

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

## Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

## Dating the arthropod tree based on large-scale transcriptome data

Peter Rehm<sup>a</sup>, Janus Borner<sup>a</sup>, Karen Meusemann<sup>b</sup>, Björn M. von Reumont<sup>b</sup>, Sabrina Simon<sup>c</sup>, Heike Hadrys<sup>c</sup>, Bernhard Misof<sup>b</sup>, Thorsten Burmester<sup>a,\*</sup><sup>a</sup> Biozentrum Grindel & Zoologisches Museum, Martin-Luther-King Platz 3, D-20146 Hamburg, Germany<sup>b</sup> Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160, D-53113 Bonn, Germany<sup>c</sup> Stiftung Tierärztliche Hochschule Hannover, ITZ, Ecology & Evolution, Bünteweg 17d, D-30559 Hannover, Germany

## ARTICLE INFO

## Article history:

Received 18 March 2011

Revised 1 September 2011

Accepted 5 September 2011

Available online 17 September 2011

## Keywords:

Arthropoda

Bayesian methods

Cambrian explosion

CIR process

Molecular clock

## ABSTRACT

Molecular sequences do not only allow the reconstruction of phylogenetic relationships among species, but also provide information on the approximate divergence times. Whereas the fossil record dates the origin of most multicellular animal phyla during the Cambrian explosion less than 540 million years ago (mya), molecular clock calculations usually suggest much older dates. Here we used a large multiple sequence alignment derived from Expressed Sequence Tags and genomes comprising 129 genes (37,476 amino acid positions) and 117 taxa, including 101 arthropods. We obtained consistent divergence time estimates applying relaxed Bayesian clock models with different priors and multiple calibration points. While the influence of substitution rates, missing data, and model priors were negligible, the clock model had significant effect. A log-normal autocorrelated model was selected on basis of cross-validation. We calculated that arthropods emerged ~600 mya. Onychophorans (velvet worms) and euarthropods split ~590 mya, Pancrustacea and Myriochelata ~560 mya, Myriapoda and Chelicerata ~555 mya, and 'Crustacea' and Hexapoda ~510 mya. Endopterygote insects appeared ~390 mya. These dates are considerably younger than most previous molecular clock estimates and in better agreement with the fossil record. Nevertheless, a Precambrian origin of arthropods and other metazoan phyla is still supported. Our results also demonstrate the applicability of large datasets of random nuclear sequences for approximating the timing of multicellular animal evolution.

© 2011 Elsevier Inc. All rights reserved.

## 1. Introduction

Dating of diversification and speciation events is a major aim of evolutionary studies. For a long time, fossil remains were the prime source of such time estimates. The fossil record, however, is far from complete and in many cases the taxonomic assignment of fossil specimens is uncertain (Benton and Donoghue, 2007). DNA and protein sequences provide a complementary source of information for the inference of life history. Although there is an ongoing debate whether such a molecular clock approach is actually valid (Graur and Martin, 2004), many studies have obtained reasonable time estimates for a broad range of taxa (for review, see: Benton and Ayala, 2003; Hedges and Kumar, 2003).

*Abbreviations:* mya, million years ago; myr, million years; LOG, log-normal autocorrelated clock model; CIR, Cox–Ingersoll–Ross model; UGM, uncorrelated gamma multipliers.

\* Corresponding author. Address: Institute of Zoology and Zoological Museum, Biocenter Grindel University of Hamburg, Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany. Fax: +49 40 42838 3937.

E-mail address: [thorsten.burmester@uni-hamburg.de](mailto:thorsten.burmester@uni-hamburg.de) (T. Burmester).

1055-7903/\$ - see front matter © 2011 Elsevier Inc. All rights reserved.  
doi:10.1016/j.ympev.2011.09.003

In theory, molecular clock calculations have the power to be more precise than fossil dates because latter usually are underestimates. At best fossils provide an approximation to the oldest member of the taxon in question (cf. Benton and Ayala, 2003). In fact, sequence-derived dates tend to be older than the fossil dates (Hedges and Kumar, 2003). This is particularly true for deep divergence times. For example, the first conclusive fossil evidence for crown group bilaterians dates ~550–530 mya (Benton and Donoghue, 2007), but molecular estimates suggest an emergence of bilaterians between 1300 and 670 mya (e.g., Blair and Hedges, 2005; Lynch, 1999; Otsuka and Sugaya, 2003; Peterson et al., 2008). The discrepancy of molecular and fossil dates, and among different molecular clock approaches can be attributed to insufficient data, wrong taxonomic assignment or dating of fossils, and, most importantly, to rate heterogeneity among lineages over time and between genes (e.g., Adachi and Hasegawa, 1995; Benton and Donoghue, 2007; Bromham et al., 1998; Graur and Martin, 2004).

The undisputed fossil record of the phylum Arthropoda dates back to the early Cambrian period (Budd and Jensen, 2000; Budd and Telford, 2009; Edgecombe, 2010). The identity of possible representatives of arthropods from the earlier Ediacaran period is questionable (Nielsen, 2001). Based on fossils and geological

considerations, Benton and Donoghue (2007) assumed an earliest date of 581 mya for the divergence of Arthropoda and Nematoda. Molecular clock analyses, however, usually support much older time estimates that range from 1200 to 625 mya for the origin of Arthropoda (Blair, 2009; Blair et al., 2005; Douzery et al., 2004; Hausdorf, 2000; Lee, 1999; Sanders and Lee, 2010; Wang et al., 1999). Due to the lack of sequence data from important taxa, calculations of internal divergence times within arthropods are sparse, with the exception of the insects (e.g., Gaunt and Miles, 2002; Regier et al., 2004, 2005).

Here we analyze the divergence times of major arthropod taxa based on a superalignment spanning 37,476 amino acid positions, which had been derived from Expressed Sequence Tags (ESTs) (Meusemann et al., 2010). This is – to the best of our knowledge – the largest dataset that has ever been used for molecular clock studies.

## 2. Materials and methods

### 2.1. Sequence data and phylogenetic tree

In a previous study (Meusemann et al., 2010), 775 orthologous genes from 214 euarthropods, three onychophorans, two tardigrades, eight nematodes, three annelids and three mollusks, derived from EST data and selected genomes were identified with the HaMStR approach (Ebersberger et al., 2009). Single multiple protein alignments were generated with MAFFT L-INSI (Katoh and Toh, 2008). Alignment masking was conducted with ALISCORE and ALICUT (Kück et al., 2010; Misof and Misof, 2009). An optimal subset of data was selected by MARE 01-alpha (MATRIX REDuction; <http://mare.zfmk.de>). The finally selected superalignment spans 37,476 amino acid positions, comprised 129 genes and 117 taxa, including 101 arthropods (available at TreeBase, <http://www.treebase.org>, under study accession no. S10507). A Bayesian phylogenetic tree was inferred with PhyloBayes (Lartillot et al., 2009). For details, refer to Meusemann et al. (2010).

