus* (Copepoda) are grouped together ($PP_B = 1$), and they are sister-group of scorpions (*Euscorpius* and *Mesobuthus*) ($PP_B = 1$).

3.3. "Neutral Transitions Excluded" model

Because multiple reversals of asymmetric mutational patterns are expected to mislead phylogenetic inferences based on mtDNA sequences, Hassanin et al. (2005) have used the "Neutral Transitions Excluded" (NTE) model for improving tree reconstruction. The Bayesian tree obtained with the NTE model (Fig. 2A) is more in agreement with current views on arthropod evolution, as several higher taxa are now monophyletic. However, Bayesian Bootstrap proportions (BP_B) show that most of these deep divergences are poorly supported: Pancrustacea ($PP_B = 0.65$; not supported by BP_B), Paradoxopoda ($PP_B = 1$; $BP_B = 32$), Chelicerata ($PP_B = 1$; $BP_B = 26$), and Myriapoda ($PP_B = 1$; $BP_B = 69$). This raised the suspicion that the lack of robustness was due to the presence of several highly divergent, long-branched sequences, and this possibility will now be explored.

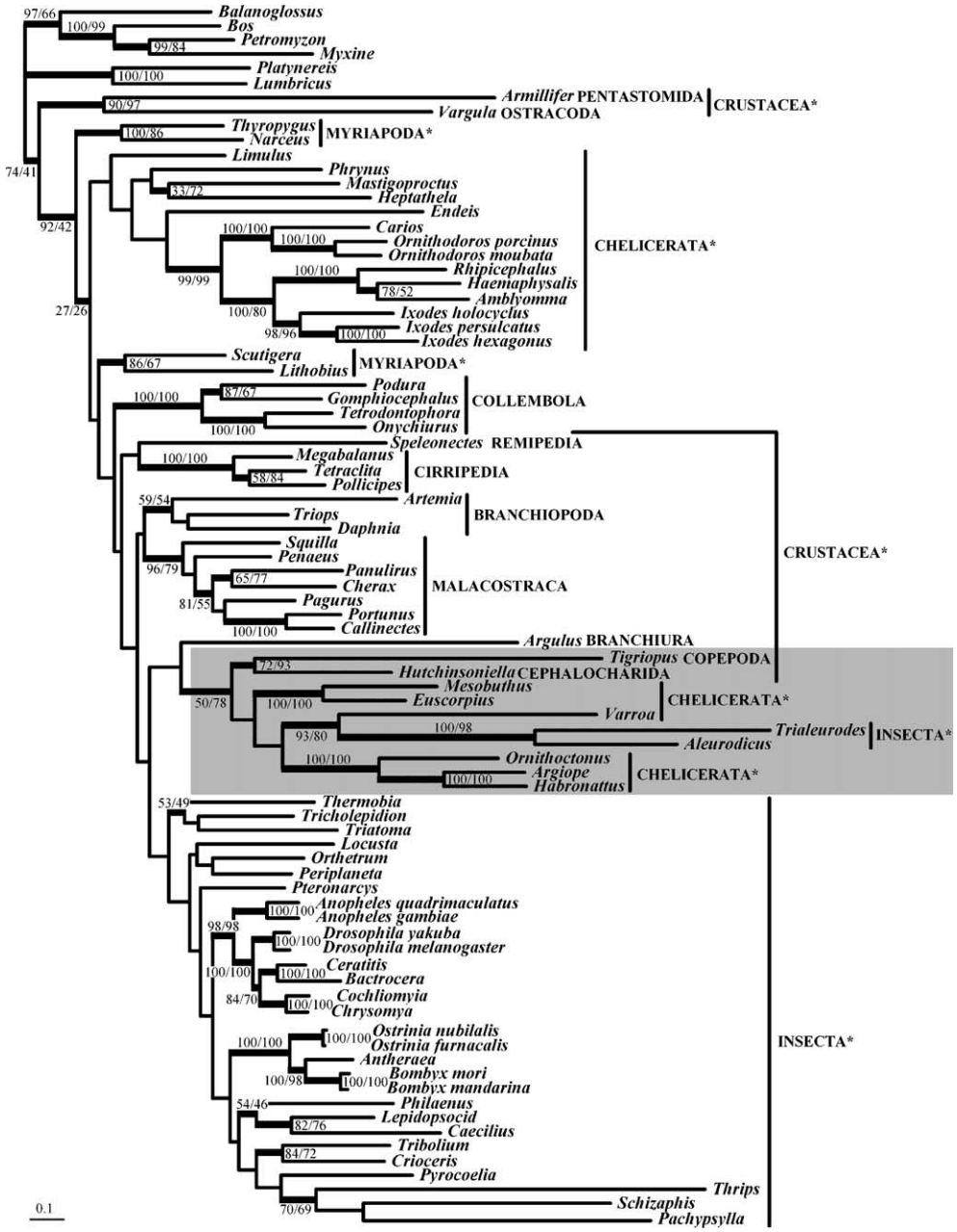

Fig. 1. Maximum-likelihood tree. The model used is the one selected by MODELTEST 3.06, i.e., GTR + I + Γ. The values indicated on the branches correspond to the Bootstrap percentages obtained either with the maximum-likelihood analysis (BP_{ML}) (to the left of the slash), or with the maximum parsimony analysis (BP_{MP}) (to the right of the slash). Bold lines indicate the nodes supported by BP_{ML} and BP_{MP} values superior to 50; that is, the best supported nodes. Taxa enclosed into the gray box are characterized by a reverse strand-bias. Higher taxa with an asterisk (*) were not found to be monophyletic.

3.4. Detection of long branches

Bayesian relative rate tests calculated under the NTE model reveal that 13 of the 84 sequences fall outside of an area defined by the mean branch length and its standard deviation (highlighted in gray in Fig. 2B). For these 13 taxa, the rates of substitution are significantly faster than in other species. They include all the four insects of Sternorrhyncha, all the three spiders of Opisthothelae, *Argulus*, *Armillifer*, *Thrips*, *Tigriopus*, *Vargula*, and *Varroa*. Interestingly, these lineages present unusual features, as all can be placed into at least one of the five following categories: (1) taxa with a reverse strand-bias, such as genera of Aleyrodidae and Opisthothelae, *Tigriopus*, and *Varroa*; (2) taxa in which one or several genes of the studied fragment (N2-C1-C2-A8-A6-C3 in Fig. 2B) have been translocated, such as *Thrips* and *Tigriopus*; (3) taxa in which the mt genome has two putative control regions, such as *Argulus*, *Thrips*, and *Vargula*; (4) taxa which have a parasitic lifestyle, such as Branchiura (*Argulus*), which are ectoparasites on fishes, Pentastomida (*Armillifer*), which inhabit respiratory tracts of tetrapods, and Sternorrhyncha, which are parasites of

