BMC Biology

Research article

The colonization of land by animals: molecular phylogeny and divergence times among arthropods

Davide Pisani[1], Laura L Poling[1,2], Maureen Lyons-Weiler[1,3] and S Blair Hedges*[1]

Address: [1]NASA Astrobiology Institute and Department of Biology, The Pennsylvania State University, University Park, PA, 16802, USA, [2]Dana-Farber Cancer Institute, Harvard Medical School 44 Binney Street, Boston MA 02115, USA and [3]Department of Pathology, University of Pittsburgh Medical School, 5230 Centre Avenue, Pittsburgh, PA, 15232 USA

Email: Davide Pisani - D.Pisani@nhm.ac.uk; Laura L Poling - Laura_Poling@student.hms.harvard.edu; Maureen Lyons-Weiler - mal26@pitt.edu; S Blair Hedges* - sbh1@psu.edu

* Corresponding author

Abstract

Background: The earliest fossil evidence of terrestrial animal activity is from the Ordovician, ~450 million years ago (Ma). However, there are earlier animal fossils, and most molecular clocks suggest a deep origin of animal phyla in the Precambrian, leaving open the possibility that animals colonized land much earlier than the Ordovician. To further investigate the time of colonization of land by animals, we sequenced two nuclear genes, glyceraldehyde-3-phosphate dehydrogenase and enolase, in representative arthropods and conducted phylogenetic and molecular clock analyses of those and other available DNA and protein sequence data. To assess the robustness of animal molecular clocks, we estimated the deuterostome-arthropod divergence using the arthropod fossil record for calibration and tunicate instead of vertebrate sequences to represent Deuterostomia. Nine nuclear and 15 mitochondrial genes were used in phylogenetic analyses and 61 genes were used in molecular clock analyses.

Results: Significant support was found for the unconventional pairing of myriapods (millipedes and centipedes) with chelicerates (spiders, scorpions, horseshoe crabs, etc.) using nuclear and mitochondrial genes. Our estimated time for the divergence of millipedes (Diplopoda) and centipedes (Chilopoda) was 442 ± 50 Ma, and the divergence of insects and crustaceans was estimated as 666 ± 58 Ma. Our results also agree with previous studies suggesting a deep divergence (~1100 − 900 Ma) for arthropods and deuterostomes, considerably predating the Cambrian Explosion seen in the animal fossil record.

Conclusions: The consistent support for a close relationship between myriapods and chelicerates, using mitochondrial and nuclear genes and different methods of analysis, suggests that this unexpected result is not an artefact of analysis. We propose the name Myriochelata for this group of animals, which includes many that immobilize prey with venom. Our molecular clock analyses using arthropod fossil calibrations support earlier studies using vertebrate calibrations in finding that deuterostomes and arthropods diverged hundreds of millions of years before the Cambrian explosion. However, our molecular time estimate for the divergence of millipedes and centipedes is close to the divergence time inferred from fossils. This suggests that arthropods may have adapted to the terrestrial environment relatively late in their evolutionary history.

Background

The terrestrial environment has been greatly altered by the actions of organisms over Earth's history. Prokaryotes were probably the first organisms to colonize land, and this occurred as early as 2.6 billion years ago [1-3]. The presence of organisms on exposed land will accelerate weathering through physical and chemical processes and may in turn affect the global atmosphere and climate [4]. Therefore, it is of interest to know when different groups of organisms colonized land to better understand their effect on the biosphere.

The earliest undisputed fossils of terrestrial plants, animals, and fungi are all from the early Palaeozoic (Ordovician and Silurian; 495 – 417 million years ago, Ma) [5-9]. However, the earliest animal fossils are known from ~600 Ma [10-12] and the earliest representatives of the "plant lineage", such as green algae [13] and red algae [14], from even earlier (1000 – 750 Ma and 1200 Ma, respectively). This raises the possibility that land was colonized by multicellular eukaryotes prior to the Ordovician. A previous molecular clock analysis addressed the question of land colonization by plants and fungi, resulting in early time estimates of about 700 and 1000 Ma, respectively [15]. However, molecular clock analyses have not addressed the colonization of land by animals.

Myriapods (centipedes, millipedes) and chelicerates (e.g., arachnids, horseshoe crabs) have figured prominently in the earliest evidence of terrestrial animals. For example, the first taxonomically identifiable body fossils of terrestrial animals are arachnids and chilopods from the late Silurian (~419 Ma) of England [6], and the oldest unambiguous evidence of sub-aerial animal activity is of arthropod (diplopod-like) trackways from ~450 Ma sediments [16]. Older arthropod traces (possibly from the latest Cambrian) have been reported from terrestrial sediments, eolian dune deposits accumulated in a sandy beach environment of southern Canada [17], although the terrestrial nature of the trace makers is uncertain. Older, marine myriapod-like fossils are known [18], but no marine remains attributable to Chilopoda or Diplopoda have been found.

Here, we use molecular clock and phylogenetic methods to place temporal constraints on the early history of arthropods and the colonization of land by animals. Our focus is on the divergence of millipedes and centipedes because they represent the most ancient living lineages of terrestrial animals (based on the fossil record) whose common ancestor presumably was terrestrial. In turn, this provides a minimum estimate for the time when land was colonized by animals. With the use of a phylogenetic framework derived from analyses of all available nuclear genes, we also estimate times of other major divergences

in the history of living arthropods, and the divergence of arthropods and deuterostomes. These additional data help to constrain a maximal time for the colonization of land by arthropods.

