melanogaster (U37541), mosquito *Anopheles quadrimaculatus* (L04272), mosquito *Anopheles gambiae* (L20934), medfly *Ceratitis capitata* (CCA242872), *Cochliomyia hominivorax* (AF260826), locust *Locusta migratoria* (X80245), honey bee *Apis mellifera* (L06178), brine shrimp *Artemia franciscana* (X69067), water flea *Daphnia pulex* (AF117817), shrimp *Penaeus monodon* (AF217843), hermit crab *Pagurus longicarpus* (AF150756), horseshoe crab *Limulus polyphemus* (AF216203), tick *Ixodes hexagonus* (AF081828), tick *Rhipicephalus sanguineus* (AF081829). For outgroup comparison, sequences were retrieved for the annelid *Lumbricus terrestris* (U24570), the mollusc *Katharina tunicata* (U09810), the nematodes *Caenorhabditis elegans* (X54252), *Ascaris suum* (X54253), *Trichinella spiralis* (AF293969) and *Onchocerca volvulus* (AF015193), and the vertebrate species *Homo sapiens* (J01415) and *Xenopus laevis* (M10217). Additional sequences were analysed for gene arrangements: *Boophilus microplus* (AF110613), *Euhadra herklotsi* (Z71696), *Cepaea nemoralis* (U23045) and *Pupa strigosa* (NC_002176).

Multiple alignments were prepared for all putative protein sequences using Clustal W[24] at default settings. Consistent with previous studies[25], preliminary analyses revealed obvious tree estimation artefacts due to extremely accelerated substitution rates or protein composition bias in the nematode species, the honeybee *Apis mellifera* and the tick species *Rhipicephalus sanguineus* and *Ixodes hexagonus*. With the exception of *Ixodes hexagonus* all of these taxa were therefore excluded from further analyses, reducing the total number of species considered to 18. Sequence alignment was repeated and inspected by eye for sufficient levels of sequence conservation, which resulted in the exclusion of the ATPase 8 gene (see Supplementary Information for single protein alignments). We used Gblocks[26] to extract regions of defined sequence conservation from the gene specific alignments and generate a single file of concatenated conserved regions. Default settings yielded the 18P2560 alignment. Modified parameter settings for generating the 18P1528 alignment were: minimum number of sequences for a conserved position: 15; maximum number of contiguous nonconserved positions: 2; minimum length of a block after gap cleaning: 5. Alignments can be retrieved from the EBI webserver (ftp://ftp.ebi.ac.uk/pub/databases/embl/align) under accession numbers ALIGN_000111 and ALIGN_000112. Maximum-likelihood mapping was carried out as described in ref. 16. Pairwise relative rate tests were carried out with the Hy-Phy program package[27]. Protein composition homogeneity test and maximum likelihood tree estimation was carried out using the TREE-PUZZLE program[28] applying the mtREV24 sequence evolution model for mitochondrial proteins[29] and a four rate approximated gamma distribution of among-site rate heterogeneity. Maximum-likelihood trees were determined by likelihood ratio tests between competing topologies. Maximum-parsimony tree reconstruction and neighbour-joining analysis with Dayhoff PAM matrix distances were performed using the respective algorithms implemented in Phylip 3.5 (ref. 30). Non-parametric bootstrapping analyses were based on 100 replicate data sets.

Received 22 February; accepted 13 July 2001.

