universität wien

DIPLOMARBEIT

Did myriapods evolve from a single common ancestor? Optimizing the phylogenetic signal of 28S rRNA sequences for myriapods

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Verfasserin:	Daniela BARTEL
Matrikel-Nummer:	9802023
Studienrichtung /Studienzweig (lt. Studienblatt):	Biologie / Zoologie
Betreuer:	Ao. Univ.- Prof. Dr. **Günther PASS**

Wien, im Mai 2009

Contents

List of Figures

List of Tables

1 Summary

The present study assesses the ability of the ribosomal gene encoding for the 28S rRNA to resolve the position of myriapods inside the Euarthropoda as well as the earliest split events inside this problematic group from phylogenetic point of view.

This study is based on 26 myriapod sequences from 22 taxa (6 newly sequenced) covering all major subgroups of myriapods. Additionally, 16 representatives of Hexapoda, Crustacea, Chelicerata, and as outgroup *Milnesium tardigradum* (Tardigrada), were used to gain more insight. Besides the manual alignment strategy the present study investigates the reliability of a fully automated alignment strategy (`RNAsalsa`) and alignment masking (`ALISCORE`). Tree reconstruction methods encompass Maximum Parsimony, Maximum Likelihood, and Bayesian inference, as well as a tree reconstruction method, which accounts for the dependence of base substitution in stem regions of ribosomal genes (DNA/RNA mixed models) was conducted and implemented in a Bayesian framework.

The results reveal high statistical support for a number of commonly accepted monophyletic taxa, including Symphyla, Pauropoda, Chilopoda, Penicillata and Helminthomorpha, as well as Euchelicerata, Pancrustacea and Hexapoda. The monophyly of Myriapoda is highly dependent on the alignment strategy used to construct the data matrix. The monophyletic status of Diplopoda is not supported under most data sets, but clearly divided into two subgroups (Helminthomorpha and Penicillata). The position of myriapod subgroups inside the euarthropod tree remain instable through all performed data sets and tree reconstruction methods. Even with the correction for known problems in tree reconstruction, e.g. low taxon sampling, heterogeneity in base composition, or violation of the assumed character independence, an optimization of the phylogenetic signal of 28S rRNA for the reconstruction of the phylogenetic position of all myriapod subgroups seems hardly possible. An investigation of the secondary structure in myriapods could test whether the lack of unambiguous resolution can be explained by a deviation from the "standard euarthropod structure ".

2 Zusammenfassung

Die vorliegende Arbeit soll die Verwendbarkeit der ribosomalen RNA 28S zur Rekonstruktion der Myriapodenphylogenie hinsichtlich der Stellung innerhalb der Euarthropoda sowie der Verwandtschaftsbeziehungen zwischen den vier Myriapodengruppen bewerten.

Diese Studie basiert auf 26 Myriapodensequenzen von denen sechs neu amplifiziert und sequenziert wurden. Zusätzlich wurden insgesamt 16 Sequenzen von Vertretern der Hexapoda, Crustacea und Chelicerata, sowie *Milnesium tardigradum* (Tardigrada) verwendet. Neben der manuellen Alinierungsstrategie wurden zusätzlich eine vollständig automatisierte Alinierungs- und Merkmalsauswahlmethode (`RNAsalsa` und `ALISCORE`) auf ihre Eignung für die Euarthropodenphylogenie mittels 28S rRNA getestet. Als Methoden zur Baumrekonstruktion wurden Maximum Parsimony, Maximum Likelihood und Bayesian inference verwendet. Darüber hinaus wurde eine Bayesian Analyse durchgeführt, welche die Abhängigkeit von Basensubstitutionen in gepaarten Bereichen innerhalb der rRNA (stem regions) berücksichtigt (DNA/RNA mixed models).

Die Ergebnisse dieser Studie bestätigen eine Reihe von allgemein akzeptierten monophyletischen Gruppen, wie z. B. Symphyla, Pauropoda, Chilopoda, Penicillata und Helminthomorpha, sowie Euchelicerata, Pancrustacea und Hexapoda. Die Monophylie der Myriapoda hängt stark von der angewandten Alininierungsmethode ab. In den meisten Datensätzen sind die Diplopoda nicht monophyletisch, sondern werden in zwei getrennte Taxa aufgespalten (Penicillata und Helminthomorpha). Die Positionen der einzelnen Myriapodengruppen innerhalb der Euarthropoda sind in allen verwendeten Datensätzen und angewandten Baumrekonstruktionsmethoden instabil. Auch nach der Korrektur bekannter Probleme von verschiedensten Baumrekonstruktionsmethoden (zu geringes Taxon sampling, Heterogenität der Basenzusammensetzung, Verletzung der angenommenen Positionsunabhängigkeit innerhalb des Datensatzes) konnte das phylogenetische Signal der 28S rRNA zur Rekonstruktion der Verwandtschaftsverhältnisse der Myriapodengruppen nicht optimiert werden. Eine weiterführende Untersuchung der Sekundärstruktur der 28S rRNA von Myriapoden könnte Aufschluss darüber geben, ob diese von der allgemein verwendeten Euarthropoda Sekundärstuktur abweicht, und dadurch das geringe Auflösungsvermögen verursacht.

3 Introduction

The placement of the four subgroups of myriapods is hotly debated. The present study aims to elucidate the placement of these groups inside the Euarthropoda.

3.1 Background

The taxon Myriapoda comprises 13000 species worldwide, is a very diverse and ecologically important group of terrestrial arthropods. Traditionally, four groups have been united as Myriapoda: the chilopods (centipedes) and diplopods (millipedes) containing the vast majority of species, and the very poorly investigated species of pauropods and symphylans. As an ancient group, the myriapods inhabited Pangeae, resulting in their occurrence all over the world, except Antarctica, with the highest abundance in the tropics and subtropics. Even if the myriapods play an important role in terrestrial ecosystems, not much is known about their diversity and morphology likewise the phylogeny of myriapods is still a topic of open questions, both regarding their position within the euarthropods, as well as the earliest splits inside the myriapod lineage.