Results

All phylogenetic analyses resulted in significant (>95%) clustering of myriapods and chelicerates (figure 1). Pancrustacea (insects and crustaceans) was found in most analyses, the only exceptions being minimum evolution with paralinear distances, and both weighted and unweighted parsimony analyses of nuclear + mitochondrial genes, and the unweighted parsimony analysis of only nuclear genes (Table 1). However, because these exceptional analyses used models that are less complex than recommended [19], it is possible that they were influenced by substitutional biases. Support for Pancrustacea was significant in most analyses of nuclear genes and the Bayesian analysis of nuclear + mitochondrial genes. The two representative crustaceans (branchiopods and malacostracans) formed a group in the nuclear gene analysis and the Bayesian all-gene analysis. However, insects joined malacostracans in the remaining all-gene analyses. This uncertainty of relationships within Pancrustacea has been encountered previously [20] and may be related to different rates of evolution among genes.

Divergence time estimates for the millipede-centipede split ranged from 357 – 537 Ma across different molecular clock methods, with an average of 442 ± 50 Ma (Table 2). Average time estimates for the other divergences among arthropods were 475 ± 53 Ma for Xiphosura-Arachnida, 642 ± 63 Ma for Myriapoda-Chelicerata, 614 ± 23 Ma for Branchiopoda-Malacostraca, 666 ± 58 Ma for Insecta-Crustacea, and 725 ± 46 Ma for Pancrustacea-(Chelicerata-Myriapoda). As expected, the use of different calibrations resulted in different time estimates, and although the penalized likelihood method (SGG_{PL}) showed the greatest sensitivity to calibration differences, most methods yielded similar time estimates.

The use of exclusively fossil calibration points (Chilopoda-Diplopoda, Xiphosura-Arachnida) or molecular calibration points (Arthropoda-Deuterostomia) did not result in substantially different time estimates (table 2). Therefore, time estimates using the largest number of calibration points and proteins (averaged across methods) were chosen to summarize the time of divergence for each node (Table 2, Figure 2).

Using the assumption that the ancestral arthropod was aquatic (marine) and that terrestrialization is derived within arthropods, myriapods colonized land after the origin of the myriapod lineage (chelicerate-myriapod divergence; 642 ± 63 Ma) and before the millipede-centi-

(a)

(b)
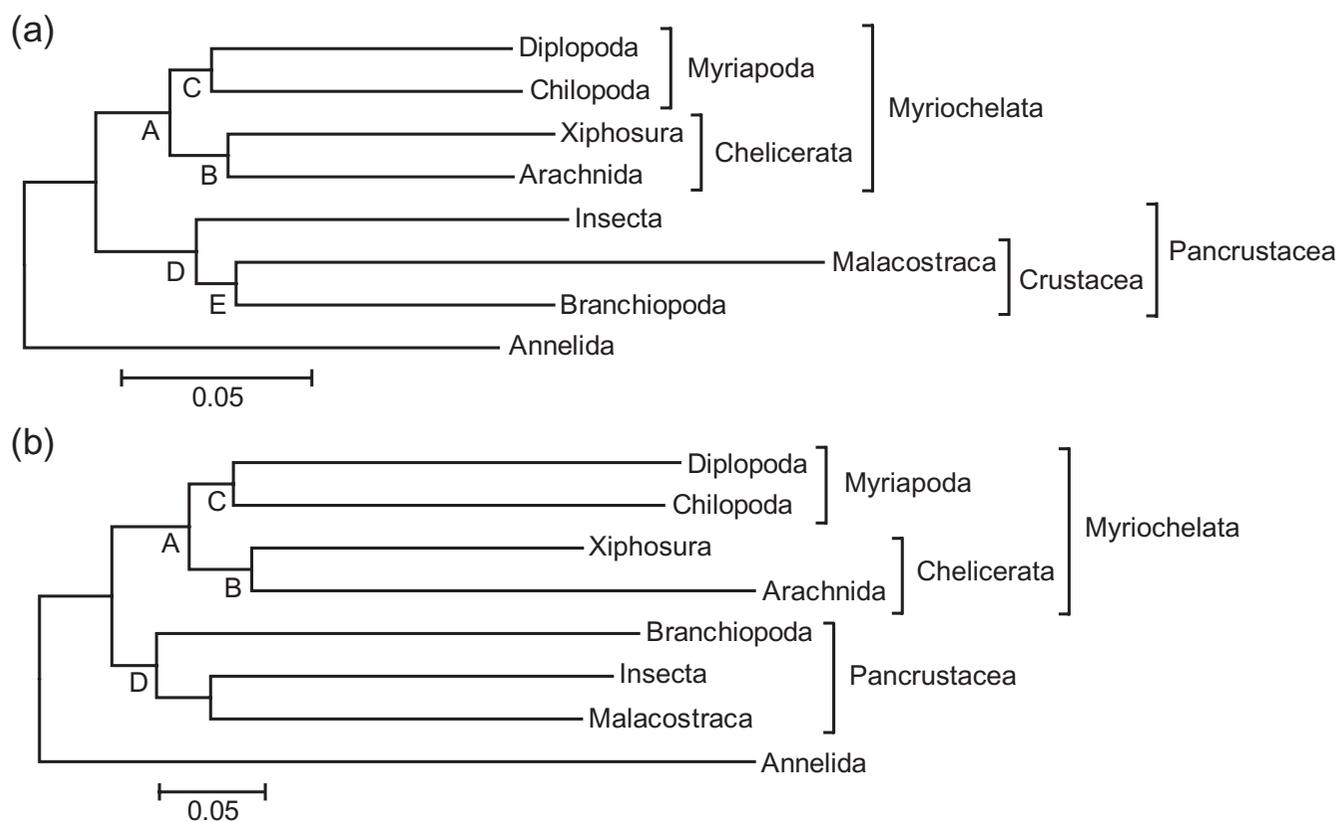
Figure 1
Phylogenetic relationships of the arthropods (a) Minimum evolution tree of the nine concatenated nuclear genes. (b) Minimum evolution tree of the 24 concatenated nuclear and mitochondrial genes. Both trees were obtained using gamma corrected Kimura 2-parameter (transversion) distances. The same (or similar) tree topologies were obtained using other methods (Table 1).