1. Akam, M. Arthropods: developmental diversity within a (super) phylum. *Proc. Natl Acad. Sci. USA* **97,** 4438–4441 (2000).
2. Boore, J. L., Labrov, D. V. & Brown, W. M. Gene translocation links insects and crustaceans. *Nature* **392,** 667–668 (1998).
3. Shultz, J. W. & Regier, J. C. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. *Proc. R. Soc. Lond. B* **267,** 1011–1019 (2000).
4. Friedrich, M. & Tautz, D. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376,** 165–167 (1995).
5. Snodgrass, R. E. Evolution of Annelida, Onychophora and Arthropoda. *Smithson. Misc. Coll.* **138,** 1–77 (1938).
6. Cisne, J. L. Trilobites and the evolution of arthropods. *Science* **186,** 13–18 (1974).
7. Giribet, G. & Ribera, C. A review of arthropod phylogeny: new data based on ribosomal DNA sequences and direct character optimization. *Cladistics* **16,** 204–231 (2000).
8. Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L. & Brown, W. M. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376,** 163–165 (1995).
9. Curole, J. P. & Kocher, T. D. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **14,** 394–203 (1999).
10. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27,** 1767–1780 (1999).
11. Wilson, K., Cahill, V., Ballment, E. & Benzie, J. 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 (2000).
12. Garcia-Machado, E. *et al*. Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta. *J. Mol. Evol.* **49,** 142–149 (1999).
13. Aguinaldo, A. *et al*. Evidence for a clade of nematodes, arthropods and other molting animals. *Nature* **387,** 489–493 (1997).
14. Sidow, A. & Thomas, W. K. A molecular evolutionary framework for eukaryotic model organisms. *Curr. Biol.* **4,** 596–603 (1994).
15. Hausdorf, B. Early evolution of the Bilateria. *Syst. Biol.* **49,** 130–142 (2000).
16. Strimmer, K. & von Haeseler, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl Acad. Sci. USA* **94,** 6815–6819 (1997).
17. Peterson, K. J. & Eernisse, D. J. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA sequences. *Evol. Dev.* **3,** 170–205 (2001).
18. Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A. & Akam, M. *Hox* genes and the phylogeny of the arthropods. *Curr. Biol.* **11,** 759–763 (2001).
19. Kusche, K. & Burmester, T. Diplopod hemocyanin sequence and the evolution of the Myriapoda. *Mol. Biol. Evol.* **18,** 1566–1573 (2001).
20. Casares, F. & Mann, R. S. Control of antennal versus leg development in *Drosophila*. *Nature* **392,** 723–726 (1998).
21. Delle Cave, L. & Simonetta, A. M. in *The Early Evolution of Metazoa and the Significance of Problematic Taxa* (eds Simonetta, A. M. & Conway Morris, S.) 189–244 (Cambridge Univ. Press, Cambridge, 1991).
22. Telford, M. J. & Thomas, R. H. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* **95,** 10671–10675 (1998).
23. Damen, W. G., Hausdorf, M., Seyfarth, E. A. & Tautz, D. A conserved mode of head segmentation in arthropods revealed by the expression pattern of *Hox* genes in a spider. *Proc. Natl Acad. Sci. USA* **95,** 10665–10670 (1998).
24. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
25. Foster, P. G. & Hickey, D. A. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48,** 284–290 (1999).
26. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17,** 540–552 (2000).
27. Muse, S. V. & Kosakovsky Pond, S. L. *Hy-Phy 0.7 β* (North Carolina State Univ., Raleigh, 2000).
28. Strimmer, K. & von Haeseler, A. Quartet puzzling—a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13,** 964–969 (1996).
29. Adachi, J. & Hasegawa, M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* **42,** 459–468 (1996).
30. Felsenstein, J. PHYLIP (Phylogeny Inference Package). (Univ. Washington, Seattle, 1995).

Supplementary information is available from *Nature*'s World-Wide Website (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank N. Muqim for technical assistance and A. Minelli and T. Burmester for comments on the manuscript. Most computation was performed on the Biological Software Server of the Institute Pasteur Paris. This study was in part supported by a DFG grant to D.T. and a Brain Korea 21 Project grant to W.K. U.W.H. was supported by fellowships from Deutscher Akademischer Austauschdienst, Korea Science and Engineering Foundation, and the Brain Korea 21 Project (Medical Sciences, Yonsei University).

Correspondence and requests for materials should be addressed to U.W.H. (e-mail: uwhwang@knu.ac.kr) or M.F. (e-mail: mf@biology.biosci.wayne.edu).

Arthropod phylogeny based on eight molecular loci and morphology

Gonzalo Giribet*, Gregory D. Edgecombe† & Ward C. Wheeler‡

** Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA*
† Australian Museum, 6 College Street, Sydney, New South Wales 2010, Australia
‡ Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

The interrelationships of major clades within the Arthropoda remain one of the most contentious issues in systematics, which has traditionally been the domain of morphologists[1,2]. A growing body of DNA sequences and other types of molecular data has revitalized study of arthropod phylogeny[3–7] and has inspired new considerations of character evolution[8,9]. Novel hypotheses such as a crustacean–hexapod affinity[4,10–12] were based on analyses of single or few genes and limited taxon sampling, but have received recent support from mitochondrial gene order[13], and eye and brain ultrastructure and neurogenesis[14,15]. Here we assess relationships within Arthropoda based on a synthesis of all well sampled molecular loci together with a comprehensive data set of morphological, developmental, ultrastructural and gene-order characters. The molecular data include sequences of three nuclear ribosomal genes, three nuclear protein-coding genes, and two mitochondrial genes (one protein coding, one ribosomal). We devised new optimization procedures[16,17] and constructed a parallel computer cluster with 256 central processing units[18] to analyse molecular data on a scale not previously possible. The optimal 'total evidence' cladogram supports the crustacean–hexapod clade, recognizes pycnogonids as sister to other euarthropods, and indicates monophyly of Myriapoda and Mandibulata.

