ISSN 0026-8933, Molecular Biology, 2009, Vol. 43, No. 5, pp. 804–818. © Pleiades Publishing, Inc., 2009.
Original Russian Text © V.V. Aleshin, K.V. Mikhailov, A.V. Konstantinova, M.A. Nikitin, L.Yu. Rusin, D.A. Buinova, O.S. Kedrova, N.B. Petrov, 2009, published in Molekulyarnaya Biologiya, 2009, Vol. 43, No. 5, pp. 866–881.

UDC 575.852'112:595.2

To the cherished memory of A.S. Antonov

On the Phylogenetic Position of Insects in the Pancrustacea Clade

V. V. Aleshin[a], K. V. Mikhailov[b], A. V. Konstantinova[a], M. A. Nikitin[a], L. Yu. Rusin[c, d], D. A. Buinova[e], O. S. Kedrova[d], and N. B. Petrov[a]

[a] *Belozersky Institute of Physicochemical Biology, Moscow State University, Moscow, 119991 Russia; e-mail: Aleshin@genebee.msu.su*

[b] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119991 Russia*

[c] *Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia*

[d] *Biological Faculty, Moscow State University, Moscow, 119991 Russia*

[e] *Russian Scientific Center of Roentgenoradiology, 117997 Russia*

Received February 4, 2009
Accepted for publication March 14, 2009

Abstract—The current views on the phylogeny of arthropods are at odds with the traditional system, which recognizes four independent arthropod classes: Chelicerata, Crustacea, Myriapoda, and Insecta. There is compelling evidence that insects comprise a monophyletic lineage with Crustacea within a larger clade named Pancrustacea, or Tetraconata. However, which crustacean group is the closest living relative of insects is still an open question. In recent phylogenetic trees constructed on the basis of large gene sequence data insects are placed together with primitive crustaceans, the Branchiopoda. This topology is often suspected to be a result of the long branch attraction artifact. We analyzed concatenated data on 77 ribosomal proteins, elongation factor 1A (EF1A), initiation factor 5A (eIF5A), and several other nuclear and mitochondrial proteins. Analyses of nuclear genes confirm the monophyly of Hexapoda, the clade uniting entognath and ectognath insects. The hypothesis of the monophyly of Hexapoda and Branchiopoda is supported in the majority of analyses. The Maxillopoda, another clade of Entomostraca, occupies a sister position to the Hexapoda + Branchiopoda group. Higher crustaceans, the Malacostraca, in most analyses appear a more basal lineage within the Pancrustacea. We report molecular synapomorphies in low homoplastic regions, which support the clade Hexapoda + Branchiopoda + Maxillopoda and the monophyletic Malacostraca including Phyllocarida. Thus, the common origin of Hexapoda and Branchiopoda and their position within Entomostraca are suggested to represent bona fide phylogenetic relationships rather than computational artifacts.

DOI: 10.1134/S0026893309050124

Key words: phylogeny, molecular evolution, cladistics, EF1A, eIF5A, RpS28e, Arthropoda, Crustacea, Insecta

Insects are among the most abundant groups of living organisms on earth. From this fact, when joking that all animals are roughly insects [1] we might not be telling lies. However, the phylogenetic relationships of insects with other arthropod taxa remain vague [2]. The traditional grouping of insects with the Myriapoda in one subphylum, Tracheata, or Atelocerata [1–3], gains no support from DNA sequence analyses, which instead favor the hypothesis of a close relationship between insects and crustaceans [4–14] – a clade named Pancrustacea [15], or Tetraconata [16]. This phylogenetic view conforms with a number of common features in anatomy and nervous system development [7, 13–22], which synapomorphic nature, however, is yet questionable [23]. So far, the hypothesis of Pancrustacea has not superseded the traditional, more familiar system [24, 25].

If the Pancrustacea does exist, an immediate question [17] is which crustacean group is the closest relative of insects and, if insects diverged from the stem of the crustaceans, the monophyly of all extant crustaceans is to be verified. Until recently, neither zoological nor molecular data sufficed to comprehensively solve this problem [24]. Massive arrival of cDNA sequence data for a wide range of non-model organisms has quickly changed the situation. A phylogenetic tree obtained for the first time with a really large set of genes (133 predicted proteins, over 30000 amino acid residues) containing insects and crustaceans was published two years ago[26]. In this tree, two groups of crustaceans—Decapoda and Cla-

docera—diverge from the common stem with insects as two independent branches, thus making crustaceans paraphyletic with respect to insects. However, this remarkable fact was dropped from the discussion, as the work [26] aimed at studying systematic errors associated with analyses of large molecular data, and the phylogenetic relationships within Pancrustacea clearly fell beyond the scope. During past two years, cDNA sequence data became available for a wide range of arthropod taxa providing for the possibility to re-evaluate the phylogenetic relationships between insects and crustaceans and clarify whether the paraphyly of crustaceans is the reality or a computational artifact.

EXPERIMENTAL

Nucleotide sequences of arthropods and the outgroup were obtained from GenBank, NCBI Trace Archive (www.ncbi.nlm.nih.gov), and NEMBASE (http://www.nematodes.org) [27]. Orthologous sequences were selected with the BLAST algorithms [28], translated according to the universal genetic code and aligned using MUSCLE [29]. The alignment was manually corrected with BioEdit [30]. The nucleotide sequences were used further only for frameshift error corrections in conserved protein regions. Erroneously annotated sequences were screened off in analyses of trees constructed for each protein family with the TREEFINDER and SEMPHY programs [31, 32]. Individual gene family alignments were concatenated using SCaFoS [33], and hypervariable regions that could not be unambiguously aligned were removed. The following chimeric operational taxonomic units were formed for closely related species with missing data: Peracarida (comprising sequences of the amphipods *Gammarus pulex* and *Parhyale hawaiiensis* and isopod *Eurydice pulchra*) and Onychophora (*Epiperipatus* sp. and *Euperipatoides kanangrensis*).

At the stage of selecting data for phylogenetic analyses, we pursued two goals: first, to minimize missing data in the matrix and, second, to avoid paralogs. Ribosomal proteins proved to be nearly ideal for achieving the both. They are easy to classify into groups of orthologs, their paralogs are seldom and usually easily detectable, and the amount of ribosomal transcripts is very high in the cell and so well represented in cDNA libraries of many species. We concatenated 77 ribosomal proteins, an almost complete kit of a "typical" eukaryotic ribosome, except for the short protein L41 and ribosomal stalk proteins P0, P1, and P2. Other nuclear encoded proteins used in the study were tested for orthology by a unidirectional BLAST search over completely sequenced genomes. Aligned sequences of mitochondrial proteins were obtained from the NCBI database (www.ncbi.nlm.nih.gov/genomes/ORGANELLES/organelles.html) and OGRe (http://drake.physics.mcmaster.ca/ogre/index.shtml) [34]. The alignment was manually inspected and corrected. Highly variable regions were removed with Gblocks [35]. All predicted mitochondrial proteins, except for variable ATP6, ATP8, and NAD6, were taken in analyses.

Preliminary analysis of the concatenated set was conducted with maximum likelihood using Phyml [36]; a more refined analysis was done with MrBayes 3.1.2 [37]. The optimal matrix of amino acid substitutions was selected with ModelGenerator [38] using distributed computing [39] or using the mixed model option of MrBayes 3.1.2. All parameters, except for branch topology and lengths, were calculated independently for all proteins in the concatenated dataset (the *partition* function). Potential synapomorphies were detected in a semiautomatic mode. At the first stage, the *protpars* program of the PHYLIP package [40] was used to display predicted sequences at nodes of the tree, then all substitutions at the node were checked for changes in sites with low levels of homoplasy.

Alternative topologies (generated using TreeView [41]) were evaluated with TREE-PUZZLE 5.2 [42], and statistical approximately unbiased (AU) test [43] was carried out with CONSEL [44]. The mitochondrial protein tree was visualized with the Treecon software package [45].