3.2 Myriapods placement within the Euarthropoda and the Monophyly versus paraphyly of Myriapoda

Before the advent of molecular phylogenetics, traditional morphological classifications differentiated two main lineages within the Euarthropoda, namely Chelicerata and Mandibulata, with Mandibulata comprising Crustacea and Atelocerata, the latter including Myriapoda and Hexapoda (Snodgrass 1938, Ax 1999, Bitsch & Bitsch 2004).

In the frame of this Mandibulata/Atelocerata hypothesis, several authors strongly advocated the paraphyly of myriapods and proposed either the Dignatha (Pauropoda + Diplopoda), the Progoneata (Dignatha + Symphyla), or Symphyla alone to be the sister group of Hexapoda (Pocock 1893, Snodgrass 1938, Sharov 1966, Kraus & Kraus 1994, Kraus 1998, Dohle 2001, Willmann 2003).

After decades of nearly universal acceptance, the Atelocerata hypothesis, sometimes also called Tracheata hypothesis, (e.g. Ax 1999, Bäcker et al. 2008), has been challenged starting with molecular studies based on ribosomal genes (Turbeville et al. 1991, Friedrich & Tautz 1995, Zrzavý & Štys 1997), which gave rise to two novel hypotheses. As a result of their molecular analyses Zrzavý & Štys (1997) introduced the clade Pancrustacea, lately, Dohle (2001) renamed this group Tetraconata. Compelling evidences for this clade result from investigations based on molecular data (Shultz & Regier 2000, Cook et al. 2001, Mallatt et al. 2004, Giribet et al. 2005, Mallatt & Giribet 2006, Regier et al. 2008, Bourlat et al. 2008, von Reumont et al. in review) as well as morphological data (Dohle 2001, Harzsch et al. 2005, Ungerer & Scholtz 2008). The myriapods are placed either together with the Pancrustacea, maintaining the taxon Mandibulata (e.g. Edgecombe 2004) or the phylogenetic position of the four subgroups remain instable inside the Euarthropoda tree (e.g. Giribet & Ribera 2000, Mallatt & Giribet 2006).

The second challenging hypothesis proposes a close relationship between Myriapoda and Chelicerata by uniting them in the taxon Myriochelata (Pisani et al. 2004) or Paradoxopoda (Mallatt et al. 2004). In contrast to the Pancrustacea, the Paradoxopoda has received support from molecular studies based on ribosomal (Mallatt et al. 2004), mitochondrial (Hwang et al. 2001, Pisani et al. 2004), and Hox genes (Cook et al. 2001), but still lacks imposing morphological support.

Despite the crucial role of myriapods within these conflicting hypotheses (Mandibulata versus Paradoxopoda), molecular analyses and in particular studies based on nuclear ribosomal RNA genes provide conflicting results (see Mallatt et al. 2004, Mallatt & Giribet 2006). Moreover, when dealing with 28S rRNA genes to infer phylogenetic relationships among the Euarthropoda, the taxon sampling among myriapods is exiguous (Giribet & Ribera 2000, Mallatt et al. 2004, Mallatt & Giribet 2006, Gai et al. 2006), particularly Pauropoda and/or Symphyla are often neglected (e.g. Friedrich & Tautz 1995).
In regard of all these ambiguities about myriapod phylogeny it is not surprising that Giribet (2001) point out in his article that " . . . the most bizarre case within arthropods (and perhaps for the entire Metazoa) are myriapods."

3.3 Relationships among and inside the myriapod subgroups

Regardless of these ambiguities, several authors suspect the Myriapoda as monophylum, and each of the four subgroup as monophyletic clade. Pure morphology-based studies consistently unite the Diplopoda with the Pauropoda in a clade Dignatha. Together with the Symphyla they build the clade Progoneata, and Chilopoda is considered as sister group to all myriapods (see Ax 1999, Edgecombe & Giribet 2002, Edgecombe 2004, Giribet et al. 2005). Considering the interrelationships inside each subgroup, the two best-investigated clades are Chilopoda and Diplopoda while Symphyla and Pauropoda received less attention and showed various placements in most molecular studies (Giribet & Ribera 2000, Regier et al. 2005b, Mallatt & Giribet 2006, Podsiadlowski et al. 2007, von Reumont et al. in review).

Chilopoda

The clade Chilopoda is composed of five extant orders (Scutigeromorpha, Lithobiomorpha, Geophilomorpha, Scolopendromorpha and Craterostigmomorpha). Almost all studies based on various data sets and methods agree that Chilopoda constitute a natural clade and consider each of the five centipede orders as monophylum. Additionally, most authors accept the basal split of the five subgroups between Notostigmophora (=Scutigeromorpha) and Pleurostigmophora (Kraus 1998, Giribet et al. 2001, Edgecombe & Giribet 2002, Regier et al. 2005b, Mallatt & Giribet 2006, Gai et al. 2006). Furthermore, relationships within Pleurostigmomorpha are well established in morphological consideration, with Lithobiomorpha as sister group to the remaining four orders (called Phylactometria). Inside the Phylactometria, Craterostigmomorpha represents the sister group to a clade composed of Scolopendromorpha and Geophilomorpha (the latter referred as Epimorpha) (see Dohle 1985). Inconsistencies result rather from investigations based on different genes than between morphology and molecular studies, such as nuclear ribosomal genes support morphology-based hypothesis (Edgecombe & Giribet 2002, Edgecombe 2004), whereas nuclear protein coding genes produce discordant results (Regier et al. 2005a).

Diplopoda

Currently, 16 extant orders are united in the clade Diplopoda. To date, the study of Sierwald & Bond (2007) is the most comprehensive phylogenetic analysis to solve relationships of diplopod orders based on morphological and molecular data. The authors recognize the basal position of Penicillata, followed by a clade of Pentazonia and Helminthomorpha (Chilognatha). These finding is in concordance of the division of Diplopoda from both earliest morphological (Enghoff 1984) and recently published molecular studies (Sierwald et al. 2003, Regier et al. 2005a).

3.4 Ribosomal gene encoding for the 28S rRNA

Ribosomal genes have become the most widely used nuclear molecular markers, which are mostly due to their ubiquity in all organisms and their presence in high copy number in each cell, facilitating amplification and sequencing. Different rates of evolution detectable inside and between these genes allow the inference of phylogenetic questions across a broad taxonomical range (Hillis & Dixon 1991).