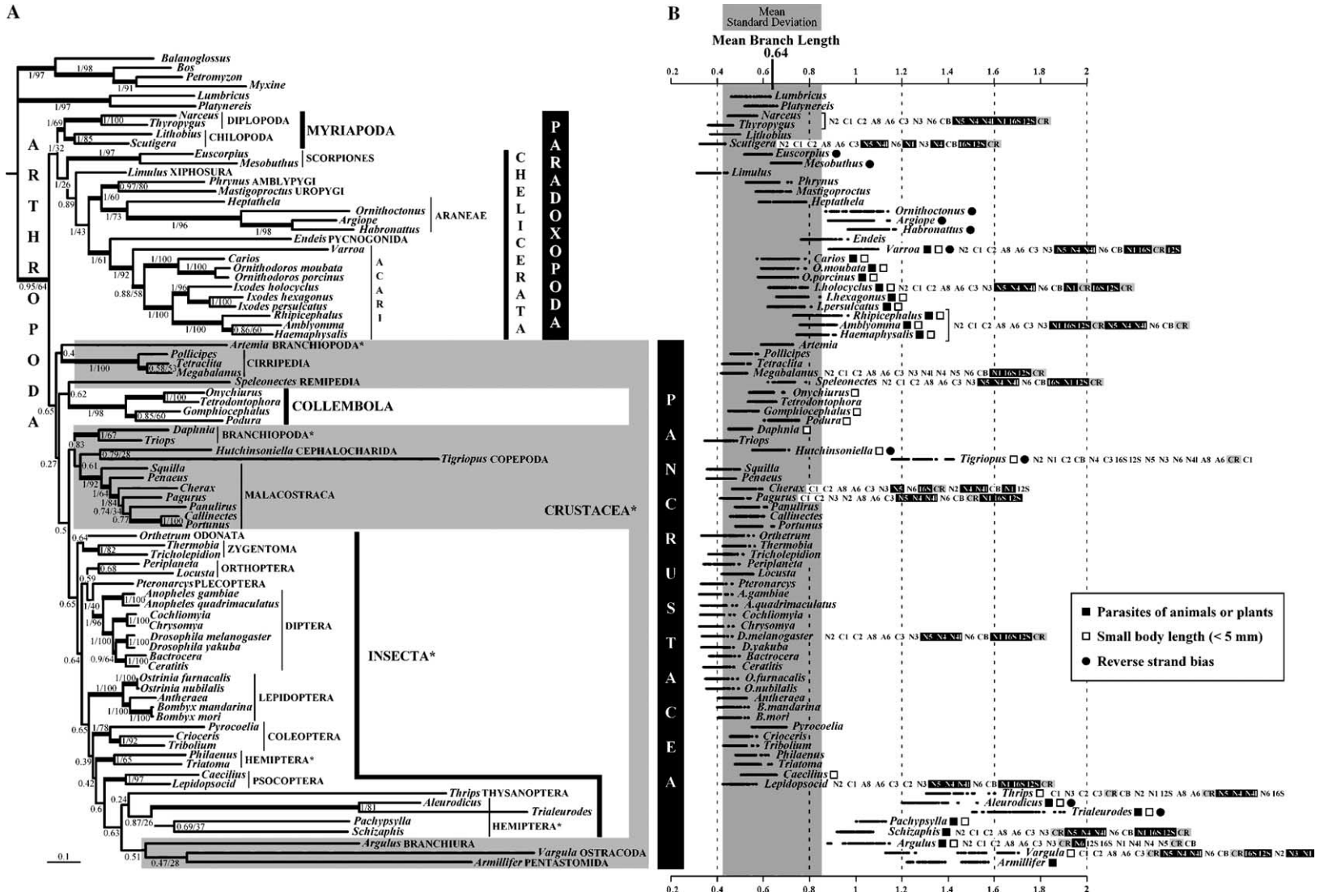

Fig. 2. Bayesian tree obtained with the "Neutral Transitions Excluded" model. (A) The Bayesian tree was obtained using all the 84 taxa, and with the "Neutral Transitions Excluded" model, which codes purines by R and pyrimidines by Y at all third codon-positions, at first positions of the codons CTN and TTN, and at first and second positions of the codons ACN, ATN, GCN, and GTN, and which apply a GTR + I + Γ model for first and second codon-positions, and a two-state substitution model + I + Γ for third codon-positions. The values indicated on the branches are posterior probabilities (to the left of the slash), and bootstrap proportions (BP_B) (to the right of the slash; values less than 20% are not shown). Bold lines indicate that the clade was supported by a BP_B value superior to 50. Higher taxa with an asterisk (*) were not found to be monophyletic. The big gray bars indicate the Crustacea. (B) Bayesian relative rates tests were used to test the hypothesis that taxa have been diverging from their common ancestor at an equal rate (Wilcox et al., 2004). For each species of *Protostomia*, the distribution of branch lengths (from the most recent common ancestor of *Protostomia* to the terminal taxa) is indicated for 100 trees sampled during the Bayesian search. Symbols indicate parasites (■), small-bodied organisms (□), or species possessing a mitochondrial genome with a reverse strand-bias (●). Most arthropod species have a mitochondrial genome where genes coding for proteins and rRNAs are arranged as in *Drosophila melanogaster*. All exceptions are indicated in the figure. Abbreviations used are the following: N1, N2, N3, N4, N4l, N5, and N6 for NADH dehydrogenase subunit 1, 2, 3, 4, 4l, 5, and 6; C1, C2, and C3 for cytochrome oxidase subunit 1, 2 and 3; CB for cytochrome *b*; A8 and A6 for ATP synthase F0 subunit 8 and 6; 12S and 16S for 12S and 16S ribosomal RNAs; and CR for control region.