RESULTS

Complete Set of Ribosomal Proteins

The analyzed set contained the aligned and concatenated amino acid sequences of 77 ribosomal proteins. After elimination of variable regions with ambiguous alignment, the concatenated alignment comprised 11 349 positions. The completeness of data for each operational taxonomic unit is shown in Fig. 1 as a percent rate that represents the proportion of filled positions to the total number of positions in the concatenated alignment. Individual ribosomal proteins considerably differ in the degree of conservation: the calculated fraction of invariant positions varies from 0.02 (RpS12, RpL30, and RpL18) to 0.19–0.21 (RpL13, RpL11, and RpL10). They also differ in the evolutionary rate of individual sites: the α-parameter of Γ-distribution in our set (containing a well represented outgroup) varies from 0.17–0.38 (RpS28, RpL40, RpS14, RpS23, RpS9, and RpS5) to 1.31–1.49 (RpL28, RpS12, RpS19, RpL24, and RpL24-like). For the majority of proteins (54 of 77), the pattern of amino acid substitutions is best described by the rtREV model [46]; for 16 proteins, by the WAG [47]; for 5 proteins, by the JTT [48]; and for two proteins, the rtREV and WAG models are approximately equally adequate. A long-term computation using a Monte Carlo procedure (ngen = 10000000), coupled

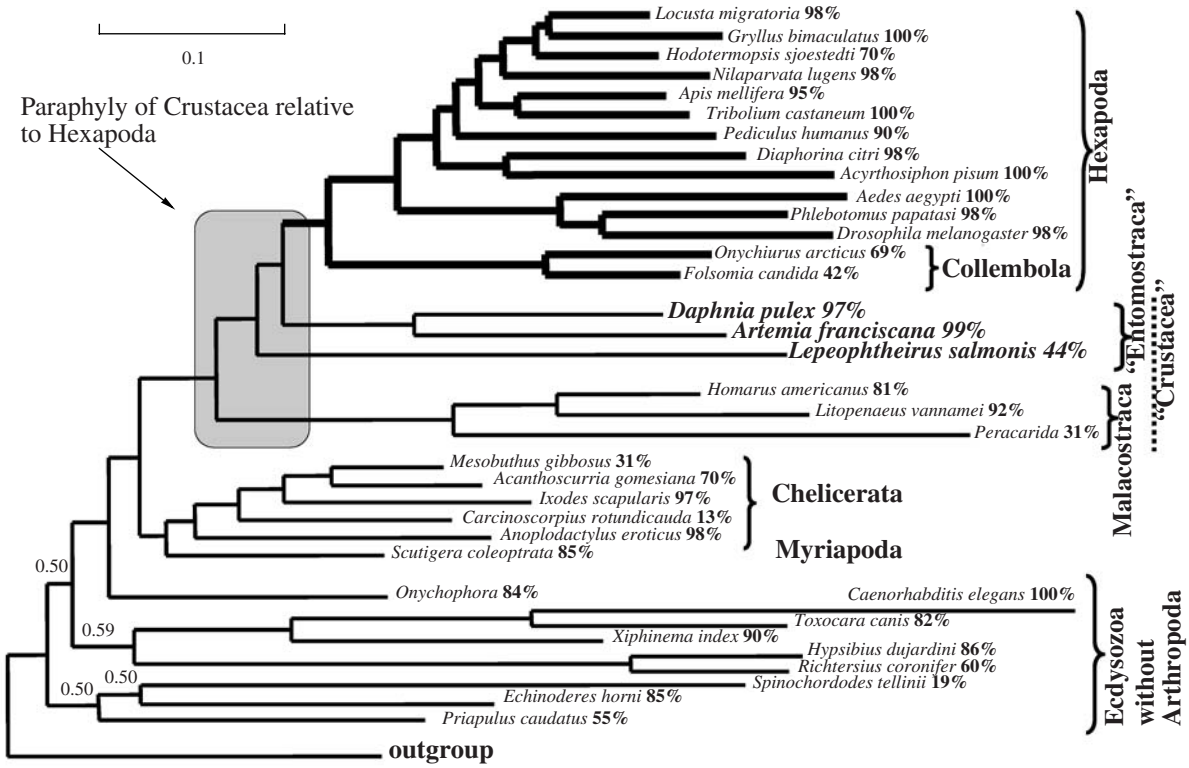
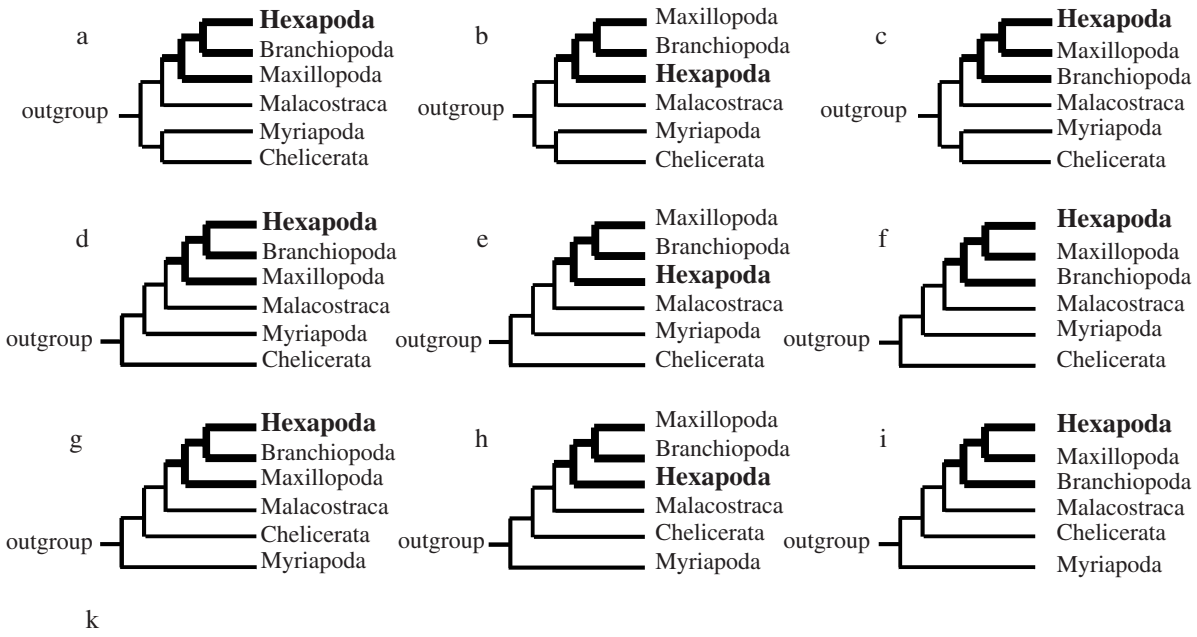


Fig. 1. Bayesian tree of concatenated sequences of 77 ribosomal proteins. The total length of the alignment after the removal of poorly alignable regions is 11349 amino acid residues. The percentage of filled positions in the alignment is given after the names of operational taxonomic units. Posterior probabilities over two independent runs are shown if not 1.0. Run parameterization: nruns = 2, nchains = 4, rates = invgamma, ngammacat = 8, aamodelpr = mixed, ngen = 10000000, burnin = 5000000, partition = by_gene, partition by_gene = 77, unlink statefreq = (all), shape = (all), pinvar =(all), and aamodel = (all). The chimeras are described in the text. The branches of insects and collembolans are in bold; the species of Entomostraca (Branchiopoda and Maxillopoda) are in semi-bold.

with the Markov chains in the MrBayes program, failed to converge in two parallel runs. In each run, we observed the stabilization of an alternative topology, leading to a posterior probability of 0.5 for some nodes of the tree. However, these differences only concern the relationship in the outgroup (Fig. 1). In the resulting tree, the insects cluster together with springtails forming a Hexapoda clade, which in turn clusters with Branchiopoda. This is not a fundamentally new result; it supports the hypothesis on a monophyly of insects and brachiopods that was proposed earlier [11, 14, 49, 50]. The branch of copepods on the tree is somewhat more distant from insects; this is the only Maxillopoda group sufficiently represented in the cDNA database. Finally, the branch of "higher crustaceans", Malacostraca, diverges at the base of Pancrustacea. The posterior probability for all groups within Pancrustacea is 1.0.

Taking the tree constructed by MrBayes (Fig. 1) as a basis, we tested 63 alternative topologies changing the positions of insects (including Collembola) and Myriapoda (represented in our set by a single species). To speed up the computation, partitioning of the alignment was not performed, and the concatenated alignment was evaluated as a single sequence. The statistical significance of the differences between topologies were calculated with Γ-distribution approximated by eight categories plus the invariant positions in the WAG model of amino acid substitutions. According to AU test [43], only six of the 63 alternative topologies overcame a 5% significance threshold and only three more according to Kishino–Hasegawa (KH) test [51]. The topology recognized as the best according to Bayesian analysis received the highest likelihood value.