The gene encoding for the 28S rRNA is composed of highly variable regions (Divergent Domains) nested inside conserved regions (Core Regions), in which different substitution rates and the probability of insertion and deletion events vary greatly. One fundamental property of the ribosomal RNA molecule is the formation of a complex secondary structure; due to the interaction of base pairs, building double-stranded stems between single-stranded regions. Since the secondary structure is dependent on Watson-Crick and Wobble base pairings, mutations on one side of the hemi helix of a stem region lead to a back mutation on the complementary side, leading to the preservation of both the secondary structure and function (Dixon & Hillis 1993). However, when compared across taxa, these so called compensatory mutations (Misof et al. 2007, Hancock et al. 1988) affect the primary sequences in a way that the sequence similarities decrease. Therefore the conservation of the secondary structure exceeds primary sequence conservation.

3.5 Structure based alignment strategies

Based on this structure-function dependence, it has been shown that secondary structure information can provide frame homology statements for positional homology in the alignment of rRNA molecules and the inclusion of this information in the alignment process might improve the accuracy of multiple sequence alignments (Kjer 1995). Therefore, the application of sequence-based alignment programs should not be considered, since they use scoring strategies that maximize sequence similarities, but entirely neglect structure information (Morrison 2006).

Dealing with long rRNA sequences, like the 28S, the prediction of secondary structure, either for single sequences or for a multiple alignment, appears as difficult task and the next paragraph is meant to illustrate current major problems.

During the last decades the 'comparative sequence analysis' approach was the only method to predict hypothetical consensus secondary structures and the determination of phylogenetic relationships for a set of homologous genes (Noller & Woese 1981, Gutell et al. 1985, Gutell & Cannone 2002). This method relied solely on the search for compensatory base changes to predict base-pairing interactions across a given taxa set. Recently, X-ray crystallography structures and NMR experiments of ribosomes have corroborated several rRNA structure models that were predicted by comparative analyses (Ban et al. 2000, Schluenzen et al. 2000, Yusupov et al. 2001).

Using the manual strategy, following the 'comparative sequence analyses' approach, the multiple alignment is generated by eye with reference to already published structural models obtained from RNA databases (e.g. RNAcomparative website) or secondary structure prediction of *Apis mellifera* (Gillespie et al. 2006) and *Saccharomyces cerevisae* (Spahn et al. 2001), and align these sequences to the reference sequence. This alignment procedure aligns and folds the sequences simultaneously and takes evolutionary information into account. During this process a consensus secondary structure is predicted under the available taxa set, and the researcher decides which characters are subsequently used in the phylogenetic analysis.

Nowadays, phylogenetic analyses consist of hundreds of sequences and the use of manual alignment strategies is no longer practicable. Recently, multiple alignment programs for structural RNA sequences have been developed, varying in the method how they implement secondary structures of the sequences and in the performance of the alignment (Siebert & Backofen 2005, Fontaine et al. 2008, Moretti et al. 2008). Major drawbacks of most of these programs are that they are not designed for long sequences. The usage of most programs is either bound to few input sequences or to a certain threshold of sequence length. Additionally, some programs need not only the sequences as input for the alignment process but also already predefined single secondary structure information from the primary sequences, which for most species are still unavailable.
The development of the software `RNAsalsa` (Stocsits et al. 2008) is the first attempt for long sequences to produce single secondary structure predictions of the input sequences to generate a multiple structure alignment based on both sequence and structure information of the individual sequences. The software produces a consensus secondary structure for a given data set, even for long ribosomal genes like 28S rRNA as input data. Furthermore, `RNAsalsa` takes into account sequence and structure information from an already published sequence for conserved regions and predicts variable regions of single sequences relying on thermodynamical folding algorithm, using the program `RNAfold` (Hofacker et al. 1994) and `RNAalifold` (Hofacker et al. 2002). For more details of the algorithm used in `RNAsalsa` I refer to Stocsits et al. (2008) and to the homepage of the Forschungsmuseum Alexander Koenig (Bonn, Germany): `http://rnasalsa.zfmk.de`

3.6 Aim of the study

For the present study, the choice of the ribosomal gene encoding for the 28S rRNA, as relatively slow evolving molecular marker, was based on its suitability in studies, which inferred deep level phylogenetic relationships (Mallatt et al. 2004, Mallatt & Giribet 2006, Gai et al. 2006, Dell'Ampio et al. 2009, von Reumont et al. in review). Apart from the publication of Gai et al. (2006), the present study was the first attempt to provide a comprehensive phylogenetic analysis of Myriapoda using the almost complete 28S rRNA gene with a broad taxon sampling covering all subgroups of the Myriapoda clade. In contrast to the study of Gai et al. (2006) the present study (i) enlarges the myriapod taxon sampling to a final amount of 26 myriapod sequences, (ii) tests for misleading effects of tree reconstruction by including/excluding characters, (iii) incorporates information of character interdependence in a Bayesian framework, (iv) and evaluates the data sets of non-homogeneous base frequencies across taxa.

The idea behind these approaches is to correct for already recognized misleading effects in reconstruction of phylogenetic relationship and in particular to deduce the position of myriapods inside the Euarthropoda tree. Phylogenetic tree reconstruction is performed with Maximum Parsimony and Maximum Likelihood, as well as with Bayesian inferences for DNA and mixed DNA/RNA substitution models. In this study sequences of the nuclear ribosomal 28S rRNA gene from 42 terminal taxa are used including almost all ordinal taxa of the myriapods along with sequences from Hexapoda, Crustacea, Chelicerata and one sequence of *Milnesium tardigradum* (Tardigrada).