Table 1: Support values for the nodes in the phylogenetic trees (Figure 1). Numbers are posterior probabilities for Bayesian inference and bootstrap confidence values for all others methods; dashes indicate that the node was not present in the tree.

		Support values				
Data	Method of analysis	Node A	Node B	Node C	Node D	Node E
Nuclear genes	Minimum evolution (gamma, transversions)	96	98	99	100	63
	Minimum evolution (transversions)	100	100	95	100	61
	Minimum evolution (paralinear distances)	100	100	47	77	68
	Maximum likelihood	100	100	90	99	91
	Bayesian inference	100	100	100	100	100
	Maximum parsimony	100	100	64	-	86
	Weighted parsimony	100	100	92	74	86
All genes	Minimum evolution (gamma, transversions)	100	89	98	70	-
	Minimum evolution (transversions)	100	100	97	61	-
	Minimum evolution (Paralinear distances)	100	100	91	-	-
	Maximum likelihood	100	99	100	52	-
	Bayesian inference	100	100	100	100	100
	Maximum parsimony	97	80	96	-	-
	Weighted parsimony	100	100	99	-	-

Table 2: Divergence times for the major groups of arthropods. Divergence times from least squares (LS) methods are means (<30 proteins) or modes of time estimates from individual proteins. Summary times are averages across methods for comparisons that maximize the calibrations and proteins used, indicated in bold. Abbreviations: CD= Chilopoda-Diplopoda calibration (423 Ma); DA= Deuterostomia-Arthropoda calibration (993 Ma); N/A = not applicable; XA= Xiphosura-Arachnida calibration (480 Ma).

		Number of proteins		Divergence time (Ma) and standard error (if available)							
Node	Calibration point	Total	Rate constant	MGG_{LS}	MGL_{LS}	SGG_{LS}	SGL_{LS}	SGL_{DT}	SGL_{MDT}	SGL_{PL}	Summary
Chilopoda-Diplopoda	DA	5	5	433 ± 39	464 ± 68	433 ± 39	468 ± 68	389 ± 59	451 ± 45	381	442 ± 50
	XA	5	4	398 ± 79	N/A	357 ± 79	N/A	499 ± 57	474 ± 40	537	
	XA & DA	5	4	398 ± 38	425 ± 79	376 ± 38	384 ± 79	469 ± 44	469 ± 36	439	
	XA &/or DA	5	5	430 ± 43	457 ± 69	404 ± 39	426 ± 73	469 ± 44	469 ± 36	439	
Xiphosura-Arachnida	DA	5	4	521 ± 71	534 ± 44	511 ± 71	526 ± 44	372 ± 58	448 ± 40	373	475 ± 53
	CD	5	4	563 ± 92	N/A	568 ± 92	N/A	403 ± 48	418 ± 36	467	
	CD & DA	5	4	526 ± 73	534 ± 48	519 ± 73	532 ± 48	394 ± 41	432 ± 35	394	
Branchiopoda-Malacostraca	DA	9	3	625 ± 8	551 ± 19	640 ± 8	591 ± 19	670 ± 58	631 ± 37	655	614 ± 23
	XA	5	0	N/A	N/A	N/A	N/A	680 ± 87	643 ± 56	806	
	XA & DA	5	0	N/A	N/A	N/A	N/A	594 ± 50	638 ± 33	595	
	XA &/or DA	5	3	625 ± 8	551 ± 19	640 ± 8	591 ± 19	632 ± 54	638 ± 33	625	
	CD	5	0	N/A	N/A	N/A	N/A	581 ± 75	561 ± 51	717	
	XA & D	5	0	N/A	N/A	N/A	N/A	635 ± 74	608 ± 47	729	
	CD & DA	5	0	N/A	N/A	N/A	N/A	554 ± 51	616 ± 34	563	
Insecta-Crustacea	DA	47	36	765 ± 74	700 ± 72	669 ± 34	643 ± 33	616 ± 53	788 ± 22	577	666 ± 58
	XA	5	4	619 ± 132	N/A	534 ± 132	N/A	744 ± 89	802 ± 69	875	
	XA & DA	5	4	617 ± 92	646 ± 99	562 ± 92	576 ± 99	647 ± 46	763 ± 17	645	
	XA &/or DA	47	36	741 ± 87	698 ± 70	615 ± 63	609 ± 66	631 ± 49	763 ± 17	611	
	CD	5	5	650 ± 98	N/A	568 ± 98	N/A	637 ± 77	687 ± 56	779	
	XA & D	5	4	612 ± 95	N/A	573 ± 93	N/A	695 ± 75	737 ± 48	729	
	CD & DA	5	5	641 ± 69	665 ± 70	578 ± 69	593 ± 70	605 ± 48	758 ± 18	611	
Myriapoda-Chelicerata	DA	6	4	594 ± 51	664 ± 84	630 ± 51	716 ± 84	630 ± 62	664 ± 32	589	642 ± 63
	XA	5	2	674 ± 110	N/A	605 ± 110	N/A	751 ± 67	728 ± 49	847	
	XA & DA	5	2	582 ± 57	668 ± 166	576 ± 57	594 ± 166	675 ± 32	689 ± 26	669	
	XA &/or DA	6	4	602 ± 51	664 ± 80	603 ± 54	655 ± 125	652 ± 47	689 ± 26	629	
	CD	5	3	560 ± 18	N/A	544 ± 18	N/A	640 ± 56	647 ± 40	757	
	XA & D	5	2	594 ± 29	N/A	638 ± 29	N/A	700 ± 47	687 ± 34	772	
	CD & DA	5	3	608 ± 49	668 ± 83	558 ± 49	576 ± 83	618 ± 34	642 ± 25	621	
(Myriapoda-Chelicerata)-Pancrustacea	DA	24	22	747 ± 40	731 ± 41	717 ± 40	692 ± 41	690 ± 61	727 ± 20	602	725 ± 46
	XA	5	4	783 ± 139	N/A	703 ± 139	N/A	892 ± 97	812 ± 59	1051	
	XA & DA	5	4	784 ± 65	826 ± 111	740 ± 65	758 ± 111	768 ± 35	738 ± 19	764	
	XA &/or DA	24	22	745 ± 41	729 ± 41	728 ± 52	725 ± 76	729 ± 48	738 ± 19	683	
	CD	5	5	791 ± 79	N/A	727 ± 79	N/A	767 ± 82	734 ± 56	940	
	XA & D	5	4	747 ± 86	N/A	754 ± 86	N/A	836 ± 77	769 ± 51	956	
	CD & DA	5	5	784 ± 42	815 ± 66	761 ± 42	780 ± 66	727 ± 38	721 ± 19	728	
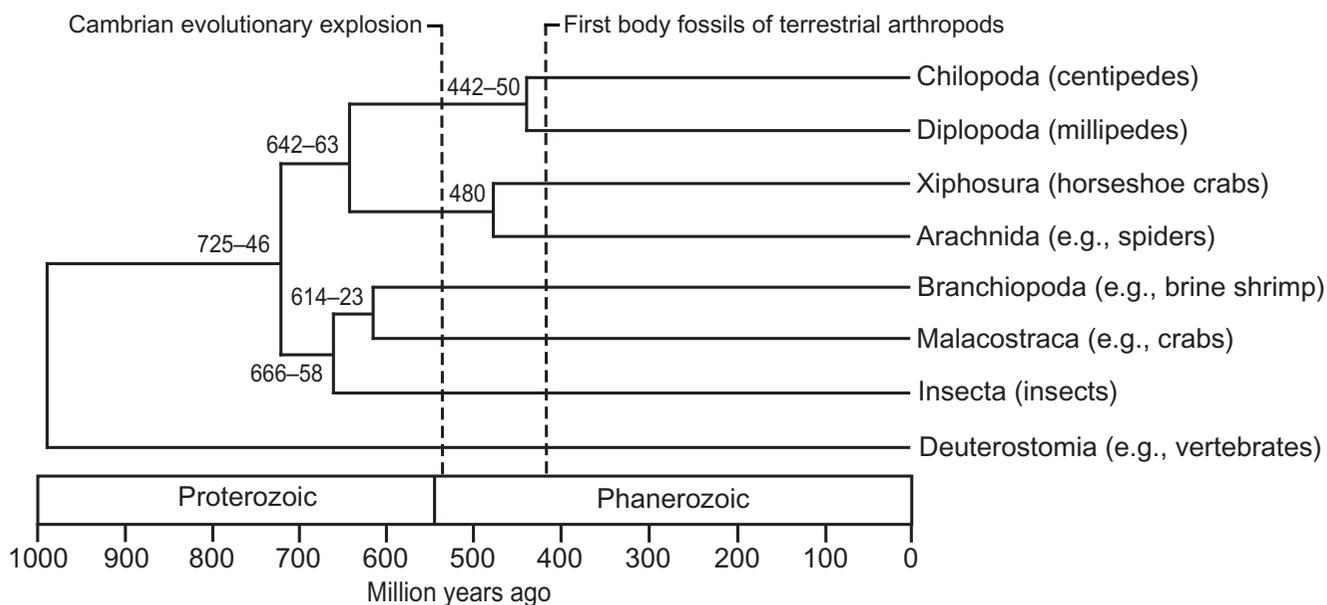
Figure 2
A timescale of arthropod evolution Numbers associated with nodes are divergence times (Ma) and their standard errors (Table 2). Three calibration points were used: the fossil-based divergence of Chilopoda and Diplopoda (423 Ma), the fossil-based divergence of Xiphosura and Arachnida (480 Ma), and the 993 Ma deuterostome-arthropod divergence estimated from a previous molecular clock analysis.