The nine highest scoring phylograms are shown in Fig. 2. All of them demonstrate the tendency of insects to cluster with "lower" crustaceans, Entomostraca, while none of them recovers a group with Myriapoda. Myriapods are found outside the Pancrustacea, i.e., the taxon Tracheata, or Atelocerata, appears polyphyletic. In the best scoring tree, myriapods cluster with chelicerates; nonetheless, their phylogenetic position still remains unsettled, because there are two alternative variants that are insignificantly worse than the "best" topology, namely, when myriapods are placed



k

#	rank	item	obs	au	np	bp	pp	kh	sh	wkh	wsh
#	1	*а*	–9.1	0.861	0.616	0.612	1.000	0.806	0.994	0.806	0.999
#	2	*б*	9.1	0.312	0.118	0.115	1e–004	0.194	0.912	0.194	0.761
#	3	*в*	9.6	0.298	0.109	0.109	7e–005	0.178	0.902	0.178	0.726
#	4	*г*	19.2	0.271	0.086	0.086	5e–009	0.152	0.830	0.152	0.697
#	5	*д*	25.0	0.191	0.038	0.040	1e–011	0.117	0.771	0.117	0.644
#	6	*е*	25.2	0.172	0.034	0.035	1e–011	0.115	0.765	0.115	0.622
#	7	*ж*	35.1	0.021	0.002	0.002	6e–016	0.016	0.636	0.016	0.126
#	8	*з*	40.9	0.014	0.001	0.001	2e–018	0.016	0.574	0.016	0.151
#	9	*и*	41.4	0.017	0.001	0.001	1e–018	0.015	0.569	0.015	0.144
#	10	–	140.0	1e–007	2e–007	0	2e–061	0	0.054	0	6e–005

Fig. 2. Alternative topologies based on 77 concatenated ribosomal proteins, which are not significantly different according to the AU and/or KH statistical tests. The topologies were constructed on the basis of the Bayesian tree (Fig. 1) by reshuffling the insects and myriapods with respect to other arthropod branches. Site likelihoods were calculated with TREE-PUZZLE 5.2 under WAG + I + Γ with 8 rate categories. (a–i) Nine best phylograms ranged by the AU test,(j) the top ten lines of the CONSEL output with statistics for the 10 best topologies out of the 63 tested. In bold are common branches of insects and Entomostraca.

within the Mandibulata (Fig. 2e, f) or when they are placed at the base of Arthropoda (Fig. 2g–i). Previous studies have also failed to resolve the position of Myriapoda [10, 12, 52].

Synapomorphies in Individual Genes

(1) Protein RpS28. The consistent grouping of insects with crustaceans, more specifically Entomostraca(Figs. 1 and 2), is expected to be a result of similarities in their ribosomal protein sequences. However, the results briefed in the previous section fail to answer the question whether similar states of characters had been inherited from the remote common ancestor (are symplesiomorphic) or originated in the nearest common ancestor and are therefore true synapomorphic indicators of kinship. Yet another possibility is their independent emergence via homoplastic changes in genetic material. The logic of phylogenetic analysis ascribes different meanings to symplesiomorphic and synapomorphic similarities (for review, see [53, 54]). To assess the characters common for insects and Entomostraca from the standpoint of cladistics, we used *protpars* program to find the particular characters supporting this group. First and foremost, we were interested in the similarities at conserved positions, where any substitutions are rare, and the substitutions in the sites with varying evolutionary rate.

RpS28 is a small relatively conserved ribosomal protein. The calculated fraction of invariant sites for our set of RpS28 sequences is 0.089; this value is similar to the mean value for ribosomal proteins. However, RpS28 differs from the other ribosomal proteins by having the highest level of among site rate hetero-

geneity: the value of the α-parameter of the Γ-distribution for this protein is 0.169, which is a minimal value over the entire set. In other words, RpS28 contains both highly conserved sites (but not invariant sites, the number of which in RpS28 does not considerably deviate from the mean across all set; see above) and highly variable sites. This is a favorable combination for phylogenetic markers.

Figure 3 shows a small (according to the number of species) fragment of the RpS28 alignment, highlighting three synapomorphic substitutions in Hexapoda and Entomostraca. The established evolutionary direction from plesiomorphic Lys6, Agr27, and Asp32 to the apomorphic states Val6, Lys27, and Gly32 (the residues are numbered according to the *Drosophila melanogaster* RpS28) in the arthropod RpS28 sequences is fairly sound. Although, the sample of Entomostraca species with the known RpS28 sequences is currently confined to the sequences shown (two Branchiopoda and one Maxillopoda species); the number of insects, higher crustaceans (Malacostraca), and representatives of the outgroup (Chelicerata and nonarthropod invertebrates) with the determined sequences of this protein is large. The substitution pattern of the highlighted sites (their apomorphic or plesiomorphic states according to the expectations for the group) is maintained on a much larger sample.

(2) Protein eIF5A. The initiation factor 5A (eIF5A) is a highly conserved and vitally important protein of eukaryotes and Archaea with undetermined functions [55]. The inhibition of eIF5A gene expression with specific microRNAs decreases the overall level of translation initiation by more than one-quarter; there are data demonstrating that this factor is also involved in the programmed cell death in response to the onset of pathogens in plants and in the regulation of differentiation of animal muscle and nervous tissues. A short sequence at the C end of eIF5A is variable; however, a specific motif characteristic of Hexapoda and Entomostraca is preserved in these groups (Fig. 4). In our sample, only the eIF5A of the water bear *Richtersius coronifer* matches the consensus characteristic of Hexapoda and Entomostraca, although the sequences of several other species display a certain similarity to it. Within the phylum Arthropoda, Chelicerata and Malacostraca clearly differ from Hexapoda and Entomostraca by having a truncated C end of eIF5A. The C-terminal fragment of Chelicerata and pantopode eIF5A is truncated by four amino acid residues as compared with the corresponding sequences of Hexapoda and Entomostraca or by two to three residues as compared with the majority of other animals. The C-terminal eIF5A fragment in Malacostraca is also by one–two amino acid residues shorter than in the majority of animals, which most likely suggests that it has been truncated during the evolution of this taxon. It cannot be excluded however that these differences are connected with the evolutionary changes specific to Hexapoda and Entomostraca (autapomorphies). Altogether this character supports the group of Hexapoda and Entomostraca to the exclusion of Malacostraca but the cladistic interpretation of its state is ambiguous.

(3) Protein eEF1A. The elongation factor 1A (eEF1A) is a multifunctional, vitally important, and moderately conserved protein. It has been widely used in phylogenetics, in particular, for reconstructing the phylogenetic relationships within the arthropods [13, 56]. We have noticed that the higher crustaceans (Malacostraca) have a unique set of amino acid substitutions in the EF1A, which distinguish them not only from other arthropod orthologs, but also from all the remaining animals, as well as fungi and plants. Nonetheless, the elongation factor of Malacostraca is a clear ortholog of EF1A and does not belong to the recently described family of EF-like (EFL) proteins [57]. On the EF1A tree, Malacostraca are located at its root seemingly unrelated to animals, which is obviously an artifact of long branch attraction (Fig. 5). It is generally accepted that the most primitive of higher crustaceans are the species belonging to the superorder Phyllocarida [1, 3, 58]. *Nebalia hessleri*, a representative of this superorder, displays the autapomorphies in EF1A protein common with the remaining Malacostraca taxa (Decapoda, Mysidacea, Amphipoda, Isopoda, and Stomatopoda) and clusters with them in this tree.

A monophyly of the phyllocarids and other higher crustaceans has been earlier inferred from 18S rRNA [59] and confirmed by analyzing the protein-coding genes; here the monophyly of Malacostraca is illustratively demonstrated by the set of autapomorphic substitutions in EF1A [13, 56]. It seems as if the presence of autapomorphy in this taxon in no way assists in solving the problem of the origin of insects, which retain a plesiomorphic state with other arthropods. However, the plesiomorphies in the insect EF1A protein suggest that they could not diverge from the Malacostraca branch later that the Phyllocarida, as was hypothesized by several authors [18].