4 Material and Methods

4.1 Taxa examined

The taxon sampling was focused on a comprehensive selection of specimens of the four myriapod subgroups. Therefore all 28S rDNA sequences of Myriapoda with sufficient length (at least 90% of the gene) were taken from NCBI `http://www.ncbi.nlm.nih.gov`. Additionally, to fill gaps in higher taxonomical groups inside the Myriapoda clade, sequences of the 28S rDNA from three Chilopoda (*Cryptops* sp., *Clinopodes flavidus* and *Craterostigmus tasmanianus*), three Diplopoda (*Polydesmus complanatus*, *Polyxenus lagurus* and *Glomeris hexasticha*) and one Pauropoda (Pauropoda sp.) were amplified and sequenced for the first time in this study. All specimens were either collected from our working group (Department of Evolutionary Biology, University of Vienna) or tissue or DNA was provided by cooperation partners (Forschungsmuseum Alexander Koenig, Bonn, working group Bernhard Misof; University of Shanghai working group Wen-ying Yin and University Rostock, working group Stefan Richter).

All in all the taxon sampling comprised almost the complete 28S rDNA sequences (60bp after the hypothetical beginning of the gene till 10bp before the Divergent Domain 12) of 26 individuals from 22 species of Myriapoda: ten Chilopoda, nine Diplopoda, five Symphyla and two Pauropoda were used as well as five Hexapoda, six Crustacea and five Chelicerata.[1]

All species used and corresponding accession numbers of the 28S rRNA sequences are given in Table 1. *Milnesium tardigradum* was used to root the remaining terminals in order to draw some conclusions about relationships across the euarthropods.

[1] The taxon sampling includes 26 individuals of 24 species; two individuals each were used from *C. tasmanianus*, *S. coleoptrata*, *Scutigerella* sp., and *Symphylella* sp.

Table 1. *Taxon sampling. Species indicated with ∗ were amplified and sequenced the first time; † indicates species, for which not the full sequences of the 28S rRNA are available in NCBI . '1' and '2' refer to sequences from two different individuals.*

Classification			Species	Accession nr.
MYRIAPODA	Symphyla	Scutigerellidae	*Scutigerella* sp.1	DQ666184
			Scutigerella sp.2 †	EF199982
			Hanseniella sp.	AY210821-22
		Scolopendrellidae	*Symphylella* sp.1	DQ666183
		Scolopendrellidae	*Symphylella* sp.2	Dell'Ampio (unpubl.)
	Chilopoda	Scutigeridae	*Scutigera coleoptrata*1	AY859601
			*Scutigera coleoptrata*2 †	EF199983
		Lithobiidae	*Lithobius forficatus* †	EF199984
		Cryptopidae	*Cryptpops* sp. ∗	EU376007
		Scolopendridae	*Cormocephalus hartmeyeri*	AY210812
			Otostigmus politus	DQ666180
			Scolopendra mutlilans	DQ666181
		Geophilidae	*Clinopodes flavidus*∗	EU376008
		Craterostigmidae	*Craterostigmus tasmanianus*1	AY859568-69
		Craterostigmidae	*Craterostigmus tasmanianus*2 ∗	EU376009
	Diplopoda	Julidae	*Cylindroiulus caeruleocinctus* †	EF199985
		Spirostreptidae	*Orthoporus* sp.	AY210827-28
		Polydesmidae	*Polydesmus complanatus* ∗	EU376010
			Cherokia georgiana	AY859562
		Paradoxosomatidae	Paradoxosomatidae sp. ∗	DQ666182
		Polyxenidae	*Monographis* sp. ∗	AY596389
			Polyxenus lagurus ∗	EU376011
			Polyxenidae sp.	AY859595
	Pauropoda	Pauropodidae	Pauropoda sp. ∗	EU346012
			Allopauropus sp.	DQ666185
CRUSTACEA	Branchiopoda	Triopsidae	*Triops* sp.	AY210844
	Malacostraca	Squillidae	*Squilla empusa*	AY210841-42
		Palinuridae	*Panulirus argus*	AY210832-35
	Maxillopoda	Argulidae	*Argulus* sp.	AY210804
		Cyclopidae	Cyclopidae sp.	AY210813
	Ostracoda	Cyprididae	Cyprididae sp.	AY210815
HEXAPODA	Archaeognatha	Machilidae	*Machilis hrabei*	EF199981
	Zygentoma	Lepismatidae	*Ctenolepisma longicaudata*	AY210810
	Neoptera	Mantidae	*Mantis religiosa*	AY859585
	Collembola	Sminthuridae	*Sminthurus viridis*	EF199973
	Ephemeroptera	Baetidae	*Callibaetis ferrugineus*	AY859557
CHELICERATA	Arachnida	Scorpionidae	*Pandinus imperator*	AY210830
		Eremobatidae	*Eremobates* sp.	AY859572
		Sironidae	*Siro rubens*	AY859602
	Xiphosura	Limulidae	*Limulus polyphemus*	AF212167
	Pycnogonida	Colossendeidae	*Colossendeis* sp.	AY210809
TARDIGRADA	Eutardigrada	Milnesiidae	*Milnesium tardigradum*	AY210826

Molecular procedure

4.2 DNA isolation

Tissue from collected material was frozen or ethanol preserved. Due to the small body size of Pauropoda sp. one entire individual was used for DNA extraction. Wherever applicable small pieces of tissue were taken from the abdomen or leg of the species. Extraction was performed using DNeasy®Tissue Kit (Qiagen) according to manufacturer's instructions. This method is based on a combination of enzymatic lyse of cells with proteinase K and subsequent purification and elution of genomic DNA by means of several centrifugation steps. All extraction samples were stored at -20°C.

4.3 Amplification and screening

4.3.1 PCR

The standard Polymerase Chain Reaction (PCR, Mullis et al. 1986) is a commonly used technique to amplify defined target DNA sequences within genomic DNA by in vitro replication. With this technique, only small quantities of template DNA is required to achieve a great number of copies of the desired DNA fragment. In addition to the DNA template a basic PCR solution requires several components, such as the heat-stable DNA polymerase and desoxynucleosidetriphosphates (dNTPs) to build the new DNA strand, a buffer solution to provide a suitable environment for the DNA polymerase, and two short oligonucleotides called primers, which are complementary to the 5' and 3' end of the target DNA.
The classic PCR profile consists of 20 to 35 cycles. Each cycle is composed of 3 different temperature changes: Denaturation, Annealing and Extension. During the Denaturation, at 95°C, all enzymatic activity is stopped and the genomic double stranded DNA disassociates in single strand hemi helices. In the Annealing step, at 45-55°C, the primers bind on the appropriate complementary single strand region of the DNA template forming a short fragment of double stranded DNA. During the Elongation step, at 72°C, the polymerase attaches to the short fragment of double stranded DNA, migrates on the template strand and connects the dNTPs on the growing newly formed daughter strand.
PCR was carried out in a thermocycler Primus 96 gradient from PeQLab. For the amplification of the fragments three different PCR procedures were used following the suggestion of Hillis et al. (1996).