Table 3: Divergence times between Arthropoda and Deuterostomia. Times of divergence for deuterostomes and arthropods using one or both calibrations from the arthropod fossil record. Abbreviations: CD= Chilopoda-Diplopoda calibration (423 Ma); XA= Xiphosura-Arachnida calibration (480 Ma).

Deuterostome representative	Calibration point	Number of proteins		Divergence times (Ma) and standard errors (if available)				
		Total	Rate constant	MGG_{LS}	SGG_{LS}	SGL_{DT}	SGL_{MDT}	SGL_{PL}
Vertebrata	CD	5	5	999 ± 96	969 ± 96	1080 ± 147	913 ± 83	1395
	XA	5	4	966 ± 130	931 ± 130	1243 ± 170	1022 ± 92	1554
	CD & XA	5	4	935 ± 69	999 ± 69	1170 ± 147	978 ± 80	1416
Tunicata	CD	5	4	1189 ± 123	1104 ± 123	988 ± 154	918 ± 93	1127
	XA	5	4	1205 ± 165	1151 ± 165	1320 ± 165	1041 ± 103	1362
	CD & XA	5	4	1185 ± 128	1130 ± 128	1236 ± 133	1018 ± 95	1362

pede divergence (442 ± 50 Ma) or the earliest terrestrial fossils (~420 Ma). Time estimates for the deuterostome-arthropod divergence ranged from 913 – 1554 Ma across all methods and calibrations (average of 1130 ± 120 Ma), although SGG_{PL} gave substantially higher estimates than the other methods (Table 3). The Bayesian (SGL_{DT} and SGL_{MDT}), multigene global (MGG_{LS}), and supergene glo-

bal (SGG_{LS}) methods resulted in estimates of ~1100 – 900 Ma regardless of calibration point used or whether a vertebrate or non-vertebrate (tunicate) was used as the representative deuterostome.