Based on morphological evidence, neontological[1,5,6] and palaeontological[2] hypotheses regarding deep divergences within Arthropoda differ in the monophyly of Mandibulata (arthropods with mandibles: crustaceans, myriapods and hexapods) versus

Schizoramia (arthropods with biramous appendages: crustaceans and stem-group chelicerates). Both fields have traditionally agreed that atelocerates (myriapods and hexapods) constitute a natural group. The Pancrustacea[19] hypothesis challenges a myriapod–hexapod alliance, instead resolving crustaceans as the closest relatives of hexapods. Only by combining all of the available data can the competing hypotheses be compared effectively for their ability to explain the evolution of the diverse sorts of changes that have occurred in many aspects of arthropod biology. Simultaneous analyses of morphological and molecular data addressed early conflicts of character data quality in arthropod evolution[3,5,6,20]. The investigation of a few more gene systems[12,21,22] complemented the widely used set of nuclear ribosomal genes[5,7,10]. However, to our knowledge, synthetic work combining multiple sources of molecular information, together with morphological data, has been restricted in either taxonomic or locus sampling.

We analysed sequence data of three nuclear ribosomal genes (18S ribosomal RNA, the D3 region of 28S rRNA, and the small nuclear rRNA U2), three nuclear protein-coding genes (histone H3, elongation factor-1α and the largest subunit of RNA polymerase II), and two mitochondrial loci, one protein coding (cytochrome *c* oxidase I) and one ribosomal (16S rRNA). From these data, six loci have been mostly sequenced in ours and co-workers' laboratories, and the remaining two genes (elongation factor-1α and RNA polymerase II) were obtained from GenBank (see Table 1). The sequence data comprise about 5 kilobases (kb) of nucleotide information. We chose these genes because of their dense taxonomic sampling and because they have been reported to resolve phylogenetic history at different taxonomic levels, so we expected their levels of resolution to overlap.

Non-sequence data analyses for arthropod interrelationships are few, and all rely on groundplan coding, a strategy that tends to standardize states within 'known' morphologically defined groups. This coding strategy tends to generate phylogenetic trees in agreement with classical hypotheses, but represents a biased test for relationships. On the contrary, if character states are scored only for those taxa for which real observations have been made, missing data (unknown states) may proliferate, which more accurately represents the actual knowledge of the group. Following the logic of this coding strategy, we modified, expanded and reanalysed a published

Table 1 Taxonomic categories represented in the trees, exemplar species and genes sequenced for each terminal

Peripatopsidae	*Euperipatoides/Peripatopsis*	18S	28S	H3	U2	EF	POL	COI	
Peripatidae	*Epiperipatus/Oroperipatus*	18S		H3	U2	EF	POL	COI	
Eutardigrada	*Macrobiotus/Milnesium*	18S			U2	EF	POL		
PYCNOGONIDA									
Endeis	*Endeis laevis*	18S	28S			EF	POL	COI	
Colossendeis	*Colossendeis* sp.	18S	28S			EF	POL	COI	
Ammotheidae	*Achelia/Ammothella/Tanystylum*	18S	28S	H3	U2	EF	POL	COI	16S
EUCHELICERATA									
Limulus	*Limulus polyphemus*	18S	28S	H3	U2	EF	POL	COI	16S
Carcinoscorpius	*Carcinoscorpius rotundicaudus*	18S	28S	H3	U2	EF	POL	COI	16S
Buthidae	*Androctonus/Lychas*	18S	28S	H3	U2			COI	16S
Mastigoproctus	*Mastigoproctus giganteus*	18S	28S			EF	POL	COI	16S
Mygalomorphae	*Atrax/Aphonopelma*	18S	28S	H3	U2	EF	POL	COI	16S
Opilio	*Opilio parietinus*	18S	28S					COI	16S
Nipponopsalis	*Nipponopsalis abei*	18S	28S			EF	POL		
Laniatores	*Equitius/Vonones/Dalquestia*	18S	28S	H3	U2	EF	POL	COI	
MYRIAPODA									
Scutigeridae	*Allothereua/Scutigera*	18S	28S	H3	U2	EF	POL	COI	16S
Lithobius	*L. obscurus/variegatus*	18S	28S	H3		EF	POL	COI	16S
Craterostigmus	*Craterostigmus tasmanianus*	18S	28S	H3	U2			COI	16S
Scolopendridae	*Cormocephalus/Scolopendra/Ethmostigmus*	18S	28S	H3	U2	EF	POL	COI	16S
Mecistocephalus	*Mecistocephalus* sp./*tahitiensis*	18S	28S	H3				COI	16S
Chilenophilidae	*Ribautia/Pachymerium*	18S	28S	H3	U2	EF	POL	COI	16S
Hanseniella	*Hanseniella* sp.	18S	28S	H3	U2	EF	POL	COI	16S
Scutigerella	*Scutigerella* sp.	18S	28S			EF	POL		
Pauropodinae	Pauropodinae sp.	18S	28S	H3	U2				
Polyxenidae	*Polyxenus/Unixenus*	18S	28S	H3	U2	EF	POL	COI	
Sphaerotheriidae	Cyliosomatini/*Epicyliosoma* sp.	18S		H3	U2			COI	16S
Proteroiulus	*Proteroiulus fuscus*	18S	28S			EF	POL	COI	16S
Narceus	*Narceus americanus*	18S	28S			EF	POL	COI	16S
CRUSTACEA									
Remipedia	*Speleonectes/Lasionectes*	18S	28S	H3	U2	EF	POL	COI	16S
Hutchinsoniella	*Hutchinsoniella macracantha*	18S	28S	H3		EF	POL	COI	16S
Calanoida	*Calanus/Eurytemora*	18S				EF	POL	COI	16S
Anostraca	*Artemia/Branchinella*	18S	28S	H3	U2	EF	POL	COI	16S
Triops	*T. longicaudatus/australiensis*	18S		H3	U2	EF	POL		16S
Limnadia	*Limnadia* sp./*lenticularis*	18S				EF	POL		
Daphnia	Daphnia galeata/pulex/sp.	18S			U2			COI	16S
Nebalia	*Nebalia* sp./*longicornis/hessleri*	18S	28S	H3	U2	EF	POL		
Balanidae	*Balanus/Semibalanus*	18S	28S	H3	U2	EF	POL	COI	
Stomatopoda	*Kempina mikado/Gonodactylus*	18S		H3	U2			COI	16S
Anaspides	*Anaspides tasmaniae*	18S	28S					COI	16S
Oniscidea	*Armadillidium/Trichoniscus/Australophiloscia*	18S	28S			EF	POL	COI	16S
Reptantia	*Homarus americanus/Libinia*	18S	28S	H3	U2	EF		COI	16S
HEXAPODA									
Protura	*Acerenthulus/Nipponentomon/Acerentomon*	18S	28S	H3		EF			
Arthropleona	*Podura/Archisotoma/Tomocerus*	18S	28S	H3	U2	EF	POL	COI	16S
Campodeidae	*Campodea/Eumesocampa*	18S	28S	H3		EF	POL	COI	16S
Japygidae	*Catajapyx/Metajapyx/Heterojapyx*	18S	28S			EF	POL	COI	16S
Meinertellidae	*Allomachilis/Machiloides*	18S	28S	H3	U2	EF	POL	COI	16S
Machilidae	*Dilta*/Petrobiinae/*Pedetontus*	18S	28S	H3		EF	POL	COI	16S
Tricholepidion	*Tricholepidion gertschii*	18S	28S	H3	U2				
Lepismatidae	*Thermobia/Ctenolepisma*	18S	28S			EF	POL	COI	16S
Callibaetis	*Callibaetis ferrugineus*	18S		H3	U2			COI	16S
Periplaneta	*Periplaneta americana*	18S		H3	U2	EF	POL	COI	
Locusta	*Locusta migratoria*	18S	28S	H3	U2			COI	16S
Drosophila	*Drosophila melanogaster*	18S	28S	H3	U2	EF	POL	COI	16S