Basic		**Touch down**			**Step up**		
95 °C	1'	95 °C	1'	95 °C	30"	95 °C	1'
45 °C	1'	56 °C	1' and	45 °C	30"	45 °C	30" and
72 °C	1'	72 °C	1'	72 °C	45"	72 °C	1'
25x		**10x**		**15x**		**15x**	**10x**

Exclusively, before the first Denaturation step a predenaturation of 5 minutes at 95 °C and a last extension step for 10 minutes at 72 °C and a terminal cooling at 4 °C were conducted. Amplification was carried out in 25μl working solution. Each reaction consisted of 13.912μl ddH_20, 2μl DNA, 2μl Primer Forward [10μM], 2μl Primer Reverse [10μM] (Sigma-Genosys), 2.5μl dNTPs [2.5mM

each] (Fermentas Life sciences), 2.5μl PCR Buffer (see Table 2), 0.088μl Taq-Polymerase [5u/μl] (GoTaq® Flexi DNA Polymerase; Promega). For the amplification of the 28S rDNA gene different primer pairs were used. The positions of these primers are shown in Figure 1; and sequences of the primers are listed in Table 3.

Table 2. *PCR buffer. Chemicals and volumes of the self-made PCR buffer with different concentration of $MgCl_2$.*

PCR buffer 10x			
Final $MgCl_2$	20	25	30
TrisHCl	$7ml$	$7ml$	$7ml$
BSA	200μl	20μl	200μl
$MgCl_2$	200μl	250μl	300μl
H_20	$2.6ml$	$2.55ml$	$2.5ml$
Final volume	$10ml$	$10ml$	$10ml$

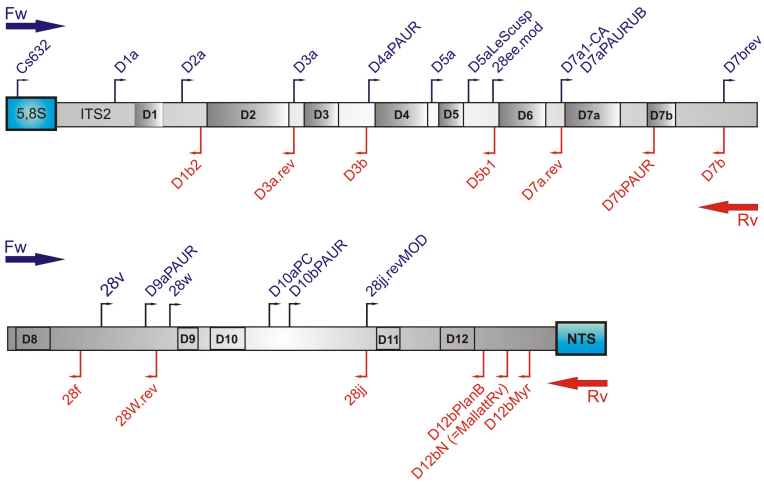

Figure 1. *Primer position. The transcription unit of the 28S rDNA is bounded by the ITS2 (Internal Transcribed Spacer 2) and on the other side from NTS (Non Transcribed Spacer). D1 till D12 indicates the 13 Divergent Domains of the 28S. Red arrows and red letters correspond to reverse primers, and blue arrows together with blue letters stand for forward primers, respectively. (changed after Dell'Ampio et al. 2009).*

4.3.2 Visualization and separation of DNA

Gel electrophoreses

The gel electrophoresis is a method for the separation of electric charged molecules according to their sizes by using an electric current. Due to the phosphate group in the DNA backbone, which carries negatively charged oxygen smaller fragments migrate faster inside the gel than larger ones to the anode. The separation of the fragments can be visualized under the UV screen.

A horizontal agarose gel electrophoresis was used to check the yields of all PCR and purification products. For fluorescent visualization of the fragments under the UV, 1% ethidium bromide solution (Boehringer Ingelheim) was added to the gel. The 1% agarose gel was embedded in 1x TBE-buffer (TBE 10x: TRIS 54g, boric acid 27.5g and 0.5M EDTA 20ml mixed with ddH_20 to 500ml). 5μl of individual PCR or purification product with about 3 μl 1:1 loading buffer (100mM

Table 3. *Primer. Sequences of the used primers; fw-forward, rv-reverse; ∗ indicates primer, which were newly designed in this study.*

Primer name	**Direction**	**Sequences**	**Specific for**
CS632	fw	CGATGAAGAACGCAGC	
D1a	fw	CCC(C/G)CGTAA(T/C)TTAAGCATAT	
D2a	fw	GATAGCGAACAAGTACC	
D3a	fw	GACCCGTCTTGAAACACGGA	
D4a.PAUR	fw	GTTCCTTCCGAAGTTTCC	Pauropoda ∗
D5a	fw	CTCAAACTTTAAATGG	
D5a.LeScusp	fw	TGGTAAGCAGGACTGG	
28ee.mod	fw	CCGCTAAGGAGTGTGTAAC	
D7a1	fw	CTGAAGTGGAGAAGGGT	
D7aCA	fw	CGATGTGGAGAAGGG	
D7a.PAURUB	fw	GCTGAAGTGGAGAAGG	Pauropoda ∗
D7b.rev	fw	ATGTAGGTAAGGGAAGTC	
28v	fw	AAGGTAGCCAAATGCCTCATC	
D9a.PAUR	fw	AATCAGCGGGGAAAG	Pauropoda ∗
28w	fw	CCT(G/T)TTGAGCTTGACTCTAATCTG	
D10aPC	fw	GGGGAGTTTGACTGGGGCGG	
28jj.rev.MOD	fw	AGGTTAGTTTTACCCTAC	
D1b2	rv	CGTACTATTGAACTCTCTCTT	
D3a.rev	rv	TCCGTGTTTCAAGACGGGAC	
D3b	rv	TCCGGAAGGAACCAGCTACTA	
D5b1	rv	ACACACTCCTTAGCGGA	
D7b	rv	GACTTCCCTTACCTACAT	
D7a.rev	rv	AAACCCTTCTCCACATCGG	
D7b.PAU	rv	ATCCTTTTCGCCGAAG	Pauropoda ∗
28f	rv	CAGAGCACTGGGCAGAAATCAC	
28W.rev	rv	CAGATTAGAGTCAAGCTCAACAGG	
28jj	rv	AGTAGGGTAAAACTAACCT	
D10b.PAUR	rv	ACCATTTGACAGATGTACCGCC	Pauropoda ∗
D12b.PLANB	rv	GAGTACGACACCCC	
D12b.MYR	rv	GTTGGTGGCTGCTCTAC	Myriapoda
Mallat.Rv1	rv	ACTTTCAATAGATCGCAG	