Mitochondrial Genes

The initial set prepared for the analysis comprised the amino acid sequences of ten concatenated mitochondrial proteins from 178 Arthropoda species, two Onychophora species, and one Priapulida species; Onychophora and Priapulus were taken as an outgroup for this set. The tree reconstructed by Phyml from this set(not shown) has demonstrated a considerable heterogeneity of evolutionary rates of mitochondrial proteins between species, which likely caused the grouping of the abnormally rapidly evolving

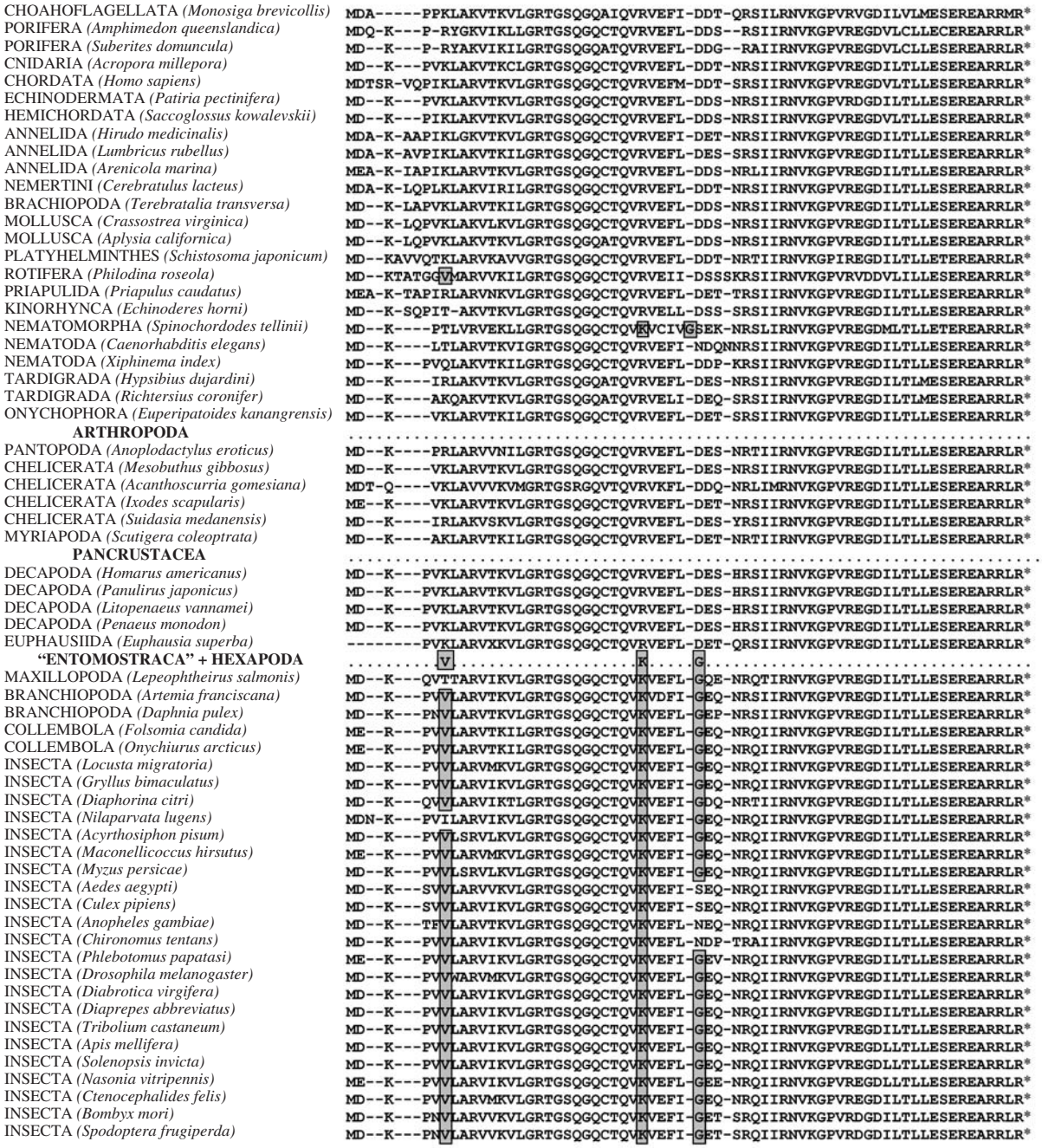


Fig. 3. A fragment of the ribosomal protein S28 alignment . Gray boxes mark the synapomorphies of Hexapoda + Entomostraca. Within the infraorder Culicomorpha successive substitutions Gly32 ⟶ Asn32 ⟶ Ser32 are shown. The residue numbering as in the RpS28 protein of *D. melanogaster.*

sequences of some insects and Chelicerata at the base of the tree. This factor also contributed to the unusual placement of other sequences, for example, of some crustaceans forming abnormally long branches.

To reduce the evolutionary rate heterogeneity of the set, we discarded the abnormally rapidly evolving sequences and difficult to align variable regions. The resulting alignment contained 82 sequences and comprised 2361 positions. Similar to the tree based on the sequences of ribosomal proteins (Fig. 1), the mitochondrial data tree constructed with MrBayes 3.1.2 (not shown) contained highly supported group Pancrustacea (all Hexapoda + all Crustacea): with an a posteriori probability of 0.95. The Myriapoda diverge

Fig. 4. A fragment of the alignment of the initiation factor 5A (eIF5A) C-terminus region. Gray boxes mark the consensus of the C-terminus in Hexapoda and Entomostraca consisting of eight amino acids $CX_5CX_3CX_7TA^{V}/_{L}{}^{D}/_{E}K$ and a stop codon; asterisks designate the stop codons.

at the base of this group, thereby supporting the Mandibulata concept, whereas in the ribosomal protein tree, they cluster with Chelicerata. A number of poorly represented groups (Cirripedia, Collembola, and Diplura) successively branch at the base of Pancrustacea between myriapods and the main group of crustaceans. The fundamental difference between the mitochondrial and nuclear protein trees is that the mitochondrial tree does not recover monophyly of Hexapoda (which implies a recurring emergence of a "Hexapod" bauplan) and does not support sister relationship between groups Branchiopoda and Hexapoda.

Low level of posterior probabilities for several groups in the mitochondrial protein tree suggests the existence of alternative topologies, which are insignificantly worse than the best Bayesian topology. Tracing the course of Bayesian analysis in two parallel runs, we found that 50% support values for some groups (the main Crustacea clade, the group Collembola + Diplura + Crustacea + Insecta *sensu stricto*, and the group Myriapoda + all Crustacea + all Hexapoda) are associated with the relocation of individual groups of sequences. In the case of Crustacea, low support value is connected with the clustering of several rapidly evolving crustacean sequences with collembolans and relocation of the ostracod sequence (also rapidly evolving) to chelicerates, which also influences the support for second group. In the case of the Mandibulata, a 50% support is a result of frequent clustering of myriapods with chelicerates in many different trees.

After the removal of another six rapidly evolving sequences, the reconstructed tree grouped myriapods and chelicerates with 64% support. The new tree also enjoyed a higher level of statistical support for the main groups (Fig. 6a). The other regions of the tree retained the same groups as the previously described tree with 82 species.

The compliance of mitochondrial data with several possible phylogenetic hypotheses was assessed by statistical tests using parametric bootstrap analysis. The topology of the tree shown in Fig. 6a was modified as described in the caption. Sixteen different topologies were compared with the initial Bayesian topology using TREE-PUZZLE and CONSEL. The results are shown in Fig. 6b. The first column in table shows the numbers of compared topology variants ordered according to the increase in their difference from the Bayesian topology. The second column lists values of corresponding topologies according to their log likelihood differences from the initial topology. The next six columns contain statistical data on the topological difference determined by various statistical tests. The value of 0.05 is a statistical threshold for assessing the significance of the difference between topologies. According to these tests, the second best topology after the initial is 7, which differs from the Bayesian topology (1) by grouping myriapods with chelicerates; the difference between these two variants is statistically insignificant. The next nine tree variants