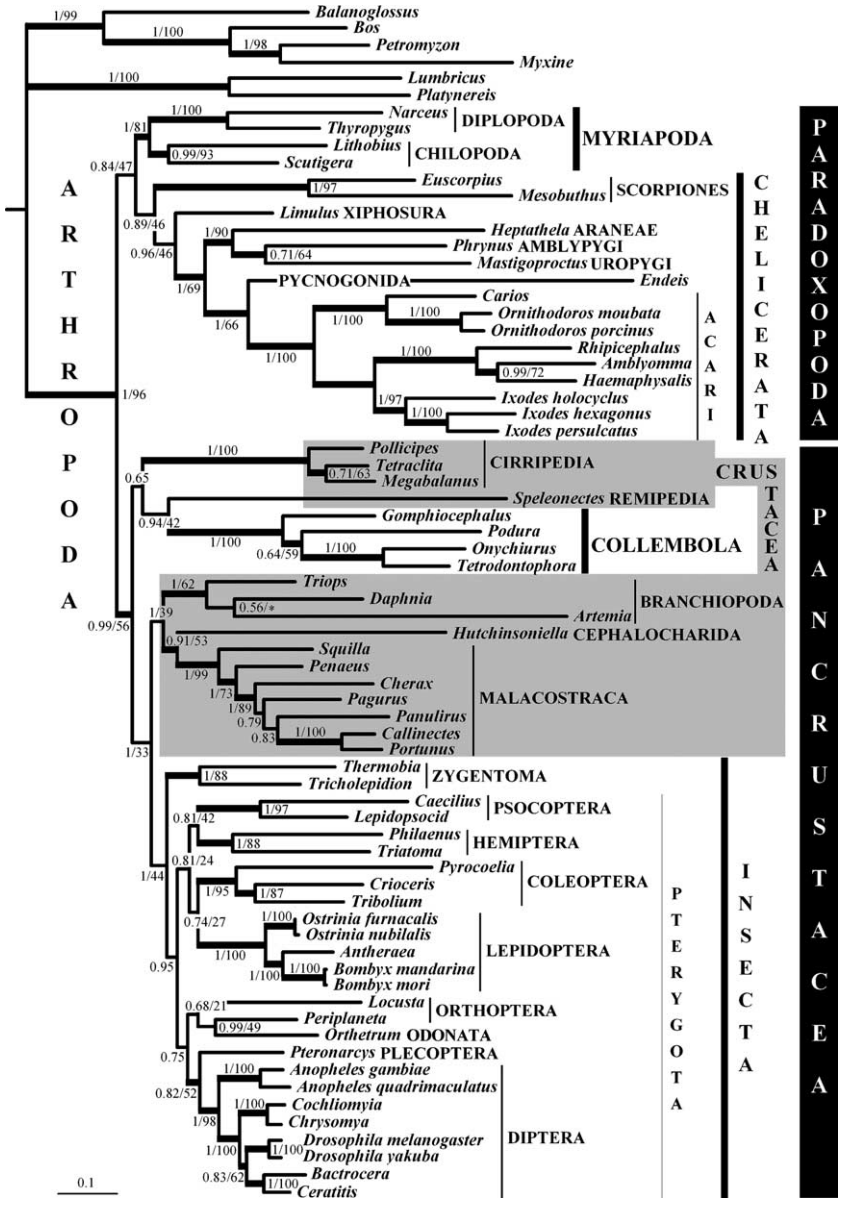

Fig. 3. Bayesian tree obtained by excluding taxa with a long branch. The analyses were done by excluding the 13 genera with branch lengths significantly longer than other taxa (see Fig. 2B); i.e., Opisthothelae spiders (*Argiope*, *Habronattus*, and *Ornithoctonus*), Sternorrhyncha insects (*Aleurodicus*, *Pachypsylla*, *Schizaphis*, and *Trialeurodes*), *Argulus*, *Armillifer*, *Thrips*, *Tigriopus*, *Vargula*, and *Varroa*. The model used is the NTE model (for more details, see in Fig. 2A and Section 2). The values indicated on the branches are posterior probabilities (to the left of the slash), and Bootstrap proportions (BP_B) (to the right of the slash; values less than 20% are not shown). Bold lines indicate that the clade was supported by a BP_B value superior to 50. Asterisk indicates that an alternative hypothesis was supported by a BP_B value greater than 50. The big gray bars indicate the Crustacea.

plants; and (5) lineages characterized by a very small body size, i.e., less than 5 mm, such as Ostracoda (*Vargula*) and Thysanoptera (*Thrips*). These data suggest, therefore, that three main factors may explain long branches: changes in the organization of the mt genome, phenotypic features of the taxa (body size), and a particular mode of speciation inherent to the parasitic lifestyle.

3.5. Exclusion of long-branch taxa

In a second approach of Bayesian tree reconstruction with the NTE model, taxa with a long branch were excluded from the phylogenetic analyses. Fig. 3 presents the tree obtained using a reduced data set excluding the 13 species with branch lengths significantly longer than other arthropods. The topology is very similar to the one previously obtained with the complete data set, but most higher taxa are now recovered with higher support: Arthropoda ($BP_B = 96$ vs. 64), Paradoxopoda ($BP_B = 47$ vs. 32), Myriapoda ($BP_B = 81$ vs. 69), Chelicerata ($BP_B = 46$ vs. 26), Tetrapulmonata, i.e., Araneae + Amblypygi + Uropygi ($BP_B = 90$ vs. 60), Pancrustacea ($BP_B = 56$ vs. not supported), Insecta ($BP_B = 44$ vs. not found), Branchiopoda ($BP_B = 62$ vs. not found), Malacostraca ($BP_B = 99$ vs.

92), and Collembola ($BP_B = 100$ vs. 98). The grouping of *Endeis* (Pycnogonida) with Acari was also retrieved with better support ($BP_B = 66$ vs. 61). The only exception with lower branch support is the group Pedipalpi, which includes Amblypygi and Uropygi ($BP_B = 64$ vs. 80), but, in this case, the presence of only one species of Araneae, rather than four in the previous analyses, may be insufficient for breaking long branches during tree reconstruction.

4. Discussion

4.1. Multiple reversals of asymmetry during the evolution of the mitochondrial genome

Typically, the mt genomes of Metazoa present a clear strand-bias in base composition; i.e., one strand is characterized by positive AT and CG skews, i.e., A (%) > T (%) and C (%) > G (%), whereas the other strand, simply because of base complementarity, is characterized by negative skews, i.e., T (%) > A (%) and G (%) > C (%). This bias is the consequence of asymmetric patterns of change where certain mutations are more common than their complements, thereby generating inequalities between the frequencies of the complementary bases A/T and C/G (Lobry, 1995; Sueoka, 1995; Wu and Maeda, 1987). Higher levels of deaminations of A and C bases on one of the two strands, during either replication or transcription processes, may be involved in establishing the strand-bias in the mt genome of animals (Hassanin et al., 2005; Reyes et al., 1998; Tanaka and Ozawa, 1994).

In the present study, 10 species from 6 unrelated lineages of arthropods have a strand-bias that is reversed compared with other taxa. They include insects of the family Aleyrodidae (*Aleurodicus* and *Trialeurodes*), spiders of the Suborder Opisthothelae (*Argiope*, *Habronattus*, and *Ornithoctonus*), scorpions (*Euscorpius* and *Mesobuthus*), *Hutchinsoniella*, *Tigriopus*, and *Varroa*. By analyzing the evolution of gene-order organization in Metazoa, Hassanin et al. (2005) have proposed that independent inversions of the control region of the mt genome can result in independent reversals of asymmetric mutational patterns. The control region, which is generally identified as an "A + T rich" region in arthropods, contains one origin of replication and all initiation sites used for transcription (Taanman, 1999). So, whatever the mechanism involved in the asymmetric patterns of mutation, i.e., replication and/or transcription, the control region appears to be the key region for determining the strand-compositional bias. Therefore, an inversion of the control region is expected to produce a reversal of asymmetric mutational patterns in the mtDNA, resulting over time in a complete reversal of strand-compositional bias. In case of inversion of the control region, the pattern of substitution is expected to reverse as in the case of gene inversion; i.e., the nucleotide-substitution types that evolved quickly in the past will now evolve slowly, and vice versa. Unfortunately, it is not easy to detect an inversion of the control region, because this is the most variable region of the mtDNA, making it impossible to align between distant species (e.g., Mardulyn et al., 2003). Consequently, the orientation of the control region cannot be determined by simple comparisons of sequences. Actually, the hypothesis for an inversion of the control region is upheld, first, by evidence from Echinoderms (Hassanin et al., 2005), and, second, by the fact that most taxa with a reverse strand-bias display a different position for the control region: that is, while the control region in most arthropods is found between 12S rRNA and I-tRNA genes, it is located between M- and Q-tRNA genes in the common ancestor of Opisthothelae, between D- and N-tRNA genes in *Mesobuthus*, between W- and S1-tRNA genes in *Hutchinsoniella* (Lavrov et al., 2004), between *W*-tRNA and CO1 genes in *Tigriopus* (Machida et al., 2002), between 12S rRNA and Q-tRNA genes in *Trialeurodes* (Thao et al., 2004), and between C-tRNA and 12S rRNA genes in *Varroa* (Navajas et al., 2002). *Aleurodicus* is exceptional because its control region is located, as in most arthropods, between 12S rRNA and I-tRNA genes. However, the control region might also have been inverted in this genus, but in this case, without implication of adjacent genes, i.e., 12S rRNA or I-tRNA.