EDTA, 43% glycerol, 0,025% brome phenole blue and 0,025% xylene cyanol) was loaded in the gel. Additionally, the molecular weight ladder GeneRulerTM DNA Ladder Mix (Fermentas) was added to the electrophoresis to quantify the length of the fragments. The runtime of each electrophoresis was approximately 20 minutes using a current of 120 Volt.

Purification

After the amplification a purification procedure was conducted following the protocol of E.Z.N.A.®Gel Extraction Kit (Classic Line) from PeQLab. The principle of this purification method is based on the linkage of DNA fragments on a silica gel membrane. Through several washing steps all undesired components are washed out and the desired DNA fragment is eluted with a clearly defined amount of ddH_20. To remove multiple unspecific bands, the complete PCR product was loaded on a gel using approximately $10\mu l$ loading buffer. The runtime of this electrophoresis was 90 minutes at 70Volt. The desired band was excised from the gel and cleaned following the protocol of E.Z.N.A.®Gel Extraction Kit (Classic Line). To measure the concentration of the purified DNA fragments a gel electrophoresis was made using MassRulerTM DNA Ladder, Mix (ready to use, Fermentas) for the quantification. This ladder provides a combination for measuring the length of the DNA fragment, in base pairs, and to quantify the concentration of DNA in the sample in

ng/μl. The purified samples were stored at -20°C.

4.3.3 DNA sequencing

Cycle sequencing reactions are usually based on the modified version of the chain-terminator method developed by Sanger et al. (1977). In a Cycle sequencing reaction, copies of the DNA fragment are produced in a similar fashion as a normal PCR. Significant differences to the PCR are the presence of only one primer in each Cycle sequencing reaction and in addition to dNTPs, the solution contains four different fluorescent-labelled didesoxynucleosidetriphosphates (ddNTP) each emit light of a different wavelength after activation. These ddNTPs or 'termination nucleotides' are lacking of a 3'-OH group, which is essential to build the next phosphodiester bond. If a fluorescent-labelled ddNTP binds on the growing strand, the DNA strand extension is terminated. The results are DNA fragments of different lengths. The subsequent acryl amide capillary electrophoresis separates the produced fragments depending on their dimension along the electrical potential towards the anode. Before the fragments complete the run and diverge into the buffer, a laser activates the fluorescence of the different dye-labelled ddNTPs at the end of each DNA fragment. According to the colour of the fluorescence, the equivalent base can be identified. An adjacent spectrophotometer detector converts the emission peaks into a digital signal, which is stored as electropherogram. The sequences can be visualized and edited by special computer programs.

For the sequencing reaction BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used, providing all required components apart from template and template-specific primers. The Cycle sequencing reactions were performed in 10μl solutions (1μl DNA, 1μl Primer, 1μl BigDye mixed with 7μl ddH_20). Condition for the sequence reaction was an initial Denaturation step at 96°C for 5 minutes, followed by 25 cycles of 20 seconds at 96°C, 10 seconds at 50°C and 4 minutes at 60°C. Cooling was performed at 4°C. The products of purification were sequenced on a sixteen capillary sequencer ABI PRISM® 310x1 Genetic Analyzer at the Department of Evolutionary Biology University of Vienna. Resulting electropherograms were visualized in ChromasLITE 2.01 `http://www.technelysium.com.au`. Sequences from the forward and reverse strand were assembled manually in BioEdit Sequence Alignment Editor (Hall 1999).

Phylogenetic reconstruction

4.4 Alignment

When dealing with sequence data the establishment of positional homology hypotheses is made during the alignment process. Alignment approaches can be divided into two main categories: (i) manual alignments and (ii) computer based alignments (Kjer et al. 2007).
In this study alignments were performed using one method from each category (for more details of the alignment procedure see below). Since secondary structure information of rRNA genes can improve sequence alignments (Kjer 1995) in both approaches the predicted secondary structure of the 28S rRNA from *Apis mellifera* (after Gillespie et al. 2006) was incorporated in each alignment process. Conclusions about evolutionary processes are dependent on accurately aligned sequences. Hence, ambiguously aligned sections in the alignment, which presumably contain no phylogenetic signal, must be eliminated before tree reconstruction. Different methods are proposed for the identification of randomly similar sections in the alignment. This is done usually by the exclusion of ambiguous positions (alignment masking), either manually, based on the judgements of the investigator, or by formal algorithms like `Gblocks` (Castresana 2000) or `ALISCORE` (Misof & Misof in press).

To test for differences of those two character exclusions strategies, two different methods were performed and tested in term of their accuracy in subsequent phylogenetic reconstruction on the basis of the phylogeny of Myriapoda:

- Manual alignment and exclusion of ambiguously aligned positions 'by eye'
- Computer based alignment performed with the test version `RNAsalsa`.0.5.8 (Stocsits et al. 2008) and character exclusion according to the suggestion of `ALISCORE`.0.4.3 (Misof & Misof in press).