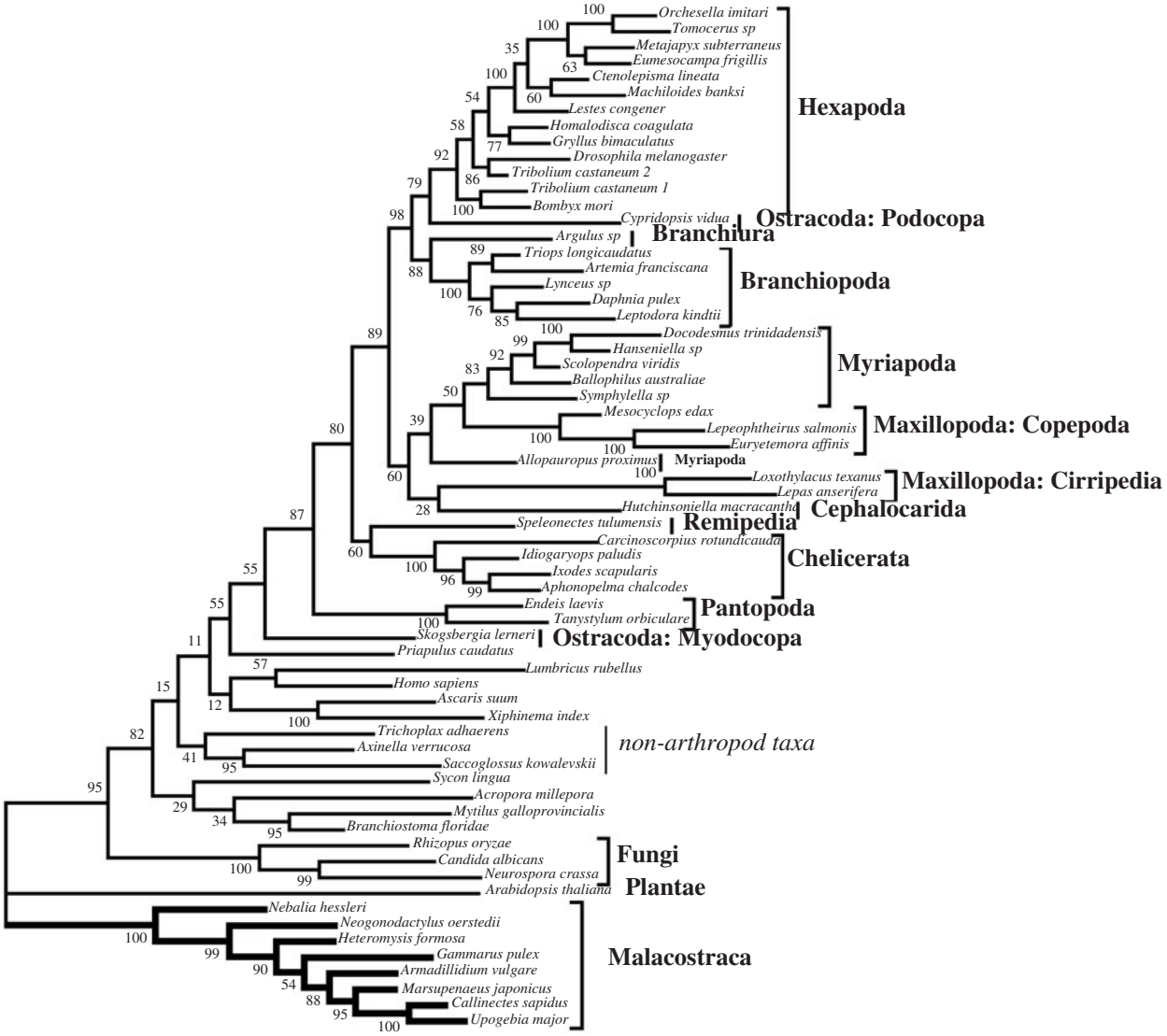


Fig. 5. Bayesian (MrBayes 3.1.2) tree of the elongation factor EF1A of various eukaryotic taxa, including the three main arthropod groups. Posterior probabilities of nodes (%) are shown. The root position of the Malacostraca is caused by the long branch attraction artifact (see text for details).

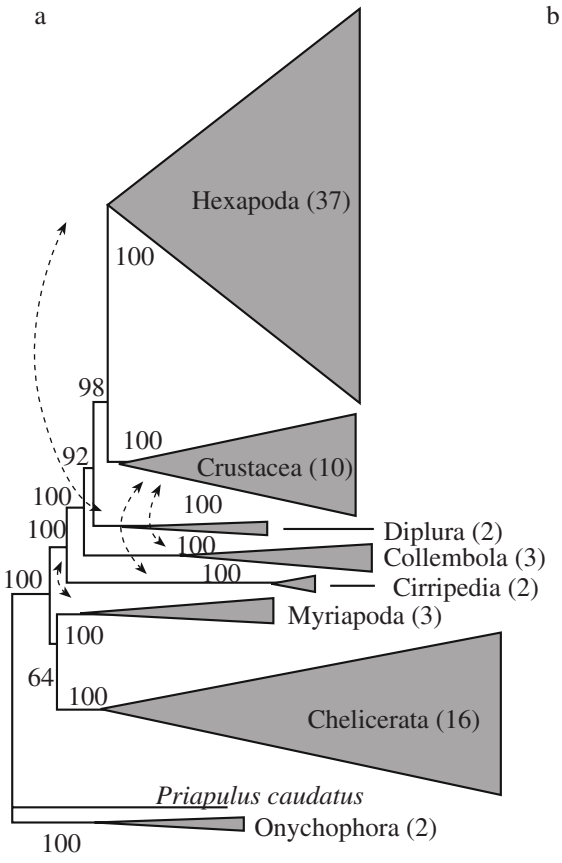
(from topology 2 to 11 excluding variant 7) are also insignificantly worse than the initial tree according to the majority of tests except for the most stringent ones (np and pp columns). All these variants have monophyletic Crustacea (Malacostraca + Branchiopoda) but varying position of Diplura and Collembola. The remaining six variants, displaying a close relationship between Hexapoda and Branchiopoda and, consequently, a paraphyly of the Crustacea independently of the position occupied by Myriapoda, differ from the best topology in a statistically significant manner according to all employed tests.

Thus, the analysis of mitochondrial protein sequences demonstrates that it is incompatible with the hypotheses for the monophyletic Hexapoda + Branchiopoda, i.e., with the hypotheses that derive hexapods from an ancestor common with branchiopods. On the other hand, the variants where all the Hexapoda and Crustacea species are united in monophyletic groups also cannot be excluded with a statistical significance; however, the variants that place Collembola at the base of Pancrustacea look statistically preferable.

DISCUSSION

The obtained results are evidently contradicting. In earlier works numerous mutually exclusive phyloge-

a



b

Topology	Values calculated by the following test						
	ΔLnL	au	np	pp	kh	c-ELW	wkh
7	–0.7	0.655+	0.271+	0.669+	0.522+	0.278+	0.522+
1	0.7	0.608+	0.203+	0.321+	0.478+	0.212+	0.478+
8	4.9	0.519+	0.220+	0.005-	0.360+	0.206+	0.360+
4	5.2	0.523+	0.138+	0.004-	0.355+	0.133+	0.336+
2	6.9	0.323+	0.075-	0.001-	0.295+	0.070+	0.254+
3	9.7	0.212+	0.020-	0.000-	0.234+	0.019+	0.182+
9	12.7	0.200+	0.041-	0.000-	0.201+	0.033+	0.199+
11	16.1	0.133+	0.015-	0.000-	0.154+	0.016+	0.154+
5	16.7	0.168+	0.017-	0.000-	0.177+	0.016+	0.138+
10	19.1	0.061+	0.003-	0.000-	0.109+	0.005-	0.100+
6	20.4	0.066+	0.003-	0.000-	0.128+	0.004-	0.098+
16	27.3	0.030-	0.003-	0.000-	0.037-	0.002-	0.031-
17	27.3	0.030-	0.003-	0.000-	0.037-	0.001-	0.031-
13	27.6	0.004-	0.001-	0.000-	0.005-	0.002-	0.005-
12	28.0	0.009-	0.001-	0.000-	0.020-	0.002-	0.005-
14	30.7	0.007-	0.001-	0.000-	0.051-	0.002-	0.009-
15	33.3	0.012-	0.001-	0.000-	0.029-	0.001-	0.019-

Fig. 6. (a) Bayesian (MrBayes 3.1.2) tree of the mitochondrial proteins from 73 arthropod species and three outgroups. Statistical support (%) for the nodes is shown. Grey triangles mark monophyletic groups, with their base proportional to the size of the group and the height – to branch lengths. In brackets are species numbers in large groups. Dashed arrows indicate the floating of groups under different taxonomic sampling. (b) Statistical significance of differences between possible topologies estimated with parametric bootstrap (TREE-PUZZLE and CONSEL programs). The tres are given in Newick format: (1) ((Cirr, (Coll, (Dipl, (Hexa, Crus)))), (Myri, Cheli)); (2) ((Cirr, (Dipl, (Coll, (Hexa, Crus)))), (Myri, Cheli)); (3) ((Cirr, (Coll, (Dipl, (Hexa, Crus)))), (Myri, Cheli)); (4) ((Cirr, ((Hexa, Crus), (Dipl, Coll))), (Myri, Cheli)); (5) (((Hexa, (Crus, Cirr)), (Dipl, Coll)), (Myri, Cheli)); (6) ((Coll, (Dipl, (Hexa, (Crus, Cirr)))), (Myri, Cheli)); (7) (Cheli, (Myri, (Cirr, (Coll, (Dipl, (Hexa, Crus)))))); (8) (Cheli, (Myri, (Coll, (Dipl, (Hexa, (Crus, Cirr)))))); (9) (Cheli, (Myri, (Dipl, ((Hexa, Coll), (Crus, Cirr))))); (10) (Cheli, (Myri, ((Coll, (Hexa, Dipl)), (Crus, Cirr)))); (11) (Cheli, (Myri, ((Hexa, (Dipl, Coll)), (Crus, Cirr)))); (12) ((Cirr, (Coll, (Dipl, ((Hexa, Bran), Mala)))), (Myri, Cheli)); (13) (((Cirr, (Coll, (Dipl, ((Hexa, Bran), Mala)))), Myri), Cheli); (14) ((Cirr, (Mala, (Bran, (Coll, (Dipl, Hexa))))), (Myri, Cheli)); (15) (Cheli, (Myri, (Cirr, (Mala, (Bran, (Coll, (Dipl, Hexa))))))); (16) (Cheli, (Myri, (Cirr, (Mala, (Bran, (Hexa, (Coll, Dipl))))))); and (17) ((Cirr, (Mala, (Bran, (Hexa, (Coll, Dipl))))), (Myri, Cheli)). The pluses stand for statistically insignificant, minuses – for statistically significant differences.

netic hypotheses were published on the evolution of arthropods and Pancrustacea in particular, partly substantiated by molecular evidence. In this situation, more reliable data or type of analysis are to be chosen. Consider several factors that might introduce technical artifacts in the tree inference and lead to inadequate test statistics.