The taxonomic category listed is the least inclusive for the species for which molecular data have been obtained. More details on exact GenBank numbers for each partition, fragments analysed and batch files for the analyses is provided as Supplementary Information.

morphology data set[6] in order to code the non-sequence (morphological and gene order) features of 51 terminals (2 onychophoran, 1 tardigrade and 48 arthropod lineages) for which at least four of the eight selected loci were available. A total of 303 characters was employed (see Supplementary Information).

The principle of total evidence is an important maxim in phylogenetic systematics because alternatives to using all available evidence must explicitly exclude parts of the data[23]. Excluding data *a priori* biases the outcome towards the set of characters retained. Combining all sources of molecular information with non-sequence characters allows phylogenetic hypotheses to be formulated based on a maximum number of independent data points and thus provides the greatest explanatory power.

The shortest cladograms based on the morphology and gene-order data alone recognize Mandibulata, Crustacea and Hexapoda as well supported clades (Bremer support ≥6). A sister-group relationship between Hexapoda and monophyletic Myriapoda is weakly favoured over the alternative grouping of Crustacea and Hexapoda, one step separating these hypotheses. The tree resulting from the combined analysis of all sources of data for the parameter set that minimizes overall incongruence among data partitions (using a character-based congruence metric as an optimality criterion[24]) shows monophyly of arthropods (Fig. 1). Pycnogonids are a sister group to all other euarthropods, as is also resolved based on the non-sequence data alone. Chelicerata (horseshoe crabs and arachnids) and Mandibulata are sister taxa within the