4.2. Long-branch attraction artifacts due to independent reversals of mutational patterns

In a recent paper, Hassanin et al. (2005) have shown that reversals of asymmetric mutational patterns can be particularly misleading for studying deep divergences with mtDNA sequences. When similar reversals of mutational patterns occurred independently in different lineages, taxa tend to acquire similar base composition, and, as a consequence, they can group together due to the long-branch attraction (LBA) phenomenon (Felsenstein, 1978). Here, this artifact of reconstruction is obvious with ML and MP methods (Fig. 1), because all taxa characterized by a reverse strand-bias fall together in the same clade in spite of their distant relationships, i.e., *Argiope*, *Habronattus* and *Ornithoctonus* (Araneae, Opisthothelae), *Euscorpius* and *Mesobuthus* (Scopriones), *Varroa* (Acari, Mesostigmata), *Aleurodicus* and *Trialeurodes* (Insecta, Aleyrodidae), *Hutchinsoniella* (Cephalocardia), and *Tigriopus* (Copepoda). This hybrid clade is better supported in parsimony than in ML based on the GTR + I + Γ model (Fig. 1; $BP_{MP} = 78$ versus $BP_{ML} = 50$), suggesting that the MP method is more sensitive to this artifact of reconstruction than are model-based methods. As model-based methods of reconstruction can deal with multiple hits (Swofford et al., 2001), and can account for heterogeneity of evolutionary rates among sites, a parameter especially important for overcoming LBA (Cunningham et al., 1998), they have a logical advantage over parsimony. This view is reinforced by the Bayesian analysis using the GTR + I + Γ model (data not shown), where several genera with reverse strand-bias have a more-reliable taxonomic position, i.e.,

Varroa with mites, Opisthothelae spiders (*Argiope*, *Habronattus*, and *Ornithoctonus*) with other spiders, and whiteflies (*Aleurodicus* and *Trialeurodes*) with other soft bugs. However, *Hutchinoniella* and *Tigriopus* are still artifactually grouped with scorpions (*Euscorpius* and *Mesobuthus*) despite their distant relationships. This very unreliable grouping shows that the LBA phenomenon due to independent reversals of strand-bias still remains problematic with the Bayesian approach. It can be argued that both Bayesian and ML methods are strongly affected because the assumed model of substitution is strongly violated. At present, most models assume (1) that the sequences have evolved with the same pattern of nucleotide substitution (homogeneity of the evolutionary process), and (2) that all lineages exhibit the same nucleotide composition (i.e., stationarity) (Rosenberg and Kumar, 2003; Swofford et al., 2001; Tamura and Kumar, 2002). If these assumptions are not satisfied, as here with the mt sequences, estimation of branch lengths is likely to be biased, which may result in erroneous groupings in the inferred phylogenetic trees (Galtier and Gouy, 1998; Hassanin et al., 2005; Rosenberg and Kumar, 2003; Tamura and Kumar, 2002).

4.3. Higher rates of substitution

In Fig. 2B, 13 genera have branches significantly longer than others. For taxa with reverse strand-bias, long branches can be attributed to the reverse pattern of substitution, and apparent inversion of the control region, as explained above. For all other taxa, long branches are probably the consequences of accelerated rates of substitution.

Four biological factors have been related to variations in rates of nucleotide substitution among evolutionary lineages: generation time, metabolic rate, population size, and DNA-repair efficiency. (1) According to the generation-time hypothesis (Kohne, 1970; Laird et al., 1969; Wu and Li, 1985), organisms with shorter generation times have a greater number of germ-cell divisions per year, and thus more replication errors per unit time leading to faster mutation rates. (2) According to the metabolic-rate hypothesis (Martin and Palumbi, 1993), organisms with higher metabolic rates have higher mutation rates, because of increased rates of DNA synthesis, and higher incidence of DNA damage generated by elevated concentrations of free radicals produced as byproducts of metabolism. Because metabolic rate and generation time both vary with body size and temperature (Gillooly et al., 2001), both hypotheses (1) and (2) predict that ectothermal organisms of small body size and/or living in warm environments will exhibit higher rates of mutations (Gillooly et al., 2005). (3) According to the population-size hypothesis, populations with small effective sizes should experience faster rates of evolution than populations with larger effective sizes, because of the increased influence of drift on selection (Ohta, 1992). Founder events should occur frequently in parasites during transmission to new host individuals. This particularity may be an important factor for explaining higher rates of speciation in parasites (Page et al., 1998), and therefore their possible elevated rates of substitution. (4) According to the DNA-repair efficiency hypothesis (Britten, 1986), the contrast in rates of sequence change is due to evolutionary variation of biochemical mechanisms such as DNA repair or replication. Obviously, it is difficult to test this hypothesis here, but, if this factor is the only one involved, changes in mutational rates are expected to be independent of other factors, such as body size or parasitism.

The present study suggests a causal link between parasitic lifestyle and faster molecular evolution, since all parasitic species are associated with long branches (Fig. 2B). A range of studies has already proposed that an increase in the rates of substitution is coincident with the adoption of the parasitic lifestyle. For example, Dowton and Austin (1995) found elevated rates of mtDNA sequence divergence in parasitic wasps compared with non-parasitic wasps. Similar results were also found in other groups of parasites, such as lice (Page et al., 1998), or leeches (Martin et al., 2000). Although generation time has been suggested as an explanation for the elevated evolutionary rates in parasites, the most likely explanation is their increased rate of speciation compared with the non-parasitic lineages, because of frequent founder events from population of small effective size (Page et al., 1998).

The inverse relationship between rate of molecular evolution and body size has been described for a range of vertebrate taxa, including mammals, birds, reptiles, amphibians, and fishes (Bromham, 2002; Bromham et al., 1996; Martin and Palumbi, 1993; Mooers and Harvey, 1994). The cause of this relationship is not certain. The two most common explanations involve difference in generation time, and variation in metabolic rate (Bromham et al., 1996; Mooers and Harvey, 1994). There is as yet no evidence that body size influences rate of molecular evolution in invertebrates. The relationship between body size and rate of molecular evolution is not clear in the present study, because several counter-examples exist, such as small-but-conservative Collembola and *Daphnia*, and because the long branches found for several tiny arthropods may be also explained by a parasitic lifestyle (i.e., Acari, Sternorrhyncha, and *Argulus*), or by important changes in the organization of their mt genome (e.g., *Argulus*, *Thrips* and *Vargula*) (see below).

It is astonishing to note that most genera associated with a long branch possess a mt genome with unusual features, including a reverse strand-compositional bias possibly generated by inversion of the control region (e.g., Opisthothelae), or the possession of two putative control regions (e.g., *Argulus*, *Thrips*, or *Vargula*), or important rearrangements in the gene order (e.g., *Tigriopus* and *Thrips*) (Fig. 2B). I suggest that genomic rearrangements, such as gene translocation and duplication of the control region, generate changes in the substitution rates. Indeed, the study of AT and CG skews in mammals has shown that the strand-compositional bias varies between mt genes according to their

distance from the two origins of replication (Reyes et al., 1998; Tanaka and Ozawa, 1994). This implies, therefore, that rates of substitution vary between genes, according to their position in the mt genome, and that, a gene translocation should change the rates of substitution, more especially if the distance separating the translocated gene from the control region has been significantly increased or decreased. According to this hypothesis, the presence of two functional control regions in the mt genome is also expected to have an impact on the rates of substitution: first, by modifying the process and/or rates of replication and transcription, and second, by changing the distance between genes and control regions. From this point of view, it is noteworthy to note that all taxa with duplicated control regions have long branches, excepting *Ixodes holocyclus* (Fig. 2B), but in the latter species, the two control regions are closely located (i.e., only separated by 16S and 12S genes), and the duplication event was also too much recent for permitting accumulation of numerous substitutions. By contrast, the metastriate ticks with duplicated control regions (i.e., the clade composed of *Rhipicephalus*, *Amblyomma*, and *Haemaphysalis*) have elevated rates of evolution compared to other ticks (Fig. 2B). In their common ancestor, a block of genes containing *nad1*, *16S*, *12S*, and the control region was translocated between *nad3* and *nad5* (Black and Roehrdanz, 1998). Therefore, the two control regions are far apart separated by several genes, including *nad5*, *nad4*, *nad4l*, *nad6*, and *cyb* (Fig. 2B). Furthermore, the duplicated control regions undergo concerted evolution, suggesting that both are indeed active in the genome (Black and Roehrdanz, 1998). However, in other "long-branch" genera possessing a mt genome with two putative control regions (i.e., *Argulus*, *Scizaphis*, *Thrips*, and *Vargula*) alternative explanations can also be proposed, because they are characterized by a parasitic lifestyle (i.e., *Argulus* and *Scizaphis*), and/or a small body size (i.e., *Argulus*, *Thrips*, and *Vargula*). Comparative analyses with nuclear genes will be of interest value to decide between these hypotheses. Indeed, if body size and parasitism are the main factors involved in accelerated rates in these genera, long branches will be also found with nuclear genes.