Most importantly, there is a sensible difference in the amount of data used in analyses of nuclear genes (over 11000 positions) and mitochondrial genes (slightly more than 2000 positions after the elimination of variable blocks). The results based on more characters can be considered more reliable. From this view, we might prefer the tree shown in Fig. 1. However, larger samples should be trusted more when no bias is expected. Here we have no reason to exclude the systematic bias at all. The tree in Fig. 1 is non-ultrametric. Thus, the branches leading to Diptera are the longest among Hexapoda. Therefore, dipterans accumulated more amino acid substitutions in protein sequences compared to other hexapods. This can be expected from the accelerated rates of molecular evolution in Diptera, which was demonstrated before [60]. The Bayesian tree in Fig. 1 is not a distance tree: the MrBayes program uses a discrete method-based

optimality criterion for tree construction. However, it is still susceptible to heterogeneity in evolutionary rates and may produce artifacts. Such is the basal position of Diptera in the insect tree, with the Orthoptera and termites in the crown (Fig. 1). It looks zoologically nonsense and is likely to be associated with the disparities in branch lengths. Similarly, Malacostraca may be placed closer to the base due to longer branches compared to the Branchiopoda.

Rejecting the nuclear ribosomal proteins tree because of their high evolutionary rates in favor of the mitochondrial proteins tree is however premature, as the mitochondrial proteins display an even sharper disparity of rates but across different lineages. Branch lengths differ considerably (data not shown) and very distant arthropod taxa are artificially placed together. Interestingly, Diptera do not display accelerated evolutionary rates of mitochondrial proteins,and occupy a presumably correct position in the crown of the insect tree. On the other side, the mitochondrial proteins of paraneopteran and hymenopterans do display an abnormal acceleration. The extremely long branches leading to these groups in the trees attract, and in some trees join the Chelicerata clade uniting there with long-branched ticks.

Another property of the mitochondrial protein set, which suggests its low resolving power is the inability to recover the monophyly of Hexapoda. Moreover, the Diplura may group with insects depending on the taxonomic sampling or tree reconstruction parameters, while the grouping of Collembola and Insecta is rejected over the entire range of studied parameters. The latter is in agreement with earlier published studies of mitochondrial proteins [61, 62] but contradicts other molecular phylogenies, which strongly suggest close relationships between collembolans and insects [49, 63, 64].

An important factor in analyses of multiple genes is the adequate choice of evolutionary models and correct estimation of rate heterogeneity across sites. We found large variation of the model parameters between different ribosomal proteins. Certainly, one could average the parameter space over the concatenated alignment. However, it seems more reasonable to parameterize the analysis for partitions individually. Only few inference programs, such as MrBayes 3.1.2 with the *partition* function, allows to do so. However, individual partitions are not accounted for during the estimation of the total likelihood of a concatenated alignment during statistical tests of topologies, which results therefore should be taken with caution.

Even more important is parameter variation not across different partitions within an alignment but between different genes (i.e. taxa) within a partition. The general thinking that orthologous proteins usually carry the same function in cells of, say, beetles and spiders, with certain substitution types being equiprobable in both taxa, is true partly. More likely, some long branches actually reflect the changes in the pattern of allowed substitutions. The evolutionary model parameters in mitochondrial proteins (substitution weight matrix and stationary amino acid frequencies) were shown to be specific for mammals, arthropods, and even Pancrustacea [62, 65, 66]. For the latter two, the models were published recently and are not incorporated in MrBayes 3.1.2. Therefore, one has to choose between using optimal substitution models or using MrBayes for partition-specific Γ-parameter estimation. This problem may aggravate if smaller subtaxa within Pancrustacea will have appeared to have their own specific substitution patterns.

Another important factor is the taxonomic sampling. Today the mitochondrial genomes are sampled for many key arthropod taxa, while nuclear gene sequence data is mainly confined to ribosomal RNA and a few proteins. Expanding the taxon sampling is vital for phylogenetic analysis not only because of getting more taxa on the tree but also for the accurate estimation of various model parameters, e.g. site-specific rate variation. Because the model specification affects the entire phylogeny, increasing the taxon sampling often improves the accuracy better than increasing the amount of characters [67, 68]. The patchiness of the genome sampling of extant taxa is temporal and will smooth over in the near future, hopefully, along with the advancement of computing hardware. Nowadays, a supercomputer is unable to run a fully parameterized likelihood or Bayesian analysis of a molecular matrix severalfold larger than the one used for computing the tree in Fig. 1.

If using the correct empiric amino acid substitution model is not always easy, maybe nucleotide sequence data is a more adequate choice for phylogenetic analyses [69]? Experimental evidence suggests that the coding sequences of mitochondrial proteins produce are informative for the comparative studies of insect taxa of family rank or higher [70]. We strongly believe that taking synonymous substitutions into account is necessary in reconstructing the phylogeny of closely related organisms; however, in the case of diverged sequences, they will mostly introduce noise. Over millions of years synonymous mutations usually become saturated, and the similarity at degenerate codon positions in lineages that diverged over 100 Ma is almost entirely due to superimposed and back mutations, which are likely to decrease the reconstruction accuracy. Other factors, such as nucleotide composition bias, typical for mitochondrial DNA, may contribute to the systematic error [71]. Therefore, only amino acid sequences or nucleotide sequences of genes encoding RNAs with complex functions, such as rRNAs, are suitable for phylogenetic analyses of

distant species. The first insect fossils are known from the Carboniferous period and unlikely to have emerged earlier than the Devonian [11, 25, 72], whereas the crustacean origins are dated back to Cambrian [73]. Because of early emergence of these groups we did not use nucleotide sequence data in this study.

Despite the above stated, it is possible to formulate a substantiated phylogenetic hypothesis. Thousands of characters available from the multiple alignment contain less homoplastic regions suitable for cladistic analysis. Thus, in ribosomal protein S28 (Fig. 3) the evolution of some characters can be unambiguously polarized, and the relationship of Hexapoda and Entomostraca firmly established from synapomorphies. The question remains, however, whether these similarities are due to convergence. The situation here is not so simple. The substitutions K ⟶ V and D ⟶ G occur at a "medium" frequency, if we consider the rtREV model, which better describes the evolution of RpS28. The substitution R ⟶ K is actually among the most frequent: among the 190 possible reciprocal substitutions only three are observed more frequently in the rtREV matrix [46]. Therefore, it may seem likely that these synapomorphies are due to the chance. However, applying the rtREV frequencies to describe positions individually is premature. The changes of lysine to arginine and back across the entire proteome are frequently neutral, which explains their high frequencies in the rtREV model. Although, their frequency at position 27 in the RpS28 alignment is very low. In fact, all the three characters in RpS28 are stabilized in Hexapoda and Entomostraca, while other groups preserve their plesiomorphic states. The changes are easily described under parsimony. For example, the autapomorphies of Diptera (infraorder Culicomorpha) can be accounted for by successive substitutions Gly32 ⟶ Asn32 ⟶ Ser32 (residue numbering as in RpS28 of *Drosophila melanogaster*). This scenario is n good agreement with the known and well grounded phylogenetic relationships within the Culicomorpha [74, 75]. The selection principles suggest that position 27 in the RpS28 protein is under strong functional constrains in different arthropod groups.

Well established statistic models are not available for phylogenetic analyses of deletions/insertions, which occur in the C-terminus of the eIF5A protein. Several groups outside Hexapoda and Entomostraca share the same sequence pattern, which implies a possibility of their independent emergence in Hexapoda and Entomostraca. But the conserved nature of this motif in Hexapoda and Entomostraca suggests common selective constraints on this protein and its common ancestry.

Conserved synapomorphies of Hexapoda+Entomostraca are not confined to the RpS28 and eIF5A proteins. In nuclear proteins, we detected other characters that support the groups Hexapoda + Branchiopoda and Hexapoda + Entomostraca, in agreement with the tree in Fig. 7. Among their large numbers, only very few appear to be conserved. Topologies 6 and 7 (Fig. 6) are not supported by any conserved regions in the alignment of mitochondrial proteins, which suggests their reconstruction due to contribution from variable homoplastic regions.