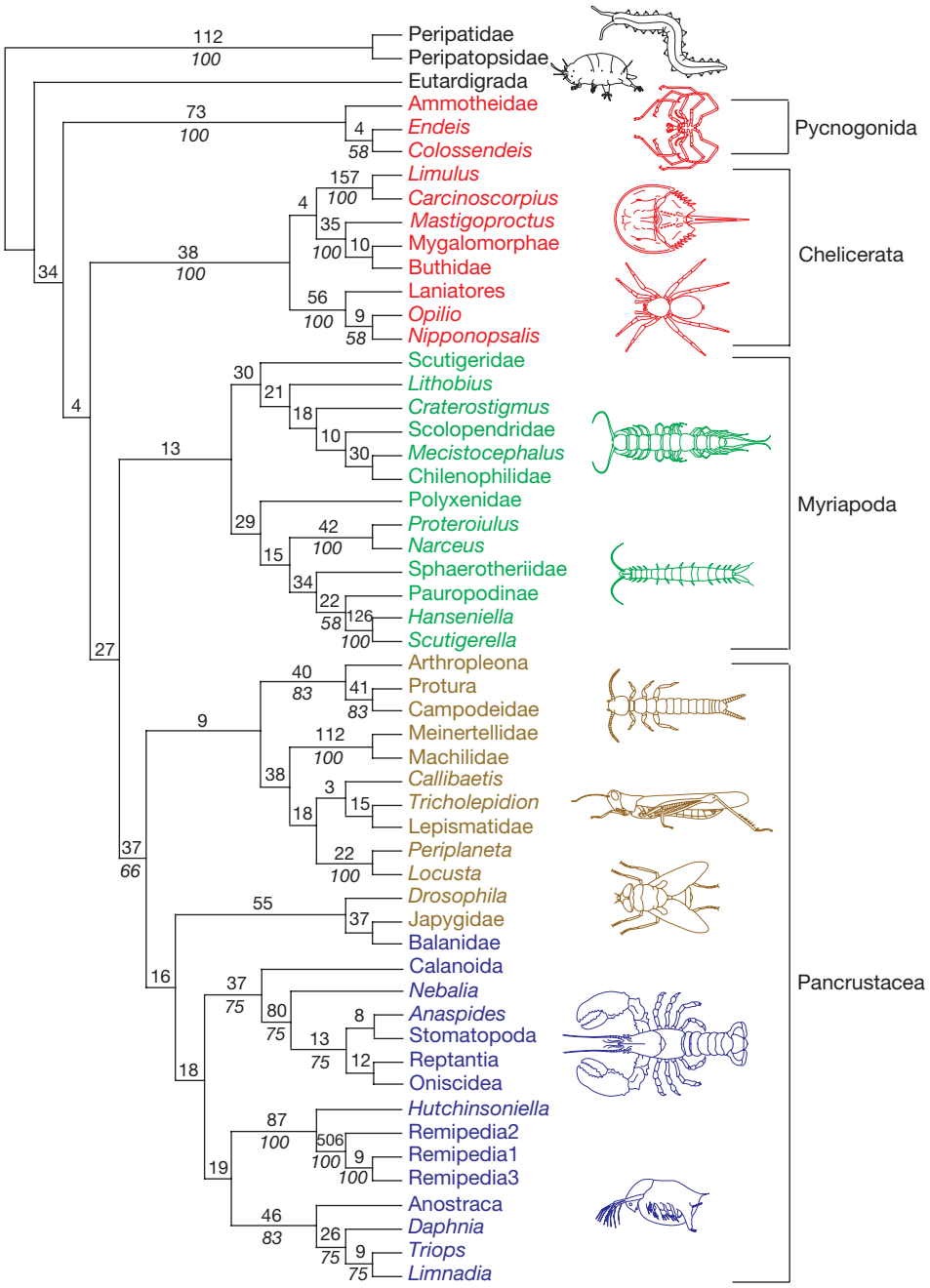

Figure 1 Phylogenetic tree of arthropod lineages based on DNA sequence data of eight loci and 303 characters of non-sequence data for the parameter set that minimizes overall incongruence among the nine partitions. The single tree, obtained when insertions and deletions equal all other genetic changes, requires 27,375 steps. Bremer support values are shown above branches while percentage of analytical parameters yielding a given topology are represented in italics. 'Chelicerates', myriapods, hexapods and crustaceans are displayed in red, green, brown and blue, respectively.

Euarthropoda. The data support neither a monophyletic Schizoramia (chelicerates + crustaceans) nor a clade composed exclusively of chelicerates and myriapods.

A monophyletic group is formed by myriapods, crustaceans and hexapods, a resolution that is consistent with the origin of mandibles being a unique event during arthropod evolution[8,9]. Relationships within Mandibulata strongly support the monophyly of Pancrustacea, although support is weak for the monophyly of myriapods. The basic division of Myriapoda into Chilopoda (centipedes) and Progoneata (millipedes, pauropods and symphylans) is retrieved by morphology as well as the combination of all data. Resolution of the internal phylogeny of the centipedes is identical between morphological and total-evidence trees, whereas relationships of the progoneates differ substantially. Inclusion of sequence data serves to unite the Symphyla and Pauropoda, which perturbs the monophyly of Diplopoda (millipedes).