In conclusion, the present analyses suggest that genomic rearrangements, a parasitic lifestyle, and perhaps body size, have an impact on the rates of substitution. In some taxa, these different causes may act in synergy to explain long branches.

4.4. Strategies for analyzing mitochondrial sequences

Because reversals of mutational patterns can considerably mislead phylogenetic inferences based on mtDNA sequences, specific strategies are recommended for avoiding misinterpretation. By comparing the strand-compositional bias in genes transcribed by opposite strands, Hassanin et al. (2005) previously found that a gene inversion will generate a reversal of asymmetric mutational patterns. For this reason, I excluded from the present study any taxa in which one or several (but not all) of the six genes were inverted.

As a preliminary to phylogenetic inferences, analyses of base composition at synonymous sites can be performed to detect taxa with reversed asymmetric mutational patterns. Then, to deal with the problem of LBA, these taxa can be excluded entirely from the phylogenetic analyses. Such strategy seems to be extreme, however, because informative taxa could be removed, limiting the value of the phylogenetic results. My analysis based on the NTE model seems a better alternative. Two arguments for adopting the NTE model, which aims at excluding neutral or quasi-neutral transitions, are: (i) the asymmetric mutational constraints act principally by the way of transitions rather than transversions (Hassanin et al., 2005); and (ii) selected transitions are expected to be less affected by changes in asymmetry than are neutral transitions. Thus, the NTE model is designated to improve signal by eliminating noise (i.e., nucleotide-substitution types likely to be more affected by homoplasy). This model should improve both parameter estimations and tree reconstruction. Here, this strategy retrieves most of the higher taxa of Arthropoda (Fig. 2A), but the presence of several taxa with accelerated rates of evolution remains problematic for tree reconstruction. Therefore, all taxa with branch lengths significantly longer than other taxa (Fig. 2B) were excluded from the next analyses. The robustness of the nodes, as expressed by Bootstrap values, was increased with this approach (Fig. 3). Consequently, these analyses suggest that phylogenetic inferences with mt sequences can be improved by using a justified approach for selecting genes, and evolutionary models, and by eliminating the most divergent taxa.

4.5. Phylogeny of Arthropods

4.5.1. A basal dichotomy separating Pancrustacea and Paradoxopoda

Traditionally arthropods have been viewed as a monophyletic group united by striking synapomorphies including a hard, jointed exoskeleton, articulated appendages, head shield, unique hemocyanin polymer, etc. (Brusca and Brusca, 2003; Snodgrass, 1938). In the past, some authors have challenged this view by proposing a polyphyletic origin of arthropods (Anderson, 1973; Fryer, 1996; Manton, 1977; Min et al., 1998). However, arthropod monophyly is now widely accepted, and most recent molecular analyses support this idea (e.g., Giribet et al., 2001; Mallatt et al., 2004; Regier et al., 2005a; Shultz and Regier, 2000), including the present study based on mtDNA sequences (Figs. 1–3).

Four main conflicting phylogenetic arrangements have been proposed for the interrelationships among the four major groups of arthropods, i.e., Crustacea, Hexapoda, Myriapoda, and Chelicerata. These arrangements are: (i) Chelicerata, [Crustacea, (Myriapoda, Hexapoda)]; (ii) Chelicerata, [Myriapoda, (Crustacea, Hexapoda)]; (iii)

(Chelicerata, Myriapoda), [Crustacea, Hexapoda]; and (iv) (Chelicerata, Crustacea), [Myriapoda, Hexapoda]. In the classification of Snodgrass (1938), which formerly was widely accepted, Myriapoda and Hexapoda are sister-groups, together forming a taxon called Atelocerata, and Crustacea are allied with Atelocerata into the taxon Mandibulata. The Atelocerata were united by several attributes: loss of the second antennae (as the name Atelocerata implies), loss of the outer branch of the limb (uniramy), a tracheal respiratory system, and Malpighian tubules for excretion. The Mandibulata were united on the basis of the mandibles, and a similar head and head-appendage structure. Some analyses combining morphological and molecular characters continue to obtain the Snodgrass' classification: Wheeler et al. (1993) with 18S rRNA and ubiquitin sequences, Edgecombe et al. (2000) with histone H3 and U2 genes. However, most paleontologists suggest that Mandibulata are not monophyletic and propose a phylogeny where Crustacea are associated with Chelicerata, together forming a taxon called Schizoramia (e.g., Cisne, 1974; Wills et al., 1998). To my knowledge, this hypothesis was never found by the molecular approach. In contrast, most molecular studies indicate that Crustacea and Hexapoda are united into the clade named Pancrustacea. This node was found in several independent studies using different molecular markers: 18S rRNA (Giribet et al., 1996; Turbeville et al., 1991), combined 18S and 28S rRNAs (Friedrich and Tautz, 1995; Mallatt et al., 2004), complete mtDNA genome (Hwang et al., 2001), elongation factor-1 α (*EF-1α*) and the largest subunit of RNA polymerase II (*Pol II*) (Shultz and Regier, 2000), and elongation factor-2 (*EF-2*) (Regier and Schultz, 2001; Regier et al., 2005a). The present mtDNA analyses also favor the Pancrustacea hypothesis (Figs. 2A and 3) . The Pancrustacea concept is also supported by some morphological characters, as hexapods and crustaceans share several features of the nervous system (Strausfeld, 1998; Whitington and Bacon, 1998), the visual system (Harzsch and Walossek, 2001; Nilsson and Osorio, 1998), and the process and control of development, especially segmentation (Averof and Akam, 1995a,b). This concept demands that the similarities shared between hexapods and myriapods (Atelocerata) represent convergences. Interestingly, most of them can be interpreted as adaptations in response to terrestrialization (Averof and Akam, 1995b).

While most molecular studies provide evidence for Pancrustacea, the question of their sister-group remains highly debated in the literature. Many authors argue that Pancrustacea are closely related to Myriapoda, supporting the notion of Mandibulata (Edgecombe et al., 2000; Giribet et al., 2001; Wheeler et al., 1993). Others propose to link Pancrustacea with Chelicerata (Regier et al., 2005a). Others consider that Pancrustacea are associated to a clade composed of Chelicerata and Myriapoda, named Paradoxopoda by Mallatt et al. (2004). Although Chelicerata and Myriapoda were never grouped together on the basis of morphological characters, three different kinds of molecular data support this clade: 18S/28S rRNA sequences (Friedrich and Tautz, 1995; Mallatt et al., 2004), mt genome (Hwang et al., 2001), and Hox genes (Cook et al., 2001). The present analyses based on mtDNA sequences support this point of view, but the node remains poorly supported (Figs. 2A and 3; $BP_B = 32$ and 47; $PP_B = 1$ and 0.84). Interestingly, the first morphological indication for this clade has been recently suggested by the fact that chelicerates and myriapods share the same developmental mechanism for neurogenesis (Dove and Stollewerk, 2003; Kadner and Stollewerk, 2004).