Discussion

Arthropod phylogeny

A major limitation of this study, with respect to phylogenetic implications, is the sparse taxonomic sampling. As in most studies, there is a trade-off in terms of taxa and genes or proteins. In this case, we have emphasized a large number of proteins to increase the statistical resolution of relationships and time estimates at the expense of taxonomic representation. Nonetheless, our results agree with most previous molecular phylogenetic analyses in supporting a close relationship between insects and crustaceans (Pancrustacea) and between myriapods and chelicerates. Of the two groups, Pancrustacea has received the strongest support in the past [20-24]. Nonetheless, a myriapod-chelicerate grouping has been found previously with mitochondrial genes [23-25], nuclear ribosomal genes [21], and nuclear protein-coding genes [26]. Additional evidence has come from hemocyanine structure [27].

We propose the name Myriochelata (in allusion to the joining of Myriapoda and Chelicerata) for the group containing myriapods and chelicerates, which otherwise is unnamed [28]. Although we are unable to identify any morphological trait diagnostic of this clade, some trends are evident that might reflect the morphological or ecological nature of the ancestral myriochelatan. For example, many species of extant myriochelatans (e.g., spiders, scorpions, centipedes) inject and immobilize prey with a poison, albeit with structures that are not homologous. Envenomation of prey is also found among pancrustaceans, but it is less broadly distributed in that group. Certainly, envenomation has arisen multiple times in arthropods, associated mostly (but not exclusively) with terrestrial predation. The significance of this trait in arthropod evolution must await sequence evidence from a greater diversity of taxa (e.g., pycnogonids, remipedes) than is currently available, and a careful examination of the early fossil record of animals (especially from the Cambrian). In general, the difficulty in finding morphological characters diagnostic of these major clades of arthropods is probably the result of deep branching of the lineages and an early fossil record that shows great morphological diversification (the Cambrian Explosion) and some important gaps [26,29,30].

Timescale for animal evolution and colonization of land

Among animals, arthropods have been considered to be the earliest colonizers of land based on fossil evidence [5-7]. However, it is possible that other animal phyla colonized land even earlier. Among them, nematodes, tardigrades, and annelids are likely candidates given their current exploitation of terrestrial environments, yet these groups have relatively poor fossil records.

Determining the number of such colonization events requires a consideration of phylogeny, the fossil record, and morphological traits associated with terrestriality. For arthropods, at least four major colonization events are inferred, leading to the chilopods + diplopods, insects, arachnids, and isopod crustaceans. Additional events may have occurred in smaller lineages [24,25,31]. Moreover, the recent discovery of a basal marine hexapod fossil from the Devonian [32] suggests that some hexapod traits previously believed to have evolved as adaptations to land may have first appeared in a marine setting. Along the same lines, it is also possible that millipedes and centipedes colonized land independently. However, because the earliest fossils of those groups are presumably terrestrial [6] and our molecular time estimate is only 5% earlier than the age of those fossils, our assumption of a terrestrial common ancestor of millipedes and centipedes has little affect on the time of colonization.

Our relatively young time estimates for the millipede-centipede (442 ± 50 Ma) and xiphosuran-arachnid (475 ± 53 Ma) divergences contrast with the much older time estimate for the deuterostome-arthropod divergence (~1100-900 Ma) (Tables 1 and 2). The first two are close to the corresponding fossil record estimates whereas the third one is ~400 million years earlier than the earliest fossil evidence for animals [12]. Most molecular clock analyses in the last three decades, including those using many genes, have resulted in similarly deep time estimates for the arthropod-deuterostome divergence [33]. Currently a debate exists as to whether divergences among animals are best represented by molecular clock dates or the fossil record [33-37]. The molecular clock dates suggest that early animals fossilized poorly, possibly because they were small and soft-bodied. Alternatively, others have argued that molecular time estimates are older than the true times because of statistical biases, rate changes, and calibration biases [36,38], although replies to those criticisms have been made [37].

The results of this study address several of these criticisms of molecular clocks in the following ways. (1) The time estimates reported here are not uniformly discordant with the fossil record suggesting that if statistical biases are present, they are not causing a directional and proportional bias in all time estimates. (2) Our time estimate for the deuterostome-arthropod divergence using the arthropod fossil record (exclusively) was similar to time estimates obtained in previous studies by using the vertebrate fossil record, suggesting that the vertebrate calibration is not obviously biasing time estimates. (3) Our use of tunicates instead of vertebrates for representing deuterostomes resulted in similar time estimates, further indicating that vertebrates (per se) are not causing a systematic bias in the time analysis. (4) Our use of a diversity

of clock methodology, including Bayesian inference and likelihood-smoothing methods, did not alter the conclusions, indicating that the global clock methods used in previous studies were not responsible for biased time estimates. (5) The use of concatenated alignments (supergenes) yielded similar results to non-concatenated (multigene) analyses, indicating that the multigene analyses of previous studies were not responsible for the discordance between molecular clocks and the fossil record.

In summary, our molecular clock analysis resulted in some time estimates (e.g. millipede-centipede) in agreement with the fossil record and others (e.g., insect-crustacean) much earlier than fossil evidence (assuming that the crustaceans are, in fact, monophyletic). However, all studies, including this one, are limited by the small size of the available sequence data (5 – 10 proteins across major groups of arthropods). In the future, it should be possible to use several hundred proteins across a diversity of arthropod taxa, and such an analysis should greatly increase the precision of phylogenetic and molecular clock analyses.

Conclusions
We have found strong statistical support for the unconventional grouping of myriapods and chelicerates, a taxon that we herein name Myriochelata. We also note that many myriochelatans immobilize their prey with venom. Using only arthropod fossil calibrations, our molecular clock analyses support earlier studies that used vertebrate calibrations in finding a deep divergence of deuterostomes and arthropods, hundreds of millions of years before the Cambrian explosion. However, our much younger molecular time estimate for the millipede-centipede divergence is close to the divergence time inferred from fossils for that node. This suggests that the colonization of land by arthropods occurred relatively late in their evolutionary history.