Pancrustacean relationships are confusing in the optimal cladogram based on combined data, because neither crustaceans nor hexapods are strictly monophyletic. *Drosophila* forms a clade with a japygid dipluran (basal hexapod) and a barnacle, and this clade receives a moderately high Bremer support value. However, a clade resolving *Drosophila* with other insects, as expected from the morphological analysis, is found in 91% of the combined analyses (all other parameter sets explored). The aberrant behaviour of *Drosophila* is mainly due to its placement based on the gene fragments of elongation factor-1α and RNA polymerase II. Unfortunately *Drosophila* was not included in previous analyses with these gene fragments.

Crustacea is monophyletic in the non-sequence data analysis as well as in 50% of the combined analyses performed for the different parameter sets. Under the remaining six analytical parameters, Balanidae (and in one instance *Daphnia*) do not group with the remaining crustaceans. This result is more readily interpreted as anomalous behaviour of the sequence data of one organism than as evidence for crustacean paraphyly, as the original Pancrustacea hypothesis proposed. Morphological and combined analyses both identify the major groups Malacostraca and Branchiopoda, with the same basal splits (Leptostraca + Eumalacostraca and Anostraca + Phyllopoda, respectively) present in each cladogram. One of the most strongly supported groups within Crustacea, a union of cephalocarids and remipedes, is unexpected morphologically but has been detected in previous sequence analyses[25].

The analysis of deep evolutionary relationships has never been easy, especially when we attempt to make global statements about the most diverse metazoan phylum using a limited sample of creatures and characters. To our knowledge, this is the largest data set compiled for such an important group of animals and was analysed using unique cluster computer technology to provide a comprehensive hypothesis of arthropod relationships. Both quality and quantity of data were optimized to create precise and robust hypotheses. Such an approach is the first step towards comprehensive genomic comparisons and the understanding of the animal group that has dominated for the past 500 million years. □

Methods

Non-sequence data

Morphological and gene-order characters were analysed with NONA v. 2.0 (ref. 26) using a standard 1,000 replicates of random addition sequence, followed by TBR (tree bisection and reconnection) branch swapping. Multistate characters were unordered except where justified in character discussions.

Direct optimization

Phylogenetic analysis was performed by combining all of the available evidence and searching for the simplest (most parsimonious) explanation for character variation. This was accomplished with the program POY (ref. 17), which is explicitly designed to reconstruct phylogeny from diverse sources of information. Molecular data were optimized using the 'direct optimization' procedure[16], which derives cladogram costs without multiple sequence alignment. This methodology accommodates sequence length variation as transformations involving the addition, deletion and substitution of nucleotides, as opposed to the conjuring of unobservable 'gaps'. Direct optimization produces more parsimonious and more congruent results than multiple sequence alignment[27], and in a vastly accelerated time frame. POY minimizes the weighted number of evolutionary changes over the entire tree, working in a one-step fashion as opposed to the more classical two-step analyses (alignment + tree search).

Node support

Bremer support[28] was used as a measure for node support. This measure records the difference in length between the favoured tree and the shortest tree that does not contain a given group. As such, the Bremer value is a measure of how decisive the data are for a given branch of the tree.

Sensitivity analysis

To test nodal stability, 12 sets of parameters (gap/change quotients of 1, 2, 4 and transversion/transition quotients of 1, 2, 4 and ∞) were analysed for each of the ten partitions (eight loci, the combined molecular data, and the combined molecular + morphological data). Each of these 120 independent analyses was executed in parallel in the 256 processors, totalling 2 months of intense computation time using extremely effective tree search algorithms[29] and an aggressive search strategy, equivalent to 42 years of computing time if analyses had to be conducted in a single-processor machine. Parallelization was executed in groups of 32 processors, the point of maximum efficiency for marshalling jobs in our cluster. This kind of analysis explores the stability of data to analysis parameters, and therefore allows hypotheses to be formulated in a more robust way than traditional phylogenetic analysis.

The simultaneous analysis of multiple genes, morphology and mitochondrial gene order hinders application of different evolutionary models to independent partitions, and even more so when multiple parameters are explored. Therefore we opted for the simplicity of treating all partitions equally. Similarly, morphological and gene-order changes were weighted equally with insertion and deletion events.

Our 'optimal tree' is that which minimizes tree length (the most parsimonious) for all the data analysed in combination (total evidence) and for the parameter set that minimizes overall incongruence among partitions[24]. In this case overall congruence is calculated as per the incongruence length difference (ILD) test[30], and it is optimized at an insertion/deletion cost equal to all other transformations (transversions or transitions).