### 2.2. Bayesian estimates of divergence times

The Bayesian phylogenetic majority rule consensus tree (Meusemann et al., 2010) was used as input for molecular clock estimates. The program PhyloBayes 3.2d was applied to calculate divergence times and 95% confidence intervals within a Bayesian framework (Lartillot et al., 2009). Three relaxed clock models, the log-normal autocorrelated clock model (LOG) (Thorne et al., 1998), the 'CIR' process (CIR) (Cox et al., 1985; see also Lepage et al., 2006) and uncorrelated gamma multipliers (UGM) (Drummond et al., 2006), were used under a uniform prior on divergence times for 50,000 cycles with a burn-in of 20,000 cycles. The CIR process is similar to the Ornstein–Uhlenbeck model, but with superior properties (Aris-Brosou and Yang, 2003; Lepage et al., 2006, 2007). Rates across sites were modeled assuming a discrete gamma distribution with four categories and with a Dirichlet process. Bayes factors were estimated using thermodynamic integration as implemented in PhyloBayes (Lepage et al., 2007) with 100,000 generations and a burn-in of 10,000. The three relaxed clock models were compared with the unconstrained model. Cross-validation of the models was performed by dividing the alignment into eight subsets (seven learning sets and one test set). Ten repetitions were run, as specified in PhyloBayes.

In a first approach, we tested the effects of gamma distributed priors for the root node on our results. To evaluate the impact of the priors, we defined different means and standard deviations (SD) of the prior distribution: mean 1000 mya (SD 1000 myr/500 myr) and 750 mya (SD 750 myr/325 myr), respectively. In

addition, a uniform root prior was assumed with a maximum limit of 5000 mya imposed by PhyloBayes. All analyses were also run under the priors (i.e. with no data) to assess whether the prior distribution was sufficiently wide (i.e. non-informative). The results were compared with those obtained when the data were analyzed.

To assess the impact of missing data, all amino acid positions in the concatenated alignment were sorted according to their taxon coverage. Only the 50% of positions with the highest taxon coverage were used in a separate molecular clock analysis with the same settings as described above. The effect of substitution rates was tested by dividing the complete superalignment (129 genes) into three subsets, each containing 43 genes with *i* lowest, *ii* intermediate, and *iii* highest substitution rates. Genes were assigned to these categories according to the mean PAM distance of all possible sequence pairs within each alignment. To avoid artifacts due to missing data, only taxa for which sequences of all genes are present were selected for the assessment of pairwise distances: *Apis mellifera*, *Bombyx mori*, *Daphnia pulex*, *Drosophila melanogaster*, and *Tribolium castaneum*. Positions with gaps were ignored. All three subsets were analyzed in separate runs according to the procedure described above.

### 2.3. Calibration of the molecular clock

Seven calibration points were evenly distributed throughout the phylogenetic tree, including one calibration point within the outgroup (Table 1; Supplemental Table S1). We aimed to cover different regions of the tree and to include calibrations for deep nodes as well as for shallow nodes. To avoid a distortion of the time estimates by systematic misplacement of fossil calibration points, we used fossils with reliable systematic placement. Numerical ages were obtained from the International Stratigraphic Chart 2009 (<http://www.stratigraphy.org>), assuming the minimum age of the respective stage interval for calibration points 1–4 and 7 (Supplemental Table S1). The minimum age of calibration point 5 was dated according to the minimal age of the Namurian A/E1 (Dusar, 2006) and calibration point 6, for which a minimum and a maximum age was obtained from Benton and Donoghue (2007). Each of the settings described above was run with seven calibration points for each dataset. In addition, the complete dataset was analyzed with six calibration points (omitting calibration point 1 within the outgroup).

Calibration point 1: the minimum age of the divergence of Mollusca and Annelida is defined by small helcionelloids of the genus *Oelandiella* from the pre-Tommotian (Cambrian, Purella Biozone, Nemakit–Daldynian) period 528 mya (Gubanov and Peel, 1999; Khomentovsky and Karlova, 1993). Calibration point 2: evidence for euarthropods is provided by *Rusophycus*-like trace fossils from the early Tommotian 521 mya (Crimes, 1987). This calibration point is confirmed by recently described Lower Cambrian euarthropods fossils (Chen, 2009), which derives from the Maotianshan Shale (Qiongzhusian) dating 521–515 mya. Calibration point 3: the oldest unambiguous myriapod fossil is the millipede *Cowiedesmus eroticopodus* (Wilson and Anderson, 2004) from the Cowie Formation, Silurian. At that time, millipedes and centipedes had separated, thus *C. eroticopodus* provides a minimal age for the divergence of Diplopoda (millipedes) and Chilopoda (centipedes) at the transition from Wenlock to Ludlow 418.7 mya (base of Ludfordian, Ludlow). Calibration point 4: the split between Entognatha and Ectognatha (true insects) dates to the early Devonian (Pragian) period 404.2 mya, delimited by the first entognathan fossil of the springtail *Rhyniella precursor* (Whalley and Jarzembowski, 1981). Calibration point 5: the minimum date for the split between paleopteran and neopteran lineages is provided by an insect wing from the Upper Silesian Basin, Czech Republic (Béthoux and Nel, 2005), which dates to the Lower Carboniferous 324.8 mya and has been

**Table 1**  
Evolutionary history of Arthropoda. Mean divergence times averaged over settings of the log-normal autocorrelated clock model (see Supplemental Table S2). The asterisks denote calibration points (for additional information, see Supplemental Table S1).

Split	Mean divergence time (mya)	Fossil record (upper bound) (mya)	References
Ecdysozoa–Lophotrochozoa	607	528 (Mollusca)	Khomentovsky and Karlova (1993)
Cycloneuralia–Arthropoda	601	521 (Arthropoda)	Chen (2009) and Crimes (1987)
Nematoda–Tardigrada	574	503 (Tardigrada)	Müller et al. (1995)
Onychophora–Euarthropoda*	589	521 (Euarthropoda)	Chen (2009) and Crimes (1987)
Myriochelata–Pancrustacea	562	Early Cambrian	Shear and Edgecombe (2010)
Myriapoda–Chelicerata	556	Early Cambrian	Shear and Edgecombe (2010)
Diplopoda–Chilopoda*	504	419 (Diplopoda)	Wilson and Anderson (2004)
Pycnogonida–Euchelicerata	546	501 (Pycnogonida)	Waloszek and Dunlop (2002)
Xiphosura–Araneae–Acari	496	445 (Xiphosura)	Rudkin et al. (2008)
Xiphosura–Araneae	473	445 (Xiphosura)	Rudkin et al. (2008)
Malacostraca/Maxillipoda–Branchiopoda/Hexapoda	520	510 (Crustacea)	Harvey and Butterfield (2008)
Copepoda–Cirripedia/Malacostraca	507	505 (Cirripedia)	Collins and Rudkin (1981)
Cirripedia–Malacostraca	495	505 (Cirripedia)	Collins and Rudkin (1981)
Branchiopoda–Hexapoda	510	404 (Collembola)	Kukalová-Peck (1991)
Entognatha–Ectognatha*	485	404 (Collembola)	Kukalová-Peck (1991)
Archeognatha–Pterygota	455	390 (Archeognatha)	Labandeira et al. (1988)
Paleoptera–Neoptera*	419	325 (Archaeorhyncha)	Béthoux and Nel (2005)
Odonata–Ephemeroptera	388	318 (Odonata)	(Brauckmann and Schneider (1996)
Hemiptera–other neopterans	397	284 (Paleorrhyncha)	Shcherbakov (2000)
Orthoptera/Blattodea/Isoptera–Endopterygota (Holometabola)	391	307 (Coleoptera)	Béthoux (2009)
Orthoptera–Blattodea/Isoptera	351	311 (Blattodea)	Jarzembowski and Schneider (2007)
Isoptera–Blattodea*	200	137 (Isoptera)	Engel et al. (2007)
Hymenoptera–Coleoptera/Lepidoptera/Diptera	372	307 (Coleoptera)	Béthoux (2009)
Coleoptera–Lepidoptera/Diptera	353	307 (Coleoptera)	Béthoux (2009)
Lepidoptera–Diptera	342	270 (Panorpida)	Minet et al. (2010)
Brachycera–Culicomorpha*	281	239 (Psychodomorpha)	Krzeminski et al. (1994)