4.4.1 Manual alignment

Positional homology among nucleotides was assessed by eye following the secondary structure prediction of *Apis mellifera* (Gillespie et al. 2006). All Core Regions as well as parts of the Divergent Domains, which were aligned unambiguously were included in the final matrix. The alignment procedure was performed with BioEdit Sequence Alignment Editor (Hall 1999). During the alignment procedure all dimensional motifs (stem, loops, bulge, pseudoknots . . .) of the consensus structure were indicated following a modified version of DCSE alignment format. At the end of the process the consensus structure was converted into the Vienna style Dot Bracket format (Hofacker 2003).

4.4.2 Computer based alignment with `RNAsalsa`

`RNAsalsa` is a recently published alignment program for structural RNA sequences (developed by Roman Stocsits, University of Leipzig, in cooperation with the work group of Bernhard Misof, Forschungsmuseum Koenig, Bonn, Germany), which is able to align ribosomal RNA sequences, by implementing comparative evidence algorithms as well as thermodynamic folding. Outputs of this computer program are individual predicted secondary structures from all input sequences, a consensus secondary structure in dot & bracket format, and a final alignment. It should be mentioned that the results obtained with alignment based on `RNAsalsa` has to be taken with

caution, simple because all tree reconstruction methods were applied with data sets obtained with a test version of RNAsalsa (RNAsalsa v0.5.8). Currently, RNAsalsa is available in the improved beta version RNAsalsa 0.7.4 and is still under development and not all results will remain reproducible.

Prealignment with MUSCLE v3.6
A prealignment is needed as input for RNAsalsa; therefore all sequences as well as the sequence of *Apis mellifera* were prealigned in MUSCLE v3.6 (Edgar 2004b,a) using default settings.

Alignment with RNAsalsa
Alignment was performed using the default parameter values for substitution costs, gap penalties and base pairing occurrence stringency (-S1: 0.6; -S2: 0.6; -S3: 0.6; gap-opening: -8; gap-extension: -4). A constraint sequence with primary and secondary structure information are required as input for RNAsalsa. In this study 28S and 5.8S was used as input for RNAsalsa. The secondary structure prediction of *Apis mellifera* (Gillespie et al. 2006) was used as constraint. The predicted structure of these rRNAs shows not only differentiation of stem and loop regions but likewise includes pseudoknots, which are long-range interactions within the molecule. The current version of RNAsalsa is not able to consider information concerning pseudoknots. To avoid possible formation artifacts during the alignment performance these sites were marked in the constraint and therefore RNAsalsa does not try to fold these positions. Five stems (H15, H31, H35, H46, and H150 according to the notation of (Gillespie et al. 2006) are builded up by interactions of sequence portions of the 28S with complementary strands from the 5.8S. These stems were likewise blocked to avoid folding artefact, and during the alignment process these positions were considered as unpaired.

4.5 Evaluation of the alignment

Prior to phylogenetic reconstruction ambiguously aligned positions and highly variable regions of the alignment should be eliminated from the data set to avoid misleading phylogenetic signals.

1. The manual alignment was screened for ambiguously aligned regions, which were excluded manually ('subjective'). In all subsequent sections all data sets based on this manual approach are abbreviated to MA.

2. For the formal approach ALISCORE v0.4.3 (developed by Bernhard Misof and Patrick Kück, Forschungsmuseum Alexander Koenig, Bonn, Germany) was chosen to identify randomness in the alignment generated with RNAsalsa v0.5.8, using default settings. ALISCORE identifies randomly similar sections in multiple sequence alignments based on pairwise comparison within a sliding window and a Monte Carlo resampling approach. Details about ALISCORE can be read in Misof and Misof (in press) or in the Manual, which can be downloaded directly with the software from the homepage of the Alexander Koenig Museum, Bonn, Germany. http://www.zfmk.de/web Default settings were used (window size w=4, gaps are treated as ambiguities -N, pairwise comparison -r=4 x number of present taxa in the alignment).
   In subsequent sections all generated data sets based on this automated approach are referred to as AA.

4.6 Data sets

On the basis of the above described alignments and character exclusion modes different data sets were constructed. In the next section these data sets are explained in detail and based on these data sets performed analyses are summarized in Table 4 at the end of this section.

4.6.1 Full character set

For each of the two alignment and character choice strategies (data sets: MA and AA) all unambiguously aligned positions of the gene were considered. These full data sets are composed of Core Regions as well as alignable parts of the Divergent Domains.

4.6.2 Reduced character set

To test for effects of character choice, subsequently a reduced character set was generated, which includes only Core Regions, according to the notation of Gillespie et al. (2006).

Tree reconstruction

4.7 Basics in phylogenetic reconstruction and dealing with rRNA genes

During the last decades numerous tree reconstruction methods were established for the analyses of molecular sequence data. Basically, all these methods can be divided into two main classes: (i) distance based methods like Neighbour-Joining (NJ) (Saitou & Nei 1987), and (ii) character based methods like Maximum Parsimony (MP) (Farris et al. 1970), Maximum Likelihood (ML) (Felsenstein 1981) and Bayesian inference. All these methods except for MP are based on explicit models of sequence evolution also named distance correction methods (for more details see section 'Models of sequence evolution and MrModeltest'). Most nucleotide substitution models were developed for DNA sequences, assuming that each nucleotide site evolved independently from all other sites. If dealing with RNA sequences this main assumption of character independence is clearly violated. RNAs build conserved secondary structures containing both paired regions (stems), and unpaired regions. Nucleotides in stem regions do not evolve independently from their counterpart. Point mutation in a given stem region can destroy Watson-Crick base pairings and may lead to the instability of a functionally important helix. Therefore Compensatory substitutions - where mutation at one site in a stem region implicates a second mutation in the corresponding site - maintain base pairing and the structure of the RNA are restored. When covariant sites are treated as independent characters, their phylogenetic information is scored two times leading to wrong inferences in the analyses (Jow et al. 2002). To account for covariation, the implementation of secondary structure information is required. Standard DNA models can be used for loop regions, and specific RNA substitution models for stem regions (mixed DNA/RNA models).
Schöninger & von Haeseler (1994) developed RNA models and are implemented in the software Mr.Bayes. MP and ML analyses were performed using PAUP*4.0B10 (Swofford 1998). In addition Bayesian analyses were conducted with the program MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003).