Search strategy

Tree search commands executed in POY included random addition sequence followed by a fast parallelized tree-building step and by SPR (subtree pruning and regrafting) and TBR branch swapping. When classical swapping algorithms did not improve tree length, the data were submitted to several rounds of tree drifting and tree fusing[29] to decrease tree length. The entire search strategy was repeated up to 1,000 times or until the results converged on the same result at least three times in independent replicates. A typical command line used for a given stepmatrix (-molecularmatrix 111) is as follows:

```
poy -parallel -jobspernode 2 -noleading -norandomizeoutgroup
 -molecularmatrix 111 -multibuild 10 -buildspr -buildtbr
-approxbuild -buildmaxtrees 1 -random 1000 -stopat 3 -minstop 10
-multirandom -sprmaxtrees 1 -tbrmaxtrees 2 -numdriftchanges 30
-driftspr -numdriftspr 10 -drifttbr -numdrifttbr 10 -Ûtchtrees
-holdmaxtrees 100 -controllers 32 -maxtrees 20 -treefuse
-fuselimit 10 -fusemingroup 5 -seed -1 ''input Ûles''
-prealigned ''input protein-coding Ûles'' > tot111.out 2>
tot111.err
```

The specific command lines and data files are provided as Supplementary Information.

Received 31 January; accepted 31 July 2001.

1. Snodgrass, R. E. Evolution of the Annelida, Onychophora and Arthropoda. *Smithson. Misc. Collect.* **97,** 1–159 (1938).
2. Wills, M. A., Briggs, D. E. G., Fortey, R. A., Wilkinson, M. & Sneath, P. H. A. *Arthropod Fossils and Phylogeny* (ed. Edgecombe, G. D.) 33–105 (Columbia Univ. Press, New York, 1998).
3. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. Arthropod phylogeny: A combined approach. *Cladistics* **9,** 1–39 (1993).
4. Regier, J. C. & Shultz, J. W. Molecular phylogeny of arthropods and the significance of the Cambrian "explosion" for molecular systematics. *Am. Zool.* **38,** 918–928 (1998).
5. Wheeler, W. C. *Arthropod Fossils and Phylogeny* (ed. Edgecombe, G. D.) 9–32 (Columbia Univ. Press, New York, 1998).
6. Edgecombe, G. D., Wilson, G. D. F., Colgan, D. J., Gray, M. R. & Cassis, G. Arthropod cladistics: Combined analysis of histone H3 and U2 snRNA sequences and morphology. *Cladistics* **16,** 155–203 (2000).
7. Giribet, G. & Ribera, C. A review of arthropod phylogeny: New data based on ribosomal DNA sequences and direct character optimization. *Cladistics* **16,** 204–231 (2000).
8. Popadić, A., Panganiban, G., Rusch, D., Shear, W. A. & Kaufman, T. C. Molecular evidence for the gnathobasic derivation of arthropod mandibles and for the appendicular origin of the labrum and other structures. *Dev. Genes Evol.* **208,** 142–150 (1998).
9. Scholtz, G., Mittmann, B. & Gerberding, M. The pattern of *Distal-less* expression in the mouthparts of crustaceans, myriapods and insects: new evidence for a gnathobasic mandible and the common origin of Mandibulata. *Int. J. Dev. Biol.* **42,** 801–810 (1998).
10. Friedrich, M. & Tautz, D. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376,** 165–167 (1995).
11. Giribet, G., Carranza, S., Baguñà, J., Riutort, M. & Ribera, C. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Mol. Biol. Evol.* **13,** 76–84 (1996).

12. Shultz, J. W. & Regier, J. C. Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade. *Proc. R. Soc. Lond. B* **267,** 1011–1019 (2000).
13. Boore, J. L., Lavrow, D. V. & Brown, W. M. Gene translocation links insects and crustaceans. *Nature* **392,** 667–668 (1998).
14. Dohle, W. 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. France* **37,** 85–103 (2001).
15. Strausfeld, N. J. Crustacean–insect relationships: The use of brain characters to derive phylogeny amongst segmented invertebrates. *Brain Behav. Evol.* **52,** 186–206 (1998).
16. Wheeler, W. C. Optimization alignment: The end of multiple sequence alignment in phylogenetics? *Cladistics* **12,** 1–9 (1996).
17. Wheeler, W. C. & Gladstein, D. POY (2000). Program and documentation available at ftp://amnh.org/pub/molecular.
18. Gee, H. Homegrown computer roots out phylogenetic networks. *Nature* **404,** 214 (2000).
19. Zrzavý, J. & Štys, P. The basic body plan of arthropods: Insights from evolutionary morphology and developmental biology. *J. Evol. Biol.* **10,** 353–367 (1997).
20. Zrzavý, J., Hypša, V. & Vlaskova, M. *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 97–107 (Chapman & Hall, London, 1998).
21. Regier, J. C. & Shultz, J. W. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. *Mol. Biol. Evol.* **14,** 902–913 (1997).
22. Colgan, D. J. *et al.* Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust. J. Zool.* **46,** 419–437 (1998).
23. Kluge, A. G. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38,** 7–25 (1989).
24. Wheeler, W. C. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst. Biol.* **44,** 321–331 (1995).
25. Spears, T. & Abele, L. G. *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 169–187 (Chapman & Hall, London, 1998).
26. Goloboff, P. A. NONA v. 2.0 (1998). Program and documentation available at ftp://unt.edu.ar/pub/parsimony.
27. Wahlberg, N. & Zimmermann, M. Pattern of phylogenetic relationships among members of the tribe Melitaeini (Lepidoptera: Nymphalidae). *Cladistics* **16,** 347–363 (2000).
28. Bremer, K. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42,** 795–803 (1988).
29. Goloboff, P. A. Analyzing large data sets in reasonable times: Solutions for composite optima. *Cladistics* **15,** 415–428 (1999).
30. Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. Constructing a significance test for incongruence. *Syst. Biol.* **44,** 570–572 (1995).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank all those who have assisted us with morphological discussions, laboratory work (especially K. Demeo), collecting specimens, and given any other form of help or advice. D. Colgan and G. Wilson have been valued collaborators. S. Thurston provided technical illustration. J. Shultz and L. Prendini shared specimens and unpublished sequence data. Funding was mainly provided by the Fundamental Biology Program of NASA.