assigned to Archaeorhyncha based on its venation (Prokop et al., 2005). Calibration point 6: the oldest fossil of a clade including Culicomorpha and Brachycera (Diptera) is *Grauvogelia arzvilleriana* from the middle Triassic Grés-a-Voltzia Formation of France, thereby providing a minimum age of 238.5 mya. The maximum age (295.4 mya) is defined by the insect fauna of the Boskovic Furrow, Czech Republic (Krzeminski et al., 1994). This deposit harbors a wide range of insects, but no representative of the brachyceran + culicomorphan clade has been described from this or from any older deposit (Benton and Donoghue, 2007). Dating of both, minimum and maximum constraints have been explained in Benton and Donoghue (2007). Calibration point 7: first evidence for the split of isopterans (termites) from other blattodeans dates to the Berriasian (lower Cretaceous) period 137.2 mya and is defined by the isopteran fossil *Baissatermes lapideus* (Engel et al., 2007).

### 3. Results

#### 3.1. Molecular clock models

Molecular clock estimates were performed based on the Bayesian topology of the arthropod tree presented by Meusemann et al. (2010) (cf. Supplemental Fig. S1). Calculations of divergence times within the Bayesian framework were consistent within the different relaxed clock models (LOG, CIR and UGM; Supplemental Tables S2–S5). Estimates under the autocorrelated models (CIR and LOG) were similar, with CIR resulting in on average ~1% older dates (Table 2). The uncorrelated clock model (UGM) gave ~10% older dates than CIR or LOG. The choice of model for the rates across sites (gamma distribution or Dirichlet process) only had a minor effect on divergence time estimates, with mean absolute differences lower than 0.5% (Table 2).

The models were further compared by calculating the Bayes factors against the unconstrained model employing thermodynamic

integration (Lartillot and Philippe, 2006), as implemented in PhyloBayes. The logarithms of the Bayes factors were 36.75 for UGM, 31.34 for CIR and 39.09 for LOG, suggesting that the log-normal autocorrelated clock model fits the data best. Cross-validation confirmed this conclusion and showed that the LOG model outperforms CIR and UGM. The cross-validation score of LOG vs. CIR was  $2.25 \pm 11.8506$  and LOG vs. UGM was  $7.75 \pm 36.2138$  (CIR vs. UGM:  $5.5 \pm 37.5167$ ). Therefore, LOG was applied in the following analyses.

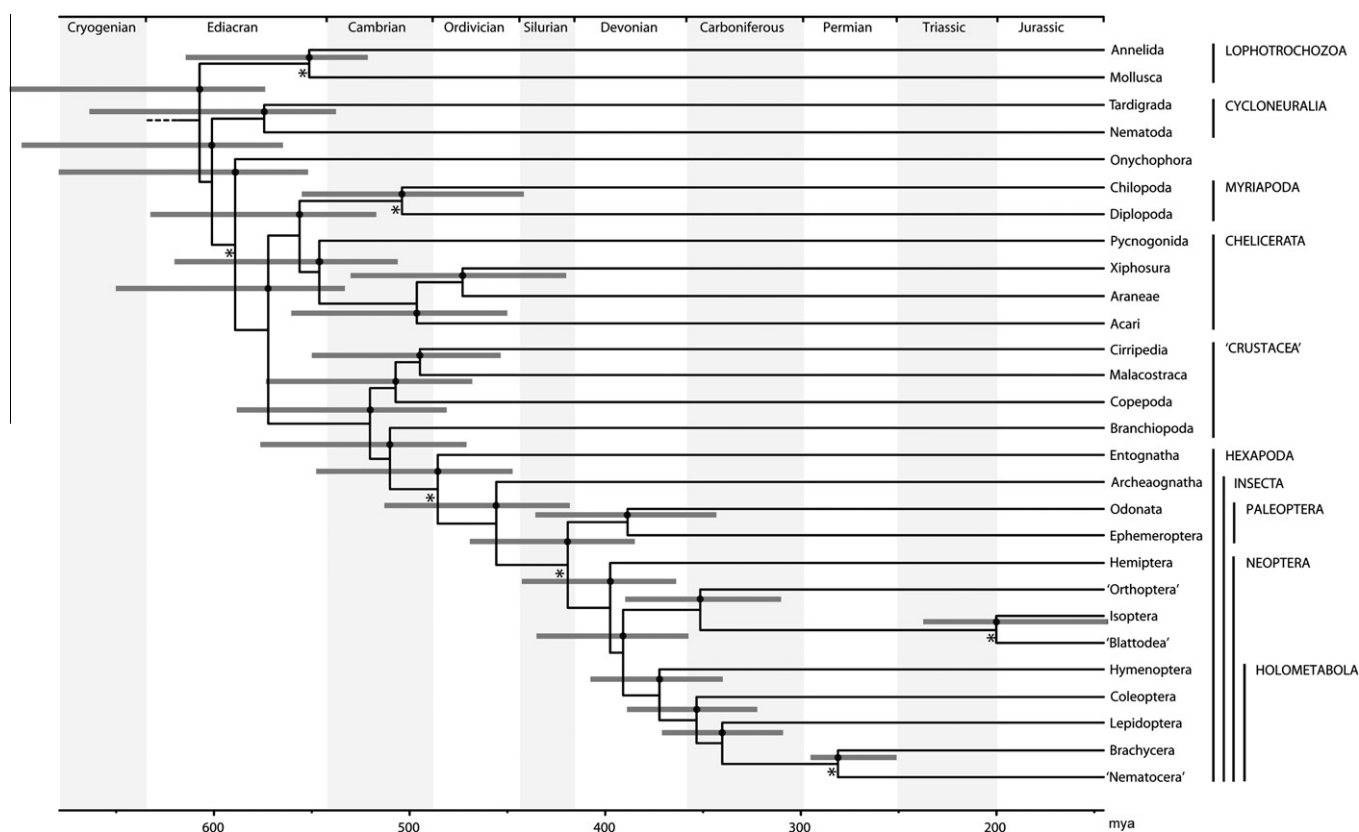
#### 3.2. Bayesian inference of arthropod divergence times