4.5.2. Relationships within Pancrustacea

Traditionally, the subphylum Hexapoda comprises the class Insecta and three wingless groups: Collembola, Diplura, and Protura. Hexapods share several remarkable features including a tripartite body plan composed of a head, thorax, and abdomen, and three pairs of thoracic legs (Brusca and Brusca, 2003). By contrast, the extant Crustacea are not clearly diagnosed, although Schram (1986) identified several features that are shared by most taxa, including a characteristic assortment of cephalic appendages with two pairs of antennae, a pair of mandibles, two pairs of maxillae, and a unique nauplius larva. The group has consistently emerged as monophyletic from cladistic treatments of morphological data (e.g., Edgecombe et al., 2000; Walossek and Müller, 1998; Wheeler et al., 1993).

By analyzing mtDNA sequences, Nardi et al. (2003) concluded that Hexapoda are paraphyletic, due to the placement of Collembola as sister-group of a clade uniting Insecta with Malacostraca and Branchiopoda. Their conclusions were drawn from ML and Bayesian analyses at the amino acid level of four of the 13 mt proteins (i.e., *cox1*, *cox2*, *cox3*, and *cob*). As pointed out in Delsuc et al. (2003), their phylogenetic analyses were potentially biased by the use of a non-adapted model of mt amino acid substitution. In addition, their taxonomic sampling was insufficient, with 18 species of Insecta, but only two species of Collembola and four species of Crustacea. At least, their analyses were carried out with several sequences evolving with a reverse mutational pattern, including taxa with gene inversion (*cob* in *Katharina* and *Loligo*, and *cox3* in *Albinaria*), and taxa with a possible inversion of the control region (i.e., *Artemia*, *Katharina*, and *Heterodoxus*; Hassanin et al., 2005). The incorporation of these sequences into the analyses may strongly affect the estimation of the parameters of the evolutionary model, and then, tree reconstruction.

The present data do not produce a strong signal for interrelationships between Insecta, Collembola, and Crustacea lineages. I therefore follow the point of view of Delsuc et al. (2003) who concluded that the monophyly of Hexapoda could not be rejected, considering the mt data currently available. By contrast, nuclear genes are clearly in favor of a monophyletic Hexapoda (Mallatt et al., 2004; Regier et al., 2005a).

4.5.3. Monophyly of Myriapoda

The subphylum Myriapoda includes four classes: Chilopoda (centipedes), Diplopoda (millipedes), Pauropoda (pauropodans), and Symphyla (symphylans) (Brusca and Brusca, 2003). As discussed in Koch (2003), the monophyly of Myriapoda is potentially supported by a considerable number of morphological characters, but none of them can presently be considered as unambiguous. Current evidence for the monophyly of Myriapoda mainly comes from molecular analyses based on nuclear markers, such as *EF-1α*, *EF-2*, and *Pol II* (Regier and Schultz, 2001; Regier et al., 2005b; Shultz and Regier, 2000), and 18S/28S rRNA genes (Mallatt et al., 2004). Recent studies based on the mt genome have however concluded to the paraphyly of Myriapoda, with Chilopoda more closely related to Chelicerata than to Diplopoda (Delsuc et al., 2003; Negrisolo et al., 2004). Actually, this unexpected result may be explained, first, by the use of a reduced taxa sample (three Myriapods and three Chelicerates in Delsuc et al., 2003; four Myriapods, but only one Chelicerate in Negrisolo et al., 2004), and, second, by the inclusion of a very inappropriate outgroup taxa, i.e., the mollusk *Katharina*, which has a mt genome affected by an inversion of asymmetric mutational constraints (see Hassanin et al., 2005). By contrast, the present mt analyses were done with a more diversified taxa sample (including 4 Myriapods and 20 Chelicerates), and by eliminating outgroup taxa possessing a mt genome with a reverse asymmetry. The results show that Myriapoda are monophyletic, as Chilopoda and Diplopoda are robustly enclosed together (Figs. 2A and 3; $BP_B = 69$ and 81; $PP_B = 1$). Although other classes of Myriapoda, such as Pauropoda and Symphyla, need to be included in further mt analyses, these results are in perfect agreement with previous studies based on morphology and nuclear genes.

4.5.4. Relationships within Chelicerata

Three classes are generally recognized in the subphylum Chelicerata: (1) Arachnida, which includes spiders, scorpions, acarids, and their allies; and two exclusively marine groups: (2) Xiphosura (horseshoe crabs) and Pycnogonida (sea spiders) (e.g., Brusca and Brusca, 2003; Firstman, 1973; Snodgrass, 1938; Weygoldt, 1986). The monophyly of Chelicerata is supported by the analyses of nuclear markers, i.e., *EF-1α* , *EF-2*, and *Pol II* genes (Regier and Schultz, 2001; Regier et al., 2005a). By contrast, Giribet et al. (2001) concluded that pycnogonids are basal to all extant arthropods by using a "total evidence" matrix combining 303 morphological characters and eight genes. Here, mtDNA analyses place pycnogonids in the Chelicerata (Figs. 2A and 3), indicating that the presence of chelicerae/chelifores is a valid autapomorphy for this clade (Waloszek and Dunlop, 2002).

The phylogenetic affinities of pycnogonids with chelicerates remain highly debated (for review, see Dunlop and Arango, 2005). Most morphologists consider, however, that Pycnogonida are a basal offshoot of the Chelicerata, with a sister-group relationship between Arachnida and Xiphosura, together forming a taxon named Eucheliceratа (e.g., Firstman, 1973; Weygoldt, 1986; Waloszek and Dunlop, 2002). The monophyly of Eucheliceratа has been retrieved with five different nuclear genes: *EF-1α*, *EF-2*, *Pol II*, 18S and 28S rRNAs (Mallatt et al., 2004; Regier and Schultz, 2001; Regier et al., 2005a), and by cladistic analyses combining molecular and morphological information (Giribet et al., 2001). The present findings conflict with these data because *Endeis* is found as the sister-group of Acari (Fig. 2A, $BP_B = 61$; Fig. 3, $BP_B = 66$). However, the association of Pycnogonida and Acari is compatible with some of the analyses published in Giribet et al. (2002). Although unusual, this grouping could be supported by two shared morphological characters: (i) some acarid species develop through a larva bearing three pairs of limbs as do the typical protonymphon larva of pycnogonids (Borradaile et al., 1958); and (ii) a similar step by step mode of limb elongation is found in some acarids and pycnogonids (Bain, 2003). However, the first character could be a plesiomorphy of the Arthropoda (Nielsen, 2001) and the second may be convergent. In addition, an LBA artifact cannot be excluded, because both *Endeis* and Acari have long branches. As a consequence, a larger species sample and additional genes are required for confirming or not the grouping of Pycnogonida with Acari.

The class Arachnida is here represented by five orders, including Araneae (spiders), Acari (mites and ticks), Scorpiones (scorpions), Uropygi (whip-scorpions), and Amblypygi (tailless whip-scorpions). Several higher taxa, previously proposed on the basis of a cladistic analysis of morphological characters (Shultz, 1990), are here retrieved with the analyses of mtDNA sequences. Indeed, Amblypygi (*Phrynus*) and Uropygi (*Mastigoproctus*) are found sister-group (Fig. 2A, $BP_B = 80$; Fig. 3, $BP_B = 64$), which supports the monophyly of Pedipalpi, a taxon morphologically determined by the possession of raptorial pedipalps (Shultz, 1990). In addition, Pedipalpi are associated with Araneae (Fig. 2A, $BP_B = 60$; Fig. 3, $BP_B = 90$), which supports the monophyly of Tetrapulmonata, a taxon named by Shultz (1990) and diagnosed by the presence of paired book lungs occupying the second and third opisthosomal segments. Book lungs are also present in Scorpiones, but these appear to have evolved well after scorpions diverged from other chelicerates (Shultz, 1990). The monophyly of the class Arachnida is not confirmed by the present analyses of mt sequences, first, because of the placement of Pycnogonida as sister-group of Acari, and second, because Scorpiones appear as being the most basal offshoot of Chelicerata. Such a topology suggests therefore that chelicerates may have acquired a terrestrial lifestyle, several times independently. As noted by Dunlop and Webster (1999), a number of characters used to support the monophyly of Arachnida (book lungs, Malpighian tubules, absence of carapace pleural margin, and anteriorly directed mouth) could also be interpreted as convergences in response to adaptations for life on land.