Morphology does not offer unambiguous evidence of the monophyly of crustaceans with respect to insects or sistership of insects with Branchiopoda or Malacostraca, although the last scenario seems more likely [17]. In this view, the monophyly of Phyllocarida and other Malacostraca is particularly important to validate. Molecular data rejects the hypothesis of the hexapod origin from Malacostraca. The only scenario still to consider is their divergence before the split of the Phyllocarida and Eumalacostraca. The Malacostraca + Phyllocarida clade has poor diagnostic characters, mainly from the body segmentation. Phyllocarida exhibit similarities with the Branchiopoda, and both were united by Schram in one class, the Phyllopoda [76], which was proved to be paraphyletic on molecular trees. Having assumed the divergence of Hexapoda before the split of Phyllocarida and Eumalacostraca, we have equip their common ancestor with plesiomorphic features of the "Phyllopoda" (with the exception of phylloid limbs), which makes it equally likely to possess plesiomorphic characteristics of the Branchiopoda as well. This logic eliminates any contradiction between the phylogenetic hypothesis of sister Hexapoda and Branchiopoda (Figs. 1 and 7) and common knowledge on the morphological evolution in Pancrustacea. There are solid reasons to consider the filtration feeding of Entomostraca and Phyllocarida a primary feeding strategy [16]. The grasping behavior, typical for many Malacostraca [47] and insects, can thus be assumed to have emerged independently in the two groups.

A new era in phylogenetics has come with the advent of genomic data on tens and hundreds (in prospect, thousands) of genes offering tens of thousands of characters. Hopefully, advancing the analytical methods will take us on a new level of reliability of phylogenetic inference. Today, however, adding new genes does not overwhelm the old known computational artifacts, e.g caused by rate disparities across lineages. Molecular evolution cannot freeze, unlike morphological change, and its severe deviations from the molecular clock-like rates hamper the inference methods. Moreover, systematic errors tent to amplify with the increase of gene sampling. In Fig. 1 the branch lengths of Diptera and other insects are not so different. Our experience tells that modern programs implementing

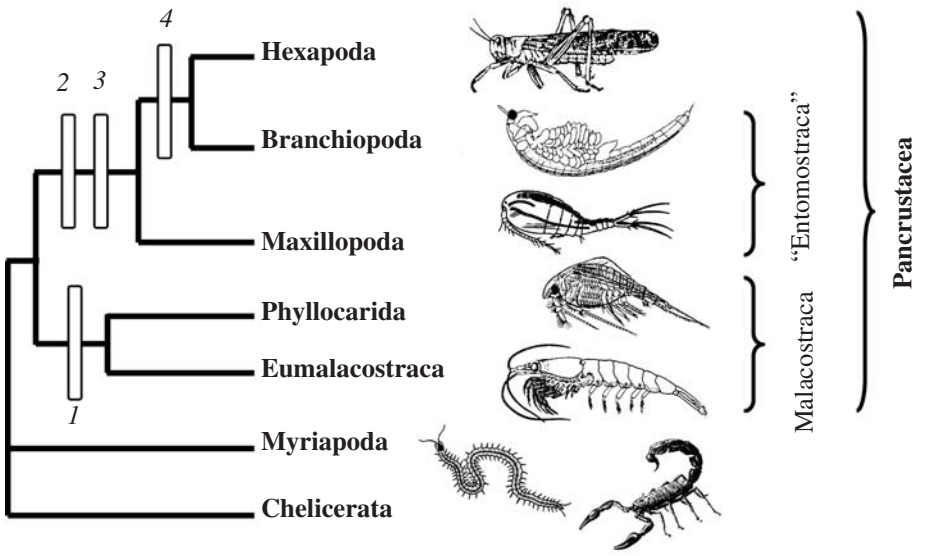


Fig. 7. The phylogenetic tree of Arthropods: (*1*) autapomorphies of the Malacostraca in elongation factor EF1A; (*2* and *3*) synapomorphies of the Hexapoda + Entomostraca in proteins RpS28 and eIF5A; (*4*) synapomorphies in moderately conserved regions of the Hexapoda and Branchiopoda proteins.

the maximum likelihood optimality criterion are able to correctly place such branches on individual gene phylogenies. However, in analyses of large sequence data the algorithm infers long branches over many partitions and estimates a longer evolutionary time to account for them, which imposes the artifact of earlier divergence of Diptera in the insect tree. The new genomic phylogenetics will require not only advanced data mining and managing software (e.g. for contigs assembly, contamination screening, orthologs identification, etc.) but will stimulate the development of new models, algorithms and their efficient implementations for building trees.

ACKNOWLEDGMENTS

In this study we used the resources of the Chebyshev Supercomputer Center of Moscow State University (http://parallel.ru/cluster) and the University of Oslo Bioportal (www.bioportal.uio.no) for tree inference, and the distributed computing platform of the National University of Ireland, Maynooth, Ireland (http://distributed.cs.nuim.ie) for selecting optimal amino acid substitution models.

The work was supported by the Russian Foundation for Basic Research (projects 08-04-01746, 09-04-01150, and 09-04-92741) and the Federal Agency for Science and Innovation of the Russian Federation (leading scientific school grant NSh-1275.2008.4).

REFERENCES

1. Ruppert E.E., Fox R.S., Barnes R.D. 2004. *Invertebrate Zoology: A Functional Evolutionary Approach*, 7th ed. Vol. 3: *Arthropods*. Belmont, CA: Brooks/Cole-Thomson Learning.
2. Kluge N.Yu. 2000. *Sovremennaya sistematika nasekomykh* (Modern Insect Taxonomy). St. Petersburg: Lan'.
3. Dogel V.A. 1981. *Zoologiya bespozvonochnykh* (Invertebrate Zoology). Ed. Polyanskii Yu.I. Moscow: Vysshaya Shkola.
4. 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.
5. Ballard J.W.O., Ballard O., Olsen G.J., Faith D.P., Odgers W.A., Rowell D.M., Atkinson P. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science*. **258**, 1345–1348.
6. Friedrich M., Tautz D. 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature*. **376**, 165–167.
7. Averof M., Akam M. 1995. Insect–crustacean relationships: Insights from comparative developmental and molecular studies. *Phil. Trans. R. Soc. London B*. **256**, 183–235.
8. Giribet G., Carranza S., Baguñà J., Riutort M., Ribera C. 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Mol. Biol. Evol*. **13**, 76–84.
9. Boore J.L., Lavrov D.V., Brown W.M. 1998. Gene translocation links insects and crustaceans. *Nature*. **392**, 667–668.