The divergence times over the means of all calculations resulting from the LOG model (Table 1; Supplemental Table S2) and 95% confidence intervals (Supplemental Table S3) were displayed in a linearized tree (Fig. 1). We estimated that the divergence of Ecdysozoa and Lophotrochozoa occurred 629–590 mya (mean 607 mya). Separation of the clade leading to Nematoda and Tardigrada on the one hand, and Panarthropoda on the other, dated to 621–584 mya (mean 601 mya). The branch leading to Onychophora diverged from Euarthropoda 607–573 mya (mean 589 mya). Myriochelata (Myriapoda + Chelicerata) and Pancrustacea diverged during the Ediacaran 580–546 mya (mean 562 mya). The split between Myriapoda and Chelicerata occurred 573–539 mya (mean 556 mya) in the late Precambrian, the radiation of Pancrustacea commenced during the Cambrian period (533–500 mya, mean 520 mya). The split between hexapods and branchiopod crustaceans occurred 521–489 mya (mean 510 mya). Ectognathan insects split from Entognatha 498–465 mya (mean 485 mya). Winged insects (Pterygota) emerged about 455 mya (468–436 mya) and about 64 myr later holometabolous (endopterygote) insects appeared (400–379 mya, mean 391 mya). Running the analyses under the priors showed that the calculated divergence times were not biased by the selected priors (data not shown).

**Table 2**

Pairwise differences of divergence time estimates with different molecular clock models (LOG, log-normal autocorrelated clock model; CIR, Cox–Ingersoll–Ross model; UGM, uncorrelated gamma multipliers). Rates across sites were modeled according to a gamma distribution with four categories ( $\Gamma$ ) or the Dirichlet process (D). Above the diagonal, absolute differences are given, below the diagonal are the mean differences. For the mean differences, positive values show older time estimates for the models in the horizontal row, negative values indicate younger dates.

	LOG $\Gamma$	LOG D	CIR $\Gamma$	CIR D	UGM $\Gamma$	UGM D
LOG $\Gamma$		0.004984	0.027751	0.028321	0.109461	0.109718
LOG D	0.003757		0.028923	0.029956	0.112200	0.112359
CIR $\Gamma$	−0.007211	−0.010987		0.003706	0.092873	0.090782
CIR D	−0.009793	−0.013702	0.002589		0.089166	0.089328
UGM $\Gamma$	−0.090723	−0.094126	−0.088827	−0.082692		0.001537
UGM D	−0.090971	−0.094367	−0.085352	−0.082959	−0.000293	



**Fig. 1.** Mean divergence times of major ecdysozoan taxa averaged over all estimations under the log-normal autocorrelated clock model shown in Supplemental Table S2. Gray bars indicate 95% mean confidence intervals (see Supplemental Table S3). Calibrated nodes are marked with an asterisk (see Supplemental Table S1 for calibration points). mya, million years ago.

### 3.3. Effect of root priors and calibration on time estimates

Bayesian time estimates were robust to changes in root priors, which determine age distributions and corresponding standard deviations for the root node. In a first approach, a uniformly distributed prior with an upper limit of the root of 5000 mya was assumed, as imposed by PhyloBayes (lane “–” in Supplemental Tables S2 and S3). To evaluate the impact of specific root priors, we defined different age means and standard deviations of the prior distribution: (a) 1000 mya and 1000 myr standard deviation, (b) 1000/500, (c) 750/750, and (d) 750/325. For all nodes, time estimates show low variation and differed – on average – less than 2% from the mean age with different root priors (standard deviation averaged over all nodes).

Standard deviations of the mean time estimates were lower with modeling the Dirichlet process (1.1%; root 1.8%) than with

gamma distributed rates across sites (2.1%; root 2.2%). The exclusion of the outgroup calibration point, *i.e.* the minimum age of 521 mya for the split between Mollusca and Polychaeta, resulted in slightly (4.1%) younger divergence time estimates (not shown). The mean age of the root, *i.e.* the origin of Ecdysozoa, was about 5.7% younger using the reduced set of calibration points. Averaged over all calculations, the ecdysozoan divergence dated to 573 mya.

### 3.4. Effect of substitution rates on time estimates

To assess the effect of variations of substitution rates among genes, the full 129 gene dataset was subdivided into three data subsets with 43 genes each. The mean PAM distances of the genes in subset *i* ranged from 0.01 to 0.17 (slowest substitution rate), in subset *ii* from 0.17 to 0.30 (intermediate rate) and in subset *iii* from 0.31 to 0.80 (fastest rate). Averaged over data subsets of the

autocorrelated LOG model, molecular clock analyses resulted in dates that were only ~0.1% younger compared to the complete dataset (Supplemental Table S6). In subset *i* the estimated dates were on average 2.9% younger (standard deviation 8.7%) as compared to the complete dataset. Subsets *ii* and *iii* showed on average slightly older dates (1.8% and 1.5%, respectively; standard deviation 6.5% and 6.9%, respectively). Most divergence times in the data subsets agree very well with the values derived from the complete dataset. Only two splits showed notable deviations. The split between Odonata and Ephemeroptera displayed high variation between subsets under all molecular clock settings, with subset *i* resulting in 22% younger dates and subsets *ii* and *iii* having 4% older dates. The results of the intermediate subset *ii* were similar with that of the entire dataset. The variations of divergence times of Isoptera and 'Blattodea' were even higher, with the intermediate subset *ii* resulting in 29% younger dates, whereas the slowly and fast evolving subsets showed older dates (subset *i*: 32%; subset *iii*: 35%).

### 3.5. Effect of missing data on time estimates

By removing 50% of the amino acid positions with the lowest taxon coverage, the relative amount of missing data in the alignment decreased from 51.0% to 38.5%. The effect of data reduction on the resulting divergence time estimates was minimal (Supplemental Table S7). Mean divergence times were 1.5% older than the results obtained from the complete dataset. Similar to the results from the subdivided datasets, the splits between Odonata–Ephemeroptera and Isoptera–'Blattodea' showed high variance. Under the autocorrelated clock models, divergence times of Odonata and Ephemeroptera were 7.7% older, whereas divergence time estimates of Isoptera and 'Blattodea' were 5.5% younger compared to the complete dataset.