Methods
Species and sequences
We sequenced the protein-coding region of two nuclear glycolytic enzymes, glyceraldehyde 3-phosphate dehydrogenase (G3PDH) and enolase, in an annelid (*Nereis macrydi*), a mollusk (*Marisa* sp.; not sequenced for enolase), and eight species of arthropods: a centipede (*Lithobius* sp.), millipede (Diplopoda sp.), blue crab (*Callinectes sapidus*), brine shrimp (*Artemia* sp.), water flea (*Daphnia magna*; not sequenced in G3PDH), muscle shrimp (Ostracoda sp.; possibly *Cypridopsis* sp.), Atlantic horseshoe crab (*Limulus polyphemus*), tarantula (*Phormictopus* sp.), and a scorpion (*Centruroides* sp.). Reverse-transcriptase PCR was used to amplify and sequence (both complementary strands) a total of approximately 1 100 base pairs (~363 amino acids) of the coding region of enolase and 890 base

pairs (~294 amino acids) of the coding region of G3PDH. Additional sequence data were obtained from the public databases (GenBank). Genes were selected if there were nucleotide sequences available from an annelid (outgroup) and at least one representative of the following arthropod taxa: Branchiopoda (e.g., brine shrimp), Malacostraca (e.g., blue crabs), Diplopoda (e.g., millipedes), Chilopoda (e.g., centipedes), Insecta (e.g., insects), Xiphosura (e.g., horseshoe crabs), and Arachnida (e.g., spiders and scorpions). Sequences were aligned with Clustal X [39]. Additional methods and sequence accession numbers are presented elsewhere (Additional file 1).

Phylogenetic methods
The alignments of all nine nuclear and 15 mitochondrial genes were concatenated (Additional file 1) because individual gene data sets were, for the most part, insufficient for statistical resolution of arthropod phylogeny (Additional file 2). This required the use of exemplars (representatives) from each of the major groups. Preliminary analyses using 16 – 54 taxa (Additional file 1) were used as a guide to choosing exemplars so that fast- or slow-evolving species were not selected.

Analyses of the concatenation of the nine nuclear genes, as well as a full data set of 24 genes, were carried out using minimum evolution, maximum parsimony, maximum likelihood, and Bayesian inference. Minimum evolution analyses were performed using Kimura two-parameter distances with transversions-only, gamma corrected Kimura two-parameters distances with transversions-only, and paralinear distances. For the paralinear distance analyses, the proportion of invariable sites in the considered alignment was estimated using likelihood and assuming a HKY85 + gamma + proportion of invariable sites model of DNA evolution. Identical sites were then removed proportionally to the base frequencies estimated from all sites (PAUP default settings). Maximum parsimony analyses were carried out with an equal weighting and a 2:1 weighting scheme favoring transversions over transitions. The model used for the maximum likelihood analyses (GTR + gamma+ proportion of invariable sites) was selected using Modeltest [19], whereas Bayesian inference was carried out under mixed models, and full parameter estimation was performed during tree search for each gene.

All minimum evolution analyses were performed using MEGA 2.1 [40] and PAUP [41], while likelihood and parsimony analyses, as well as the estimation of the specific gamma parameters for the minimum evolution analyses were carried out using PAUP. Bayesian analyses were performed with MrBayes 3.0 [42]. The branch and bound algorithm was used in the parsimony analyses, while in the likelihood analyses 100 heuristic searches were per-

formed. In the latter, starting trees were obtained using random sequence addition and swapped using the tree bisection reconnection algorithm. In the Bayesian analysis, 300 000 generations were run and trees were used only after convergence was reached.

Robustness for the nodes in the minimum evolution, maximum parsimony, and maximum likelihood trees was evaluated using the bootstrap. PAUP settings for the bootstrap analyses were as follows: 2 000 replicates with full heuristic search (one random addition sequence) and the multiple tree option turned off. Trees were swapped using the TBR algorithm. Support for the groups recovered in the Bayesian analyses were expressed as their posterior probabilities. Gapped sites were removed prior to all analyses.

Molecular clock methods
Molecular clock analyses were performed using a diversity of methods [37]: multigene global least squares (MGG_{LS}) [43,44], multigene local least squares (MGL_{LS}) [45], supergene global least squares (SGG_{LS}) [46], supergene local least squares (SGL_{LS}), supergene local divtime (SGL_{DT}; DivTime5b) [47], supergene local multidivtime (SGL_{MDT}; Multidivtime) [48], and supergene local penalized likelihood (SGL_{PL}; r8s version 1.5) [49,50]. Least squares methods are distance based, SGL_{PL} is a semi-parametric likelihood-smoothing method, and SGL_{DT} and SGL_{MDT} are Bayesian methods. Multigene methods use the mean or mode [37,44,51] of time estimates from individual proteins whereas supergene methods derive a single time estimate from the simultaneous analysis of all available proteins. SGL_{DT} and SGL_{MDT} are two different implementations of the Bayesian method of Thorne *et al.* [47]. The difference between the two being that with SGL_{DT} protein sequences are concatenated and considered as a single entity, whereas with SGL_{MDT} each considered protein maintains its individuality during the analyses.

We used the divergence of centipedes (Chilopoda) and millipedes (Diplopoda) as the minimal (most recent) time of arthropod terrestrialization and the divergence of myriapods with their closest relatives (in this case, chelicerates) as the maximal (earliest) time for myriapod terrestrialization. The maximal time for arthropod terrestrialization was considered as the divergence between insects (Insecta) and crustaceans (Crustacea). This does not preclude the possibility that land was colonized even earlier by extinct species or groups of arthropods (or other animals).