Interrelationships between arachnid orders are very confusing when are considered published nuclear analyses. *EF-1α*, *EF-2*, and *Pol II* genes argue for the monophyly of Arachnida, but only two orders were represented in the analyses, i.e., Araneae and Uropygi (Regier and Schultz, 2001; Regier et al., 2005a; Shultz and Regier, 2000). By contrast, analyses of 18S and 28S rRNA genes indicate that Arachnida are paraphyletic with *Aphonopelma* (Araneae) associated with *Limulus*, rather than with *Pandinus* (Scorpiones) (Mallatt et al., 2004). By combining molecules and morphological characters, Giribet et al. (2001) also suggest that Arachnida are paraphyletic, but Xiphosura were this time found sister-group of a clade containing Scorpiones, Araneae, and Uropygi. The study of relationships within Chelicerata needs therefore to be reconsidered by integrating more taxa and more genes.

Acknowledgments

I thank Jean Deutsch for laboratory facilities, Nelly Léger for DNA sequencing, and all people that collected arthropod specimens used for the present study: Pierre Escoubas and Eric Queinnec for *Euscorpius flavicaudis*, Anne Ropiquet for *Argiope bruennichi*, and Franck Simonnet for *Endeis spinosa*. I acknowledge Gonzalo Giribet, and two anonymous reviewers for their helpful comments and suggestions. I would particularly like to thank reviewer #2 for its important contribution.

References

Anderson, D.T., 1973. Embryology and Phylogeny in Annelids and Arthropods. Pergamon, Oxford.

Averof, M., Akam, M., 1995a. Hox genes and the diversification of insect and crustacean body plans. Nature 376, 420–423.

Averof, M., Akam, M., 1995b. Insect–crustacean relationships: insights from comparative developmental and molecular studies. Philos. Trans. R. Soc. Lond. B 347, 293–303.

Bain, B., 2003. Larval types and a summary of postembryonic development within the pycnogonids. Invert. Reprod. Dev. 43, 193–222.

Black 4th, W.C., Roehrdanz, R.L., 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Mol. Biol. Evol. 15, 1772–1785.

Borradaile, L.A., Potts, F.A., Kerkut, G.A., 1958. The Invertebrata. A manual for the use of students, third ed. Cambridge University Press, Cambridge, 795 p..

Britten, R.J., 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231, 1393–1398.

Bromham, L., 2002. Molecular clocks in reptiles: life history influences rate of molecular evolution. Mol. Biol. Evol. 19, 302–309.

Bromham, L., Rambaut, A., Harvey, P.H., 1996. Determinants of rate variation in mammalian DNA sequence evolution. J. Mol. Evol. 43, 610–621.

Brusca, R.C., Brusca, G.J., 2003. Invertebrates, second ed. Sinauer Press, 936 p..

Burger, G., Gray, M.W., Lang, B.F., 2003. Mitochondrial genomes: anything goes. Trends Genet. 19, 709–716.

Cisne, J.L., 1974. Trilobites and the origin of arthropods. Science 186, 13–18.

Cook, C.E., Smith, M.L., Telford, M.J., Bastianello, A., Akam, M., 2001. Hox genes and the phylogeny of the arthropods. Curr. Biol. 11, 759–763.

Cunningham, C.W., Zhu, H., Hillis, D.M., 1998. Best-fit maximum likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution 52, 978–987.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. TREE 14, 394–398.

Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on "hexapod origins: monophyletic or paraphyletic?". Science 301, 1482d.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Dove, H., Stollewerk, A., 2003. Comparative analysis of neurogenesis in the myriapod *Glomeris marginata* (Diplopoda) suggests more similarities to chelicerates than to insects. Development 130, 2161–2171.

Dowton, M., Austin, A.D., 1995. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. J. Mol. Evol. 41, 958–965.

Dunlop, J.A., Arango, C.P., 2005. Pycnogonid affinities: a review. J. Zool. Syst. Evol. Res. 43, 8–21.

Dunlop, J.A., Webster, M., 1999. Fossil evidence, terrestrialization and arachnid phylogeny. J. Arachnol. 27, 86–93.

Edgecombe, G.D, Wilson, G.D.F., Colgan, D.J., Gray, M.R., Cassis, G., 2000. Arthropod cladistics: combined analysis of histone H3 and U2 snRNA sequences and morphology. Cladistics 16, 155–203.

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 3.6b. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Firstman, B., 1973. The relationship of the chelicerate arterial system to the evolution of the endosternite. J. Arachnol. 1, 1–54.

Friedrich, M., Tautz, D., 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature 376, 165–167.

Fryer, G., 1996. Reflections on arthropod evolution. Biol. J. Linn. Soc. 58, 1–55.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15, 871–879.

Garcia-Machado, E., Pempera, M., Dennebouy, N., Oliva-Suarez, M., Mounolou, J.C., Monnerot, M., 1999. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. J. Mol. Evol. 49, 142–149.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M., Charnov, E.L., 2001. Effects of size and temperature on metabolic rate. Science 293, 2248–2251.

Gillooly, J.F., Allen, A.P., West, G.B., Brown, J.H., 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc. Natl. Acad. Sci. USA 102, 140–145.

Giribet, G., Carranza, S., Baguna, J., Riutort, M., Ribera, C., 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. Mol. Biol. Evol. 13, 76–84.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.

Giribet, G., Edgecombe, G.D., Wheeler, W.C., Babbitt, C., 2002. Phylogeny and systematic position of Opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. Cladistics 18, 5–70.

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

Harzsch, S., Walossek, D., 2001. Neurogenesis in the developing visual system of the branchiopod crustacean *Triops longicaudatus* (LeConte, 1846): corresponding patterns of compound-eye formation in Crustacea and Insecta?. Dev. Genes Evol. 211 37–43.

Hassanin, A., Lecointre, G., Tillier, S., 1998. The 'evolutionary signal' of homoplasy in protein-coding gene sequences and its phylogenetic