## 4. Discussion

Molecular clock analysis has become a powerful tool based on a data source largely independent from the fossil record for the inference of divergence times of organisms. Still, there is much discrepancy between time estimates of different studies (e.g., Douzery et al., 2004; Peterson et al., 2008; Pisani et al., 2004). Factors that influence the outcome of molecular clock calculations are the sampling size, the selection of taxa and genes, rate heterogeneity, the suitability of the dating method, and the accuracy of calibration points. Simultaneous analyses of a large number of orthologous genes and application of multiple fossil calibration points provide more reliable estimates of divergence times if rate heterogeneity is considered (Thorne and Kishino, 2002; Yang and Yoder, 2003).

Due to the limited availability of orthologous genes, studies on large multi-gene datasets were usually restricted to only few taxa (e.g., Aris-Brosou and Yang, 2003; Blair and Hedges, 2005; Blair et al., 2005; Douzery et al., 2004; Gu, 1998; Wang et al., 1999). An alternative source of sequence data, which had not been applied to a molecular clock approach so far, is provided by EST data.

### 4.1. Applicability of EST data for molecular clock analyses

Our supermatrix with 129 orthologous genes and 117 taxa (Meusemann et al., 2010) is – to the best of our knowledge – the largest dataset that has ever been used for molecular clock studies (cf. Aris-Brosou and Yang, 2003; Blair and Hedges, 2005; Blair et al., 2005; Douzery et al., 2004; Gaunt and Miles, 2002; Gu, 1998; Lynch, 1999; Peterson et al., 2008; Regier et al., 2005; Wang et al., 1999). Our analyses are expected to provide more reliable estimates than the inference from few genes due to rate

homogenization (Battistuzzi et al., 2010; Thorne and Kishino, 2002; Yang and Yoder, 2003). In addition, the essentially stochastic nature of ESTs further is expected to reduce sampling bias caused by the selection of specific genes.

A potential drawback of our approach may be the fragmentary nature of ESTs, which is reflected by 51% missing data in the concatenated superalignment (Meusemann et al., 2010). It has been demonstrated that large datasets are less sensitive to missing data (Philippe et al., 2004). In fact, when we reduced the amount of missing data by removing positions with low coverage, thereby increasing data density, only minimal changes in divergence time calculations were observed (Supplemental Table S7). Only two splits showed notable variation (Odonata–Ephemeroptera and Isoptera–'Blattodea'). Therefore, molecular clock analyses of our EST-dataset were essentially robust to the effect of missing data.

Another factor that may influence the outcome of a molecular clock approach are differences in substitution rates between genes. The *a priori* stochastic approach of obtaining ESTs is expected to result in large variations of rates. This is reflected by the PAM distances of the individual proteins ranging from 0.01 to 0.80. However, splitting the dataset into three subsets with different evolutionary rates shows little variation in divergence times between the three estimates (Supplemental Table S6). Only for the two splits mentioned above (i.e. Odonata–Ephemeroptera and Isoptera–'Blattodea'), the variance was notably large, which may be due to a bias introduced by gene selection. Therefore, we can conclude that – at least if the datasets are large enough – the effect of substitution rate differences in the EST dataset is low.

### 4.2. Molecular clock models

The Bayesian relaxed clock model approach to the arthropod EST-derived tree was also robust to the choice of priors and parameter settings. Neither the root priors, which specify the age of the root and its standard deviation, nor the model for the site-specific rates (discrete gamma distribution with four rate categories or the Dirichlet process) had significant effects (Table 2; Supplemental Tables S2, S4 and S5). The main factor that actually influenced the time estimates was the applied clock model.

In recent years, a whole range of different molecular clock methods that either rely on a maximum likelihood approach or on Bayesian methods have been proposed (for review, see Lepage et al., 2007). In initial tests, we applied a maximum likelihood, local clock approach, as implemented in the program r8s (Sanderson, 2002) to our data. However, the resulting time estimates were highly dependent on the age of the root and resulted in unreasonable divergence times (data not shown). The dates that derived from the Bayesian models were more consistent and not mutually exclusive, but still showed large differences in the calculated divergence times. We applied three different relaxed Bayesian clock models, two autocorrelated (LOG and CIR) and the uncorrelated UGM model. But which model is correct, i.e. fits best to the data and is thus expected to provide the best time estimate?

Both cross-validation and Bayes factors showed that the autocorrelated clock models were significantly better than UGM. Autocorrelation assumes that adjacent branches in a phylogenetic tree evolve with a similar rate, while in an uncorrelated model the individual rates cluster around the mean. Thus the assumption of autocorrelation of rates in related species appears to be more realistic than averaging rates across the branches. Our results also agree with the study by Lepage et al. (2007), who demonstrated that autocorrelated models outperform uncorrelated models, particularly when the dataset is large. Cross-validation also found that the LOG model was slightly better than the CIR process. Therefore, we discuss below only divergence times averaged over all LOG clock model settings (Table 1), as displayed in Fig. 1.

### 4.3. Arthropod origins and age of major arthropod taxa

While the fossil record suggests the emergence of the Metazoa during the Cambrian period 542–488 mya (e.g., Chen et al., 2004; Chen, 2009; Conway Morris, 1993; Crimes, 1987; Harvey and Butterfield, 2008; Shu et al., 1996), most molecular clock studies estimated much older dates of up to 1200 mya (e.g., Blair et al., 2005; Feng et al., 1997; Hausdorf, 2000; Lee, 1999; Nei et al., 2001; Peterson et al., 2008; Wang et al., 1999). We obtained notably younger estimates in our studies, which, however, still do not agree with a Cambrian origin of metazoan phyla. For example, we dated the earliest divergence time within the Arthropoda 589 mya, while the first unambiguous arthropod fossils are 521 myr old (Chen, 2009; Crimes, 1987). The gap between a possible Precambrian emergence and the Cambrian metazoan fossils may be explained by a period of cryptic evolution or detection bias, e.g., due to largely unexplored Early Cambrian and Pre-Cambrian Lagerstätten (Benton and Ayala, 2003; Conway Morris, 1993; Fortey et al., 1996; Valentine et al., 1991).

Traditionally, tardigrades have been joined with the arthropods (Brusca and Brusca, 2003). Recent molecular studies suggested a sister group relationship of tardigrades with Cycloneuralia (nematodes and allies) (Bleidorn et al., 2009; Lartillot and Philippe, 2008; Meusemann et al., 2010; Roeding et al., 2007). However, this topology has been discussed as an artifact due to long branch attraction (Rota-Stabelli et al., 2011). Previous molecular clock studies estimated an early origin of tardigrades 813–670 mya (Regier et al., 2005; Sanders and Lee, 2010), while we calculated that the divergence of Nematoda and Tardigrada took place during the Ediacaran (~575 mya). Although our estimate is in better agreement with the Cambrian fossils of crown-group tardigrades (Müller et al., 1995), it should be considered with caution because of the uncertain tardigrade relationships.