Divergence times were estimated using two calibration points from the arthropod fossil record (Additional file 1): Xiphosura-Arachnida (~480 Ma) [52] and Chilopoda-Diplopoda (423 Ma; see Additional File 1 for explanation

of date) [53,54]. A third calibration point, the divergence of deuterostomes (vertebrates) and arthropods based on a molecular clock study using 50 proteins and calibrated with the vertebrate fossil record (993 Ma) [33], was used in some analyses and tested (with the arthropod fossil record and by substituting vertebrates with a tunicate, *Ciona intestinalis*) in others (see below). Analyses were performed using calibration points separately and simultaneously, for comparison. The millipede-centipede fossil divergence was not used to estimate the molecular clock time of that divergence.

GenBank was screened for nuclear encoded proteins (>100 amino acids) in which divergence times could be estimated between two or more arthropod groups, resulting in 84 proteins. For each, reciprocal BLAST analyses were used to identify and assemble sets of available sequences, and orthology was investigated through phylogenetic analyses. In those analyses, 61 proteins (27 291 amino acid positions) were selected for further analysis after determining that orthologous relationships (divergences corresponding to speciation events) could be distinguished from paralogous relationships (divergences corresponding to gene duplication events).

The least squares-based global and local clock methods (MGG_{LS}, SGG_{LS}, MGL_{LS}, SGL_{LS}) were performed on 49 proteins (Additional file 1) that passed the relative rate test [55] as implemented in PHYLTEST [56]. Gamma-corrected Poisson distances were used for these analyses and specific gamma parameters were estimated with PAML [57]. The Bayesian and the likelihood-smoothing local clock methods (SGL_{DT}, SGL_{MDT}, and SGL_{PL}) were used with all 61 proteins that passed or did not pass the rate tests. Branch lengths for the penalized likelihood analyses were estimated with PAML (assuming a Gamma corrected JTT model of amino acid substitution). For the Bayesian analyses, branch length estimation was performed using the software AUTOestbranches [47] in the case of SGL_{DT}, and Estbranches [48] in the case of SGL_{MDT}. Standard errors for SGG_{LS} and SGL_{LS} were calculated from among-gene comparisons (i.e., MGG_{LS} and MGL_{LS}) and were not available for SGL_{PL}.

Authors' contributions
DP carried out the bioinformatic research and drafted the manuscript. LLP and ML sequenced the new G3PDH and enolase sequences. LLP carried out preliminary analyses of the sequence data. SBH directed the research and assisted with drafting of the manuscript. All authors read and approved the final manuscript.

Additional material

Additional File 1

This file (MS Word) includes: (1) additional details of the methodology, (2) accession numbers of sequences used in the phylogenetic and molecular clock analyses, (3) the aligned sequence data used in the phylogenetic analyses, and (4) additional information regarding the fossil calibration points.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1741-7007-2-1-S1.doc]

Additional File 2

This file (pdf) contains phylogenetic trees of arthropods estimated using five representative genes. The trees were constructed with neighbor-joining, using gamma-corrected Kimura two-parameter transversion distances. Bootstrap confidence values (>50%) are shown on nodes.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1741-7007-2-1-S2.pdf]