10. Aleshin V.V., Petrov N.B. 1999. Implicaciones del gen 18S ARNr en la evolución y filogenia de los Arthropoda. In: *Evolución y Filogenia de Arthropoda. Boletin de la Sociedad Entomológica Aragonesa*, no. 26. Eds. Melic A., de Haro J.J., Méndez M., Ribera I. Zaragoza: SEA, pp. 177–196.
11. Glenner H., Thomsen P.F., Hebsgaard M.B., Sorensen M.V., Willerslev E. 2006. The origin of insects. *Science*. **314**, 1883–1884.
12. Budd G.E., Telford M.J. 2009. The origin and evolution of arthropods. *Nature*. **457**, 812–817.
13. Regier J.C., Shultz J.W., Kambic R.E. 2005. Pancrustacean phylogeny: Hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc. R. Soc. London B*. **272**, 395–401.
14. Giribet G., Richter S., Edgecombe G.D., Wheeler W.C. 2005. The position of crustaceans within Arthropoda: Evidence from nine molecular loci and morphology. *Crustacean Issues*. **16**, 307–330.
15. Zrzavý J., Štys P. 1997. The basic body plan of arthropods: Insights from evolutionary morphology and developmental biology. *J. Evol. Biol*. **10**, 353–367.
16. Dohle W. 2001. Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name Tetraconata for the monophyletic unit Crustacea + Hexapoda. *Ann. Soc. Entomol. Fr. (N.S.)*. **37**, 85–103.
17. Schram F.R., Jenner R.A. 2001. The origin of Hexapoda: A crustacean perspective. *Ann. Soc. Entomol. Fr. (N.S.)*. **37**, 243–264.
18. Pavlov V.Ya. 2000. P*eriodicheskaya sistema chlenistykh* (The Periodic System of Atriculates). Moscow: VNIRO.
19. Duman-Scheel M., Patel N.H. 1999. Analysis of molecular marker expression reveals neuronal homology in distantly related arthropods. *Development*. **126**, 2327–2334.
20. Richter S. 2002. The Tetraconata concept: Hexapod–crustacean relationships and the phylogeny of Crustacea. *Org. Divers. Evol*. **2**, 217–237.
21. Harzsch S., Hafner G. 2006. Evolution of eye development in arthropods: Phylogenetic aspects. *Arthropod Struct. Dev.* **35**, 319–340.
22. Regier J.C., Shultz J.W., Ganley A.R., Hussey A., Shi D., Ball B., Zwick A., Stajich J.E., Cummings M.P., Martin J.W., Cunningham C.W. 2008. Resolving arthropod phylogeny: Exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Syst. Biol*. **57**, 920–938.
23. Stollewerk A., Chipman A.D. 2006. Neurogenesis in myriapods and chelicerates and its importance for understanding arthropod relationships. *Integr. Comp. Biol*. **46**, 195–206.
24. Backer H., Fanenbruck M., Wagele J.W. 2008. A forgotten homology supporting the monophyly of Tracheata: The subcoxa of insects and myriapods re-visited. *Zool. Anz.* **247**, 185–207.
25. Zherikhin V.V., Ponomarenko A.G., Rasnitsyn A.P., 2008. *Vvedenie v paleoentomologiyu* (Introduction to Paleoentomology). Moscow: KMK.
26. Baurain D., Brinkmann H., Philippe H. 2007. Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? *Mol. Biol. Evol*. **24**, 6–9.
27. Parkinson J., Whitton C., Schmid R., Thomson M., Blaxter M. 2004. NEMBASE: A resource for parasitic nematode ESTs. *Nucleic Acids Res*. **32**, D427–D430.
28. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389–33402.
29. Edgar R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**, 1792–1797.
30. Hall T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
31. Jobb G., von Haeseler A., Strimmer K. 2004. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol*. **4**, 18.
32. Friedman N., Ninio M., Pe'er I., Pupko T. 2002. A structural EM algorithm for phylogenetic inference. *J. Comput. Biol*. **9**, 331–353.
33. Roure B., Rodriguez-Ezpeleta N., Philippe H. 2007. SCaFoS: A tool for selection, concatenation, and fusion of sequences for phylogenomics. *BMC Evol. Biol*. **7**, Suppl. 1, S2.
34. Jameson D., Gibson A.P., Hudelot C., Higgs, P.G. 2003. OGRe: A relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res*. **31**, 202–206.
35. Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577.
36. Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. **52**, 696–704.
37. Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. **17**, 754–755.
38. Keane T.M., Creevey C.J., Pentony M.M., Naughton T.J., Mclnerney J.O. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol*. **6**, 29.
39. Keane T.M., Naughton T.J., McInerney J.O. 2007. MultiPhyl: A high-throughput phylogenomics webserver using distributed computing. *Nucleic Acids Res*. **35**, W33–W37.
40. Felsenstein J. 2005. *PHYLIP (Phylogeny Inference Package) Version 3.6. Distributed by the Author.* Seattle,

WA: Department of Genome Sciences, University of Washington.

41. Page R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci*. **12**, 357–358.
42. Schmidt H.A., Strimmer K., Vingron M., von Haeseler A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. **18**, 502–504.
43. Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol*. **51**, 492–508.
44. Shimodaira H., Hasegawa M. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics*. **17**, 1246–1247.
45. Van de Peer Y., De Wachter R. 1993. TREECON: A software package for the construction and drawing of evolutionary trees. *Comput. Appl. Biosci*. **9**, 177–182.
46. Dimmic M.W., Rest J.S., Mindell D.P., Goldstein R.A. 2002. rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J. Mol. Evol*. **55**, 65–73.
47. Whelan S., 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.
48. Jones D.T., Taylor W.R., Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci*. **8**, 275–282.
49. Mallatt J., Giribet G. 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Mol. Phylogenet. Evol*. **40**, 772–794.
50. Kashiyama K., Seki T., Numata H., Goto S.G. 2009. Molecular characterization of visual pigments in Branchiopoda and the evolution of opsins in Arthropoda. *Mol. Biol. Evol*. **26**, 299–311.
51. Kishino H., Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol*. **29**, 170–179.
52. Rota-Stabelli O., Telford M.J. 2008. A multi-criterion approach for the selection of optimal outgroups in phylogeny: Recovering some support for Mandibulata over Myriochelata using mitogenomics. *Mol. Phylogenet. Evol*. **48**, 103–111.
53. Pavlinov I.Ya., 2005. *Vvedenie v sovremennuyu filogenetiku* (Introduction to Modern Phylogenetics). Moscow: KMK.
54. Aleshin V.V., Konstantinova A.V., Mikhailov K.V., Nikitin M.A., Petrov N.B. 2007. Do we need many genes for phylogenetic inference? *Biokhimiya*. **72**, 1610–1623.
55. Zanelli C.F., Valentini S.R. 2007. Is there a role for eIF5A in translation? *Amino Acids*. **33**, 351–358.
56. Regier J.C., Shultz J.W. 1998. Molecular phylogeny of arthropods and the significance of the Cambrian 'explosion' for molecular systematics. *Am. Zool*. **38**, 918–928.
57. Keeling P.J., Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc. Natl. Acad. Sci. USA*. **101**, 15380–15385.
58. Zarenkov N.A., 1983. *Chlenistonogie. Rakoobraznye* (Arthropods and Crustaceans), part 2. Moscow: Mosk. Gos. Univ.
59. Spears T., Abele L.G. 1998. Crustacean phylogeny inferred from 18S rDNA. In: *Arthropod Relationship. The Systematics Association Special Volume Series 55*. Eds. Fortey R.A., Thomas R.H. London: Chapman & Hall, pp. 169–187.
60. Savard J., Tautz D., Lercher M.J. 2006. Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evol. Biol*. **6**, 7.
61. Nardi F., Spinsanti G., Boore J.L., Carapelli A., Dallai R., Frati F. 2003. Hexapod origins: Monophyletic or paraphyletic? *Science*. **299**, 1887–1889.
62. Carapelli A., Lio P., Nardi F., van der Wath E., Frati F. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evol. Biol*. **7**, Suppl. 2, S8.
63. Luan Y.X., Mallatt J.M., Xie R.D., Yang Y.M., Yin W.Y. 2005.The phylogenetic positions of three basal-hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. *Mol. Biol. Evol*. **22**, 1579–1592.
64. Timmermans M.J., Roelofs D., Marién J., van Straalen N.M. 2008. Revealing pancrustacean relationships: Phylogenetic analysis of ribosomal protein genes places Collembola (springtails) in a monophyletic Hexapoda and reinforces the discrepancy between mitochondrial and nuclear DNA markers. *BMC Evol. Biol*. **8**, 83.
65. Adachi J., Hasegawa M. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol*. **42**, 459–468.
66. Abascal F., Posada D., Zardoya R. 2007. MtArt: A new model of amino acid replacement for Arthropoda. *Mol. Biol. Evol*. *24*, 1–5.
67. Rosenberg M.S., Kumar S. 2003. Taxon sampling, bioinformatics, and phylogenomics. *Syst. Biol*. **52**, 119–124.
68. Heath T.A., Zwickl D.J., Kim J., Hillis D.M. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst. Biol*. **57**, 160–166.
69. Goremykin V., Moser C. 2009. Classification of the *Arabidopsis ERF* gene family based on Bayesian analysis. *Mol. Biol*. **43**, 729–734.
70. Fenn J.D., Song H., Cameron S.L., Whiting M.F. 2008. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol. Phylogenet. Evol*. **49**, 59–68.
71. DeSalle R., Freedman T., Prager E.M., Wilson A.C. 1987. Tempo and mode of sequence evolution in mito-

chondrial DNA of Hawaiian *Drosophila. J. Mol. Evol*. **26**, 157–164.

72. *Istoricheskoe razvitie klassa nasekomykh* (Historical Development of the Class of Insects). 1980. Eds. Rodendorf B.B., Rasnitsyn A.P. Moscow: Nauka. *Tr. Paleontol. Inst. Akad. Nauk SSSR*, vol. 175.
73. Walossek D., Müller K.J. 1998. Cambrian "Orsten"-type arthropods and the phylogeny of Crustacea. In: *Arthropod Relationship. The Systematics Association Special Volume Series 55*. Eds. Fortey R.A., Thomas R.H. London: Chapman & Hall, pp. 139–153.
74. Miller B.R., Crabtree M.B., Savage H.M. 1997. Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera). *Insect Mol. Biol*. **6**, 105–114.
75. Pawlowski J., Szadziewski R., Kmieciak D., Fahrni J., Bittar G. 2003. Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S RNA gene sequences. *Syst. Entomol*. **21**, 167–178.
76. Schram F.R. 1986. *Crustacea*. Oxford: Oxford Univ. Press.