The closest arthropod relative of the myriapods is uncertain. While some molecular studies either suggested a sister group relationship of Myriapoda and Pancrustacea (“Mandibulata”; e.g., Giribet and Ribera, 2000; Regier et al., 2010; Rota-Stabelli et al., 2011), others provided evidence for a common clade of Myriapoda and Chelicerata (“Myriochelata”; Pisani et al., 2004; Roeding et al., 2009). Because Meusemann et al. (2010) recovered Myriochelata in their Bayesian approach, this topology was assumed here although it may be an artifact (Rota-Stabelli et al., 2011). We inferred that Myriochelata and Pancrustacea diverged 562 mya. This is notably younger than previous calculations based on topologies supporting Myriochelata, which ranged from 672–642 mya (Pisani et al., 2004; Regier et al., 2005). The branch that joins Myriapoda and Chelicerata is comparatively short, corresponding to ~15 million years (myr) with a large confidence interval (Fig. 1; Supplemental Table S3). Thus, a rapid divergence of the three clades Myriapoda, Chelicerata and Pancrustacea may explain at least in part the problems associated with the relationships among these taxa.

There is still no conclusive myriapod record from the Cambrian, but presence of fossils from putative sister group taxa (Crustacea, Chelicerata) strongly suggests a Cambrian or earlier origin of Myriapoda (Shear and Edgecombe, 2010). While previous studies date the emergence of Myriapoda more than 600 mya (Otsuka and Sugaya, 2003; Pisani et al., 2004; Regier et al., 2005), our estimates of myriapod origin are comparatively young (~556 mya). Within the Myriapoda, our results showed an age for the split of Diplopoda and Chilopoda of ~504 mya, which is slightly older than previous molecular studies (e.g., 442 mya; Pisani et al., 2004) and the fossil record (~420 mya; Edgecombe and Giribet, 2007).

Most studies agree that Pycnogonida (sea spiders) represent the earliest branch within Chelicerata (Meusemann et al., 2010; Regier et al., 2010; Roeding et al., 2009). A larval sea spider from the upper

Cambrian (~500 mya) is the oldest fossil evidence for the split of Pycnogonida and Euchelicerata (Waloszek and Dunlop, 2002). Regier et al. (2005) suggested that this event took place 813–632 mya, but our calculation (~546 mya) is closer to the fossil record. Our estimate for the origin of Xiphosura (horseshoe crabs; ~473 mya) agrees well with the fossil dating, ~445 mya (Rudkin et al., 2008).

Traditionally, the divergence of Arachnida and Xiphosura has been considered the first split within the Euchelicerata (Regier et al., 2010; Weygoldt, 1998). However, several molecular studies did not recover monophyletic Arachnida, but suggest a basal position of the Acari (Meusemann et al., 2010; Roeding et al., 2007, 2009; Sanders and Lee, 2010). Given the low taxon sampling within the arachnids, it must remain uncertain which topology may reflect a true relationship or a long branch attraction phenomenon. This unresolved topology is the most likely explanation for the younger divergence times of Acari (~424 mya) and Araneae in other studies (~401–390 mya) (Aris-Brosou and Yang, 2002; Jeyaprakash and Hoy, 2009; Sanders and Lee, 2010).

The origin of the clade leading to Pancrustacea (‘Crustacea’ and Hexapoda) was previously estimated between 725 and 565 mya (Burmester, 2001; Otsuka and Sugaya, 2003; Pisani et al., 2004; Regier et al., 2005). Our results (~562 mya) are on the younger side. We estimated the divergence of the clade leading to the crustacean taxa Malacostraca, ‘Maxillopoda’ (Copepoda and Cirripedia) and Branchiopoda, and to the subphylum Hexapoda at ~520 mya in the early Cambrian. This timing is in line with recent findings of a crown group crustacean from the Mount Cap Formation 515–510 mya (Harvey and Butterfield, 2008).

The oldest known hexapods are collembolans (springtails) from the Lower Devonian ~400 mya (Kukalová-Peck, 1991). Previous molecular analyses estimated the split between crustaceans and hexapods (either Malacostraca–Hexapoda or Branchiopoda–Hexapoda) from 492–420 mya, but these studies relied on single or a limited number of genes (Burmester, 2001; Gaunt and Miles, 2002; Otsuka and Sugaya, 2003; Regier et al., 2005; Sanders and Lee, 2010). Based on an alignment of multiple genes, Pisani et al. (2004) proposed that the divergence of Hexapoda and Crustacea took place ~666 mya. Although our estimate (divergence of Branchiopoda and Hexapoda 510 mya) is closer to the hexapod fossil record, there is still a gap of ~100 myr. It must be considered that the true crustacean sister group of the Hexapoda is ambiguous. Recent studies have suggested that the enigmatic Remipedia may represent the closest living crustacean relatives of Hexapoda (Ertas et al., 2009; Regier et al., 2010). Unfortunately, fossil Remipedia are rare and ambiguous, and ESTs are currently not available for dating of the divergence from hexapods.

Within the insects, there is a general discrepancy between molecular time estimates and fossils. For example, we dated the divergence of Archeognatha and Pterygota ~455 mya, while first archeognath fossils derive from ~390 myr old (Labandeira et al., 1988) and first pterygotes from ~325 myr old strata (Prokop et al., 2005). Likewise, the time of the origin of Holometabola was estimated ~391 mya, while the first unambiguous fossils are ~307 mya (Béthoux, 2009). Our time estimates are actually close to previous calculations by (Gaunt and Miles, 2002), but the relatively large gap between molecular and fossil dating requires further investigations.

## 5. Conclusions

Although any molecular clock calculation for the inference of divergence times embraces problems beyond experimental control, we undertook measures to reduce potential errors to a minimum. The large amount of orthologous sequences from many

arthropod species, and the application of a relaxed Bayesian clock model using evenly distributed calibration points yielded consistent molecular divergence time estimates. Missing data had only minor effect on the estimation of divergence times highlighting the suitability of ESTs for molecular clock analyses. Likewise, selection of three data subsets (from fast, intermediate or slow evolving genes) and different model priors had only negligible influence. The application of different models (uncorrelated vs. autocorrelated models) had notable effects on divergence time calculations. Along with errors in calibration points, inappropriate data and similar problems, unsuitable models may explain in part the unreasonably early divergence times obtained in some previous molecular clock studies. Our approach resulted in divergence time estimates of the arthropods that are generally in much better agreement with the fossil record.

### Acknowledgments

We are grateful to Nicolas Lartillot for his help with PhyloBayes and J.-Wolfgang Wägele for his support. We are grateful to Falko Roeding for his help with the data and initial analyses, and Manfred Menning, Alexander Gubanov, Jacques Verniers and Chales H. Wellman for their support on stratigraphical issues. We thank Jakub Prokop for his advice on insect evolution. We acknowledge the associate editor and two anonymous reviewers for their helpful comments which improved the manuscript. This work has been supported by grants of the Deutsche Forschungsgemeinschaft within the Priority Project 1174 "Deep Metazoan Phylogeny" (Bu 956/8; Mi 649/6; Ha 1947/5; Wa 530/34).

### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ympev.2011.09.003](https://doi.org/10.1016/j.ympev.2011.09.003).

### References

- Adachi, J., Hasegawa, M., 1995. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. *J. Mol. Evol.* 40, 622–628.
- Aris-Brosou, S., Yang, Z., 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Syst. Biol.* 51, 703–714.
- Aris-Brosou, S., Yang, Z., 2003. Bayesian models of episodic evolution support a late precambrian explosive diversification of the Metazoa. *Mol. Biol. Evol.* 20, 1947–1954.
- Battistuzzi, F.U., Filipowski, A., Hedges, S.B., Kumar, S., 2010. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol. Biol. Evol.* 27, 1289–1300.
- Benton, M.J., Ayala, F.J., 2003. Dating the tree of life. *Science* 300, 1698–1700.
- Benton, M.J., Donoghue, P.C.J., 2007. Paleontological evidence to date the tree of life. *Trends Ecol. Evol.* 24, 26–53.
- Béthoux, O., 2009. The earliest beetle identified. *J. Paleont.* 83, 931–937.
- Béthoux, O., Nel, A., 2005. Some palaeozoic 'protorthoptera' are 'ancestral' orthopteroids: major wing braces as clues to a new split among the 'protorthoptera' (Insecta). *J. Syst. Palaeontol.* 2, 285–309.
- Blair, J.E., 2009. Animals (Metazoa). In: Hedges, S.B., Kumar, S. (Eds.), *The Timetree of Life*. Oxford University Press, New York, pp. 223–230.
- Blair, J.E., Hedges, S.B., 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* 22, 2275–2284.
- Blair, J.E., Shah, P., Hedges, S.B., 2005. Evolutionary sequence analysis of complete eukaryote genomes. *BMC Bioinform.* 6, 53.
- Bleidorn, C., Podsiadlowski, L., Zhong, M., Eeckhaut, I., Hartmann, S., Halaných, K.M., Tiedemann, R., 2009. On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol. Biol.* 9, 150.
- Brauckmann, C., Schneider, J., 1996. Ein unter-karbonisches Insekt aus dem Raum Bitterfeld-Delitzsch (Pterygota, Arnsbergium, Deutschland). *N. Jb. Geol. Paläont. Mh.* 1996, 17–30.
- Bromham, L., Rambaut, A., Fortey, R., Cooper, A., Penny, D., 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl. Acad. Sci. USA* 95, 12386–12389.
- Brusca, R.C., Brusca, G.J., 2003. *Invertebrates*. Second ed. Sinauer Associates, Sunderland, MA.
- Budd, G.E., Jensen, S., 2000. A critical reappraisal of the fossil record of the bilaterian phyla. *Biol. Rev. (Camb)* 75, 253–295.
- Budd, G.E., Telford, M.J., 2009. The origin and evolution of arthropods. *Nature* 457, 812–817.
- Burmester, T., 2001. Molecular evolution of the arthropod hemocyanin superfamily. *Mol. Biol. Evol.* 18, 184–195.
- Chen, J.Y., 2009. The sudden appearance of diverse animal body plans during the Cambrian explosion. *Int. J. Dev. Biol.* 53, 733–751.
- Chen, J.-Y., Bottjer, D.J., Oliveri, P., Dornbos, S.Q., Gao, F., Ruffins, S., Chi, H., Li, C.-W., Davidson, E.H., 2004. Small bilaterian fossils from 40 to 55 million years before the Cambrian. *Science* 305, 218–222.
- Collins, D., Rudkin, D.M., 1981. *Priscansermarinus barnetti*, a probable lepadomorph barnacle from the Middle Cambrian Burgess Shale of British Columbia. *J. Paleontol.* 55, 1006–1015.
- Conway Morris, S., 1993. The fossil record and the early evolution of the Metazoa. *Nature* 361, 219–225.
- Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Crimes, T.P., 1987. Trace fossils and correlation of late Precambrian and early Cambrian strata. *Geol. Mag.* 12, 97–119.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4, 699–710.
- Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Mol. Biol. Evol.* 101, 15386–15391.
- Dusar, M., 2006. Namurian. *Geol. Belg.* 9, 163–175.
- Ebersberger, I., Strauss, S., Haeseler, A., 2009. HamStr: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* 9, 157.
- Edgecombe, G.D., 2010. Arthropod phylogeny: an overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Struct. Dev.* 39, 74–87.
- Edgecombe, G.D., Giribet, G., 2007. Evolutionary biology of centipedes (Myriapoda: Chilopoda). *Annu. Rev. Entomol.* 52, 151–170.
- Engel, M.S., Grimaldi, D.A., Krishna, K., 2007. Primitive termites from the Early Cretaceous of Asia (Isoptera). *Stuttg. Beitr. Naturk. Ser. B (Geol. Paläontol.)* 371, 1–32.
- Ertas, B., von Reumont, B., Wägele, J.W., Misof, B., Burmester, T., 2009. Hemocyanin suggests a close relationship of Remipedia and Hexapoda. *Mol. Biol. Evol.* 26, 2711–2718.
- Feng, D.F., Cho, G., Doolittle, R.F., 1997. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. USA* 94, 13028–13033.
- Fortey, R.A., Briggs, D.E.G., Wills, M.A., 1996. The Cambrian evolutionary 'explosion': decoupling cladogenesis from morphological disparity. *Biol. J. Linn. Soc.* 57, 13–33.
- Gaunt, M.W., Miles, M.A., 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* 19, 748–761.
- Giribet, G., Ribera, C.A., 2000. A review of arthropod phylogeny: new data based on ribosomal DNA sequences and direct character optimization. *Cladistics* 16, 204–231.
- Graur, D., Martin, W., 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20, 80–86.
- Gu, X., 1998. Early metazoan divergence was about 830 million years ago. *J. Mol. Evol.* 47, 369–371.
- Gubanov, A.P., Peel, J.S., 1999. Oelandiella, the earliest Cambrian helcionelloid mollusc from Siberia. *Paleontology* 42, 211–222.
- Harvey, T.H.P., Butterfield, N.J., 2008. Sophisticated particle-feeding in a large Early Cambrian crustacean. *Nature* 452, 868–871.
- Hausdorf, B., 2000. Early evolution of the bilateria. *Syst. Biol.* 49, 130–142.
- Hedges, S.B., Kumar, S., 2003. Genomic clocks and evolutionary timescales. *Trends Genet.* 19, 200–206.
- Jarzewowski, E.A., Schneider, J.W., 2007. The stratigraphical potential of blattodean insects from the late Carboniferous of southern Britain. *Geol. Mag.* 144, 449–456.
- Jeyaprakash, A., Hoy, M.A., 2009. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp. Appl. Acarol.* 47, 1–18.
- Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286–298.
- Khomentovskiy, V.V., Karlova, G.A., 1993. Biostratigraphy of the Vendian–Cambrian beds and the lower Cambrian boundary in Siberia. *Geol. Mag.* 10, 29–45.
- Krzeminski, W., Krzeminski, E., Papier, F., 1994. The oldest Polyneura (Diptera) and their importance to the phylogeny of the group. *Acta Zool. Cracov.* 37, 95–99.
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., Reumont, B., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* 7, 10.
- Kukalová-Peck, J., 1991. Fossil history and the evolution of hexapod structures. In: Naumann, I.D. (Ed.), *The Insects of Australia*, vol. 1, second ed. Melbourne University Press, Carlton, Australia, pp. 125–140.
- Labandeira, C.C., Beall, B.S., Hueber, F.M., 1988. Early insect diversification: evidence from a Lower Devonian bristletail Québec. *Science* 242, 913–916.
- Lartillot, N., Philippe, H., 2006. Computing Bayes factors using thermodynamic integration. *Syst. Biol.* 55, 195–207.