Correspondence and requests for materials should be addressed to G.G. (e-mail: ggiribet@oeb.harvard.edu). GenBank accession codes for the new sequences are AF370781–AF370876.

Dissociation between hand motion and population vectors from neural activity in motor cortex

Stephen H. Scott, Paul L. Gribble*, Kirsten M. Graham & D. William Cabel

CIHR Group in Sensory-Motor Systems, Centre for Neuroscience Studies, Department of Anatomy and Cell Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

The population vector hypothesis was introduced almost twenty years ago to illustrate that a population vector constructed from neural activity in primary motor cortex (MI) of non-human primates could predict the direction of hand movement during reaching[1–6]. Alternative explanations for this population signal have been suggested[7,8] but could not be tested experimentally owing to movement complexity in the standard reaching model. We re-examined this issue by recording the activity of neurons in contralateral MI of monkeys while they made reaching movements with their right arms oriented in the horizontal plane—where the mechanics of limb motion are measurable and anisotropic. Here we found systematic biases between the population vector and the direction of hand movement. These errors were attributed to a non-uniform distribution of preferred directions of neurons and the non-uniformity covaried with peak joint power at the shoulder and elbow. These observations contradict the population vector hypothesis and show that non-human primates are capable of generating reaching movements to spatial targets even though population vectors based on MI activity do not point in the direction of hand motion.

Primary motor cortex (MI) has an important function in controlling visually guided limb movements, and a central problem in motor research has been to identify how neurons within MI participate in these tasks[4,9]. Several studies have shown that the activity of individual neurons is sensitive to many different parameters related to target, hand and limb movement[5,10–13]. However, it has been shown that if individual cells are represented as vectors, with direction defined by the cell's preferred direction (PD, the direction of movement in which the cell is maximally active) and magnitude defined as the cell's discharge rate for a given movement direction, the resulting vector sum (population vector) of all cell vectors is congruent with the direction of hand movement[1–3,6]. This ability to predict hand motion has supported the idea that MI may reflect a higher level representation related to movement direction[4,5]. Theoretical studies, however, have argued that neural activity in myriad coordinate frames related to sensory or motor features of the task would also produce population vectors that point in the direction of hand motion[7,8,14]. This debate on the neural representation of movement in MI is important not only for understanding its role in movement planning and control, but also for understanding the computational processes performed by other regions of the central nervous system, such as the spinal cord. Although some deviations between the population vector and hand motion have been observed[15], they have been small and difficult to interpret.

Considerable insight into human motor performance and learning has been gained from studies of planar limb movement where the arm is oriented in the horizontal plane, hand motion is generated only by flexion and extension motions at the shoulder and elbow, and the mechanics of movement can be easily estimated[16–21]. The mechanics of these planar movements are anisotropic with large variations that are dependent on movement direction[17]. We addressed whether the activity of MI neurons at the population level was influenced by these mechanical anisotropies.

We trained monkeys to make planar movements with roughly straight hand trajectories to spatial targets (Fig. 1a). We recorded the activity of neurons in the left contralateral MI of monkeys while they made reaching movements with their right hand from a central target to eight spatial targets that were located on the circumference of a circle. The activity of 214 neurons was found to be unimodally tuned to the direction of movement (62, 22 and 130 in monkeys *a*, *b* and *c*, respectively). As shown previously, cell activity was broadly tuned to the direction of movement. Figure 1b shows the activity of a typical cell in MI during the task where maximal activity occurred when the monkey moved its hand to the right and towards itself (PD = 326°).

We constructed population vectors from our cell sample and compared these vectors to the actual directions of hand motion. Population vectors during movement tended to be biased towards one of two directions: away and left, or towards and right (Fig. 2a). Thirteen of the 16 population vectors did not point in the direction of hand motion (Fig. 2b). There was no significant correlation

* Present address: Department of Psychology, The University of Western Ontario, London, Ontario N6A 5C2, Canada.