Acknowledgements

References

1. Horodyski RJ, Knauth LP: **Life on land in the Precambrian.** *Science* 1994, **263:**494-498.
2. Watanabe Y, Martini JEJ, Ohmoto H: **Geochemical evidence for terrestrial ecosystems 2.6 billion years ago.** *Nature* 2000, **408:**574-578.
3. Prave AR: **Life on land in the Proterozoic: Evidence from the Torridonian rocks of northwest Scotland.** *Geology* 2002, **30:**811-814.
4. Schwartzman DW: **Life, temperature, and the Earth.** *New York, Columbia University Press*; 1999:241.
5. Shear WA, Bonamo PM, Grierson JD, Rolfe WDI, Smith EL, Norton RA: **Early land animals in North America.** *Science* 1984, **224:**492-494.
6. Jeram AJ, Selden PA, Edwards D: **Land animals in the Silurian: arachnids and myriapods from Shropshire, England.** *Science* 1990, **250:**658-661.
7. Gray J, Shear W: **Early life on land.** *American Scientist* 1992, **80:**444-456.
8. Redecker D, Kodner R, Graham LE: **Glomalean fungi from the Ordovician.** *Science* 2000, **289:**1920-1921.
9. Wellman CH, Osterloff PL, Mohluddin U: **Fragments of the earliest land plants.** *Nature* 2003, **425:**282-285.
10. Xiao SH, Zhang Y, Knoll AH: **Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite.** *Nature* 1998, **391:**553-558.
11. Li CW, Chen JY, Hua TE: **Precambrian sponges with cellular structures.** *Science* 1998, **279:**879-882.
12. Barfod GH, Albarede F, Knoll AH, Xiao SH, Telouk P, Frei R, Baker J: **New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils.** *Earth Planet Sc Lett* 2002, **201:**203-212.
13. Porter SM, Knoll AH: **Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon.** *Paleobiology* 2000, **26:**360-385.
14. Butterfield NJ: **Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes.** *Paleobiology* 2000, **26:**386-404.
15. Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB: **Molecular evidence for the early colonization of land by fungi and plants.** *Science* 2001, **293:**1129-1133.
16. Johnson EW, Briggs DEG, Suthren RJ, Wright JL, Tunnicliff SP: **Non-marine arthropod traces from the subaereal Ordivician Borrowdale volcanic group, English Lake District.** *Geol Mag* 1994, **131:**395-406.
17. MacNaughton RB, Cole JM, Dalrymple RW, Braddy SJ, Briggs DEG, Lukie TD: **First steps on land: Arthropod trackways in Cambrian-Ordovician eolian sandstone, southeastern Ontario, Canada.** *Geology* 2002, **30:**391-394.
18. Robison RA, Wiley EO: **A new arthropod, Meristoma: more fallout from the Cambrian explosion.** *Journal of Paleontology* 1995, **69:**447-459.
19. Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14:**817-818.
20. Shultz JW, Regier JC: **Phylogenetic analysis of arthropods using two nuclear protein-encoding genes supports a crustacean + hexapod clade.** *Proc R Soc Lond B Biol Sci* 2000, **267:**1011-1019.
21. Friedrich M, Tautz D: **Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods.** *Nature* 1995, **376:**165-167.
22. Giribet G, Edgecombe GD, Wheeler WC: **Arthropod phylogeny based on eight molecular loci and morphology.** *Nature* 2001, **413:**157-161.
23. Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W: **Mitochondrial protein phylogeny joins myriapods with chelicerates.** *Nature* 2001, **413:**154-157.
24. Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F: **Hexapod origins: monophyletic or paraphyletic?** *Science* 2003, **299:**1887-1889.
25. Delsuc F, Phillips MJ, Penny D: **Comment on "Hexapod origins: monophyletic or paraphyletic?".** *Science* 2003, **301:**1482; author reply 1482.
26. Cook CE, Smith ML, Telford MJ, Bastianello A, Akam M: **Hox genes and the phylogeny of the arthropods.** *Curr Biol* 2001, **11:**759-763.
27. Kusche K, Burmester T: **Diplopod hemocyanin sequence and the phylogenetic position of the Myriapoda.** *Mol Biol Evol* 2001, **18:**1566-1573.
28. Štys P, Zrzavý J: **Phylogeny and classification of extant Arthropoda: a review of hypotheses and nomenclature.** *Eur J Entomol* 1994, **91:**257-275.
29. Fortey RA, Owens RM: **Evolutionary radiations in the Trilobita.** *Major evolutionary radiations* Edited by: Taylor PD and Larwood GP. Oxford, Clarendon Press; 1990:139-164.
30. Lieberman BS: **Phylogenetic analysis of some basal early Cambrian trilobites, the biogeographic origins of the Eutrilobita, and the timing of the Cambrian radiation.** *J Paleontol* 2002, **76:**692-708.
31. Dunlop JA, Webster M: **Fossil evidence, terrestrialization and arachnid phylogeny.** *J Arachnol* 1999, **27:**86-93.
32. Haas F, Waloszek D, Hartenberger R: **Devonohexapodus bocksbergensis, a new marine hexapod from the Lower Devonian Hunsrück Slates, and the origin of Atelocerata and Hexapoda.** *Organisms, Diversity and Evolution* 2003, **3:**39-54.
33. Wang DY, Kumar S, Hedges SB: **Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi.** *Proc R Soc Lond B Biol Sci* 1999, **266:**163-171.
34. Wray GA, Levinton JS, Shapiro LH: **Molecular evidence for deep Precambrian divergences among metazoan phyla.** *Science* 1996, **274:**568-573.
35. Fortey RA, Briggs DEG, Wills MA: **The Cambrian evolutionary 'explosion' recalibrated.** *Bioessays* 1997, **19:**429-434.
36. Benton MJ, Ayala FJ: **Dating the tree of life.** *Science* 2003, **300:**1698-1700.
37. Hedges SB, Kumar S: **Genomic clocks and evolutionary timescales.** *Trends in Genetics* 2003, **19:**200-206.
38. Benton MJ: **Early origins of modern birds and mammals: molecules vs. morphology.** *Bioessays* 1999, **21:**1043-1051.
39. Thompson JD, Higgins DG, Gibson TJ: **CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Research* 1994, **22:**4673-4680.

40. Kumar S, Tamura S, Jakobsen IB, Nei M: **MEGA2: Molecular Evolutionary Genetics Analysis software.** *Tempe, AZ, USA, Arizona State University*; 2001.

41. Swofford DL: **PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.** *Sunderland, Massachusetts, Sinauer Associates*; 1998.

42. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.

43. Hedges S. Blair, Parker Patrick H., Sibley Charles G., Kumar Sudhir: **Continental breakup and the ordinal diversification of birds and mammals.** *Nature* 1996, **381:**226-229.

44. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392:**917-920.

45. Schubart Christoph D., Diesel Rudolf, Hedges S. Blair: **Rapid evolution to terrestrial life in Jamaican crabs.** *Nature* 1998, **393:**363-365.

46. Nei M, Xu P, Glazko G: **Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms.** *Proceedings of the National Academy of Sciences (U.S.A.)* 2001, **98:**2497-2502.

47. Thorne JL, Kishino H, Painter IS: **Estimating the rate of evolution of the rate of molecular evolution.** *Molecular Biology and Evolution* 1998, **15:**1647-1657.

48. Thorne JL, Kishino H: **Divergence time and evolutionary rate estimation with multilocus data.** *Systematic Biology* 2002, **51:**689-702.

49. Sanderson MJ: **Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach.** *Molecular Biology and Evolution* 2002, **19:**101-109.

50. Sanderson MJ: **r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock.** *Bioinformatics* 2003, **19:**301-302.

51. Hedges SB, Shah P: **Comparison of mode estimation methods and application in molecular clock analysis.** *BMC Bioinformatics* 2003, **4:**31.

52. Dunlop JA, Selden PA: **The early history and phylogeny of the chelicerates.** *Arthropod relationships* Edited by: *Fortey RA and Thomas RH. London, Chapman & Hall*; 1998:221-235.

53. Almond JE: **The Silurian-Devonian fossil record of the Myriapoda.** *Phil Trans R Soc Lond B* 1985, **309:**227-237.

54. Shear WA, Selden PA: **Rustling in the undergrowth: animals in early terrestrial ecosystems.** *Plants invade the land* Edited by: *Gensel PG and Edwards D. New York, Columbia University Press*; 2001:29-51.

55. Takezaki Nauko, Rzhetsky Andrey, Nei Masatoshi: **Phylogenetic test of the molecular clock and linearized trees.** *Molecular Biology and Evolution* 1995, **12:**823-833.

56. Kumar S: **PHYLTEST.** 2.0.th edition. *University Park, PA, USA., IMEG, Department of Biology, The Pennsylvania State University*; 1996.

57. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *CABIOS* 1997, **13:**555-556.