- Lartillot, N., Philippe, H., 2008. Improvement of molecular phylogenetic inference and the phylogeny of bilateria. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1463–1472.
- Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288.
- Lee, M.S., 1999. Molecular clock calibrations and metazoan divergence dates. *J. Mol. Evol.* 49, 385–391.
- Lepage, T., Lawi, S., Tupper, P., Bryant, D., 2006. Continuous and tractable models for the variation of evolutionary rates. *Math. Biosci.* 199, 216–233.
- Lepage, T., Bryant, D., Philippe, H., Lartillot, N., 2007. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* 24, 2669–2680.
- Lynch, M., 1999. The age and relationships of the major animal phyla. *Evolution* 53, 319–325.
- Meusemann, K., Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walz, M., Pass, G., Breuers, S., Achter, V., Haeseler, A., Burmester, T., Hadrys, H., Wägele, J.W., Misof, B., 2010. A phylogenomic approach to resolve the arthropod tree of life. *Mol. Biol. Evol.* 27, 2451–2464.
- Minet, J., Huang, D.Y., Wu, H., Nel, A., 2010. Early Mecoptera and the systematic position of the Microptysmatidae (Insecta: Endopterygota). *Ann. Soc. Entomol. Fr.* 46, 262–270.
- Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* 58, 21–34.
- Müller, K.J., Walossek, D., Zakharov, A., 1995. 'Orsten' type phosphatized soft-integument preservation and a new record from the Middle Cambrian Kuonamka formation in Siberia. *N. Jb. Geol. Paläont. Abh.* 197, 101–118.
- Nei, M., Xu, P., Glazko, G., 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. USA* 98, 2497–2502.
- Nielsen, C., 2001. *Animal Evolution: Interrelationships of the Living Phyla*. Oxford University Press, Oxford.
- Otsuka, J., Sugaya, N., 2003. Advanced formulation of base pair changes in the stem regions of ribosomal RNAs; its application to mitochondrial rRNAs for resolving the phylogeny of animals. *J. Theor. Biol.* 222, 447–460.
- Peterson, K.J., Cotton, J.A., Gehling, J.G., Pisani, D., 2008. The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1435–1443.
- Philippe, H., Snell, E.A., Baptiste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 21, 1740–1752.
- Pisani, D., Poling, L.L., Lyons-Weiler, M., Hedges, S.B., 2004. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biol.* 2, 1.
- Prokop, J., Nel, A., Hoch, I., 2005. Discovery of the oldest known Pterygota in the Lower Carboniferous of the Upper Silesian Basin in the Czech Republic (Insecta: Archeoptera). *Geobios* 38, 383–387.
- Regier, J.C., Shultz, J.W., Kambic, R.E., 2004. Phylogeny of basal hexapod lineages and estimates of divergence times. *Ann. Entomol. Soc. Am.* 97, 411–419.
- Regier, J.C., Shultz, J.W., Kambic, R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc. R. Soc. Lond. B Biol. Sci.* 272, 395–401.
- Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463, 1079–1083.
- Roeding, F., Hagner-Holler, S., Ruhberg, H., Ebersberger, I., Haeseler, A., Kube, M., Reinhardt, R., Burmester, T., 2007. EST sequencing of Onychophora and phylogenomic analysis of Metazoa. *Mol. Phylogenet. Evol.* 45, 942–951.
- Roeding, F., Börner, J., Kube, M., Klages, S., Reinhardt, R., Burmester, T., 2009. A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Mol. Phylogenet. Evol.* 53, 826–834.
- Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G.D., Longhorn, S.J., Peterson, K.J., Pisani, D., Philippe, H., Telford, M.J., 2011. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc. R. Soc. Lond. B Biol. Sci.* 278, 298–306.
- Rudkin, D.M., Young, G.A., Nowlan, G.S., 2008. The oldest horseshoe crab: A new xiphosurid from Late Ordovician Konservat-Lagerstätten deposits, Manitoba, Canada. *Paleontology* 51, 1–9.
- Sanders, K.L., Lee, M.S.Y., 2010. Arthropod molecular divergence times and the Cambrian origin of pentastomids. *Syst. Biodivers.* 8, 63–74.
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109.
- Shcherbakov, D.E., 2000. Permian Faunas of Homoptera (Hemiptera) in Relation to Phytogeography and the Permo-Triassic crisis. *Paleontol. J.* 34 (Suppl. 3), S251–S267.
- Shear, W.A., Edgecombe, G.D., 2010. The geological record and phylogeny of the Myriapoda. *Arthropod Struct. Dev.* 39, 174–190.
- Shu, D.G., Conway Morris, S., Zhang, X.L., 1996. A Pikaia-like chordate from the Lower Cambrian of China. *Nature* 381, 157.
- Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702.
- Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647–1657.
- Valentine, J.W., Awramik, S.M., Signor, P.W., Sadler, P.M., 1991. The Biological explosion at the Precambrian Cambrian boundary. *Evol. Biol.* 25, 279–356.
- Waloszek, D., Dunlop, J.A., 2002. A larval sea spider (Arthropoda: Pycnogonida) from the Upper Cambrian 'Orsten' of Sweden, and the phylogenetic position of pycnogonids. *Paleontology* 45, 421–446.
- Wang, D.Y., Kumar, S., Hedges, S.B., 1999. 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.* 266, 163–171.
- Weygoldt, P., 1998. Evolution and systematics of the Chelicerata. *Exp. Appl. Acarol.* 22, 63–79.
- Whalley, P., Jarzembowski, E.A., 1981. A new assessment of Rhyniella, the earliest known insect, from the Devonian of Rhynie, Scotland. *Nature* 291, 317.
- Wilson, H.M., Anderson, L.I., 2004. Morphology and taxonomy of Paleozoic millipedes (Diplopoda: Chilognatha: Archipolypoda) from Scotland. *J. Paleontol.* 78, 169–184.
- Yang, Z., Yoder, A.D., 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* 52, 705–716.