consequences for weighting in phylogeny. Comptes Rendus de l'Académie des Sciences, série III 321, 611–620.
Hassanin, A., Léger, N., Deutsch, J., 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inference. Syst. Biol. 54, 277–298.
Holton, T.A., Graham, M.W., 1990. A simple and efficient method for direct cloning of PCR products using dT-tailed vectors. Nucleic Acids Res. 19, 1156.
Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.
Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.
Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.
Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J., Kim, W., 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413, 154–157.
Kadner, D., Stollewerk, A., 2004. Neurogenesis in the chilopod *Lithobius forficatus* suggests more similarities to chelicerates than to insects. Dev. Genes Evol. 214, 367–379.
Knight, R.D., Freeland, S.J., Landweber, L.F., 2001. Rewiring the keyboard: evolvability of the genetic code. Nature Rev. 2, 49–58.
Koch, M., 2003. Monophyly of the Myriapoda. Reliability of current arguments. Afr. Invert. 44, 137–153.
Kohne, D.E., 1970. Evolution of higher organism DNA. Q. Rev. Biophys. 33, 327–375.
Kraus, O., 2001. "Myriapoda" and the ancestry of the Hexapoda. Ann. Soc. Entomol. Fr. (NS) 37, 105–127.
Laird, C.D., McConaughy, B.L., McCarthy, B.J., 1969. Rate of fixation of nucleotide substitutions in evolution. Nature 224, 149–154.
Lavrov, D.V., Brown, W.M., Boore, J.L., 2004. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. Proc. R. Soc. Lond. B Biol. Sci. 271, 537–544.
Li, W.-H., 1997. Molecular Evolution. Sinauer Associates, Sunderland, 487 p.
Lobry, J.R., 1995. Properties of a general model of DNA evolution under no-strand-bias conditions. J. Mol. Evol. 40, 326–330.
Machida, R.J, Miya, M.U., Nishida, M., Nishida, S., 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). Mar. Biotechnol. 4, 406–417.
Mallatt, J.M., Garey, J.R., Shultz, J.W., 2004. Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. Mol. Phylogenet. Evol. 31, 178–191.
Manton, SM., 1977. The Arthropods. Habitats, Functional Morphology and Evolution. Clarendon Press, Oxford.
Mardulyn, P., Termonia, A., Milinkovitch, M.C., 2003. Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera: Chrysomelidae): a hierarchical analysis of nucleotide sequence variation. J. Mol. Evol. 56, 38–45.
Martin, A.P., Palumbi, S.R., 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90, 4087–4091.
Martin, P., Kaygorodova, I., Sherbakov, D.Y., Verheyen, E., 2000. Rapidly evolving lineages impede the resolution of phylogenetic relationships among Clitellata (Annelida). Mol. Phylogenet. Evol. 15, 355–368.
Min, G.S., Kim, S.H., Kim, W., 1998. Molecular phylogeny of arthropods and their relatives: polyphyletic origin of arthropodization. Mol. Cells 8, 75–83.
Mooers, A.O., Harvey, P.H., 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. Mol. Phylogenet. Evol. 3, 344–350.
Nardi, F., Carapelli, A., Fanciulli, P.P., Dallai, R., Frati, F., 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. Mol. Biol. Evol. 18, 1293–1304.
Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003. Hexapod origins: monophyletic or paraphyletic? Science 299, 1887–1889.
Navajas, M., Le Conte, Y., Solignac, M., Cros-Arteil, S., Cornuet, J.M., 2002. The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). Mol. Biol. Evol. 19, 2313–2317.
Naylor, G.J.P., Brown, W.M., 1997. Structural biology and phylogenetic estimation. Nature 388, 527–528.
Negrisolo, E., Minelli, A., Valle, G., 2004. The mitochondrial genome of the house centipede scutigera and the monophyly versus paraphyly of myriapods. Mol. Biol. Evol. 21, 770–780.
Nielsen, C., 2001. Animal Evolution: Interrelationships of the Living Phyla, second ed. Oxford Univ. Press, Oxford, 563 p.
Nilsson, D.-E., Osorio, D., 1998. Homology and parallelism in arthropod sensory processing. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Systematics Association Special Volume Series 55. Chapman and Hall, London, pp. 333–347.
Ohta, T., 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23, 263–286.
Page, R.D., Lee, P.L., Becher, S.A., Griffiths, R., Clayton, D.H., 1998. A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. Mol. Phylogenet. Evol. 9, 276–293.
Perna, N.T., Kocher, T.D., 1995. Unequal base frequencies and the estimation of substitutional rates. Mol. Biol. Evol. 12, 359–361.
Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Posada, D., Buckley, T., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Regier, J.C, Shultz, J.W., 2001. Elongation factor-2: a useful gene for arthropod phylogenetics. Mol. Phylogenet. Evol. 20, 136–148.
Regier, J.C., Shultz, J.W., Kambic, R.E., 2005a. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc. R. Soc. Lond. B Biol. Sci. 272, 395–401.
Regier, J.C., Wilson, H.M., Shultz, J.W., 2005b. Phylogenetic analysis of Myriapoda using three nuclear protein-coding genes. Mol. Phylogenet. Evol. 34, 147–158.
Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957–966.
Rosenberg, M.S., Kumar, S., 2003. Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. Mol. Biol. Evol. 20, 610–621.
Schram, F., 1986. Crustacea. Oxford University Press, New York, Oxford.
Shultz, J.W., 1990. Evolutionary morphology and phylogeny of Arachnida. Cladistics 6, 1–38.
Shultz, J.W., Regier, J.C., 2000. Phylogenetic analysis of arthropods using two nuclear protein-encoding gene supports a crustacean + hexapod clade. Proc. R. Soc. Lond. B 267, 1011–1019.
Snodgrass, R.E., 1938. Evolution of the Annelida, Onychophora and Arthropoda. Smithson. Misc. Collect. 97, 1–159.
Strausfeld, N.J., 1998. Crusteacean–insect relationships: the use of brain characters to derive phylogeny amongst segmented invertebrates. Brain Behav. Evol. 52, 186–206.
Sueoka, N., 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. 40, 318–325.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.
Taanman, J.W., 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta 1410, 103–123.

Tamura, K., Kumar, S., 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol. Biol. Evol. 19, 1727–1736.

Tanaka, M., Ozawa, T., 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics 22, 327–335.

Thao, M.L., Baumann, L., Baumann, P., 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol. Biol. 4, 25.

Turbeville, J.M., Pfeifer, D.M., Field, K.G., Raff, R.A., 1991. The phylogenetic status of arthropods, as inferred from 18S rRNA sequences. Mol. Biol. Evol. 8, 669–686.

Walossek, D., Müller, K.J., 1998. Cambrian "Orsten"-type arthropods and the phylogeny of Crustacea. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Systematics Association Special Volume Series 55. Chapman and Hall, London, pp. 139–153.

Waloszek, D., Dunlop, J.A., 2002. A larval sea spider (Arthropoda: Pycnogonida) from the Upper Cambrian 'Orsten' of Sweden, and the phylogenetic position of pycnogonids. Palaeontology 45, 421–446.

Weygoldt, P., 1986. Arthropod interrelationships—the phylogenetic-systematic approach. Z. Zool. Syst. Evol. 24, 19–35.

Wheeler, W.C., Cartwright, P., Hayashi, C.Y., 1993. Arthropod phylogeny: a combined approach. Cladistics 9, 1–39.

Whitington, P.M., Bacon, J.P., 1998. The organization and development of the arthropod ventral nerve cord: insights into arthropod relationships. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Systematics Association Special Volume Series 55. Chapman and Hall, London, pp. 349–367.

Wilcox, T.P., Garcia de Leon, F.J., Hendrickson, D.A., Hillis, D.M., 2004. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. Mol. Phylogenet. Evol. 31, 1101–1113.

Wills, M.A., Briggs, D.E.G., Fortey, R.A., Wilkinson, M., Sneath, P.H.A., 1998. An arthropod phylogeny based on fossil and recent taxa. In: Edgecombe, G.D. (Ed.), Arthropod Fossils and Phylogeny. Columbia University Press, New York, pp. 33–106.

Wilson, K., Cahill, V., Ballment, E., Benzie, J., 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods. Mol. Biol. Evol. 17, 863–874.

Winnepenninckx, B., Backeljau, T., Dewachter, R., 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet. 9, 407.

Wu, C.I., Li, W.H., 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82, 1741–1745.

Wu, C.-I., Maeda, N., 1987. Inequality in mutation rates of the two strands of DNA. Nature 327, 169–170.

Yang, Z., 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39, 105–111.

Yokobori, S., Suzuki, T., Watanabe, K., 2001. Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. J. Mol. Evol. 53, 314–326.