4.8 Phylogenetic tree reconstruction with DNA model

4.8.1 Maximum Parsimony (MP)

Maximum Parsimony algorithm (Farris et al. 1970) searches for tree(s), which require the minimum number of evolutionary changes to reconstruct the evolutionary history of the data set, and therefore is the preferred tree. Each nucleotide site of the alignment is treated as one character with four different possible character states A, T, G, C. Gaps can be treated either as missing or as fifth character state. MP considers every site directly and reconstructs the evolution of every site on a tree that comprises the fewest evolutionary changes. The total number of evolutionary changes on a tree, referred as tree length, is the sum of the number of changes at each site. Theoretically the best option to achieve the optimal tree is to use exact methods, such as searching for the best tree by evaluating every possible tree upon an underlying data set; however this can only be performed for small data sets due to the long computational time (Swofford 1998). Therefore heuristic methods are employed for larger data sets. These strategies do not explore all possible trees but only a subset of possible trees. Typically a heuristic search follows a two-step process, starting with an initial tree, which is constructed by randomly adding one taxon after another, attaching each taxon to the branch that yields the optimum tree at each step. In a second step the tree is rearranged using special branch swapping algorithms.

In the present study, for the MP tree construction the stepwise addition method was employed to obtain starting trees. Using stepwise addition, it is possible that the algorithm gets stuck in a so-called 'local optimum' and is not able to find the 'global optimum', which is the best tree under the 'optimality criterion'. A branch-swapping algorithm the TBR (Tree Bisection and Reconnection) is used to elude this problem. In this method the tree is bisected into two sub trees and scrutinized along another pair of branches. Each rearrangement is evaluated to achieve at the global optimum (Swofford 1998).

The following parameters were used for the heuristic search under the parsimony criterion with DNA model:

- Gaps treated as missing data
- Exclusion of parsimony-uninformative characters and missing data
- Starting trees obtained via stepwise addition
- Heuristic search with random sequence addition, nreps=1000
- One tree retained at each step
- Branch-swapping algorithm is tree-bisection-reconnection (TBR)

4.8.2 Bootstrap analyses

A non-parametric bootstrap test (Felsenstein 1985) is performed to test the statistical support for internal nodes of a phylogenetic tree. From the original matrix (alignment) positions are repeatedly randomly chosen to achieve multiple pseudo-replicates (bootstrap replicates) of the same size as the original alignment. The result is a new data set that contains some sites with multiple occurrences while others are not represented. Each pseudo-replicate is evaluated under the parsimony principle. The bootstrap analysis under MP was run with heuristic search and random sequence addition

for 1000 pseudo-replicates. Resulting bootstrap trees were summarized in a 50% majority rule consensus tree.

4.8.3 Consensus trees

Consensus trees summarize different topologies received by a phylogenetic analysis. Commonly used consensus trees are the majority rule and the strict. The strict consensus tree includes only those internal nodes that are present in all calculated topologies. In a majority rule consensus tree all nodes are included, which are present in a user-defined percentage of the individual topologies. After each MP analysis the most parsimonious trees were summarized to 50% majority rule consensus trees.

4.8.4 Fit measurement of homoplasy

When positions between two sequences are corresponding in base content there are two possible explanations: (1) correspondence is based on homology and derived from a common ancestor, or (2) correspondence is based on homoplasy and therefore not derived from a common ancestor. Several indices have been suggested for the estimation of the degree of homoplasy inside a given data sets and therewith for the reliability of calculated topologies. In this study two indices were performed to estimate the reliability for the MP-trees: Consistency Index (CI) (Kluge & Farris 1969) and the Retention Index (RI) (Farris 1989). The CI represents the relative rate of those characters, where no 'homoplasy' occurs. To calculate the CI the formula

$$CI = M/S$$

can be used. The less 'homoplasy' occurs, the closer the CI converges to 1. RI compares the number of observed homoplasy with the theoretically largest data set. With the formula

$$RI = (G - S)/(G - M)$$

the retention index can be calculated.
'M' indicates the minimal amount of substitution within all possible trees
'S' specifies the amount of substitution in the given tree
'G' is the minimum amount of substitution in all possible trees

For all trees under the MP criterion the CI and the RI were calculated.

4.8.5 Models of sequence evolution and MrModeltest

Considering that possible multiple hits in sequences are common, observed distances may underestimate the actual number of evolutionary changes. To overcome this problem a considerable number of 'correction'-methods has been developed. These methods differ in their assumption regarding the rate of the two different transitions and the different transversions as well as in the proportion of the four bases in the data set (Page & Holmes 1998). The simplest model is represented by the Jukes-Cantor-Model (JC) (Jukes & Cantor 1969), and is based on the assumption that all substitutions have an equal probability and the frequency of all nucleotides is the same.

The Hasegawa-Kishino-Yano-85 (HKY85) (Hasegawa et al. 1985) assumes no equal base frequencies and accounts for the difference between transitions and transversions in one parameter. The most complex method is the General-Time-Reversible-Model (GTR) (Rodríguez et al. 1990), which adopts a symmetric substitution matrix (substitutions from A to G happen at the same rate as from G to A) though each pair of nucleotide substitution has different rates, the base composition is not equal and each base has another occurrence frequency.

Homogeneity of rate of evolution among site is one of the implicit assumptions under these models. Albeit, they ignore multiple changes on fast evolving sites in the alignment. The incorporation of heterogeneity of evolutionary rates among sites yielded in the implementation of two further parameters: gamma-distribution of among-site rate variation (Γ) and proportion of invariable sites (I) in the alignment.

The program MrModeltest v2.2 (Nylander 2004) in conjunction with PAUP*4.0B10* is besides the Bayes factor test (Kaas & Raftery 1995, Nylander et al. 2004) one alternative to perform an approximation of the best-fit model for a given data set. The combination from PAUP and MrModeltest estimates the best-fit model in a three-step process: PAUP generates a neighbour joining tree of the data set and subsequently estimates the likelihood scores for 24 models of nucleotide substitution. The produced batch file is then implemented in MrModeltest and a particular model is selected under two different frameworks for model selection, (1) hierarchical likelihood ratio test (hLRT) and (2) Akaike Information Criterion (AIC).

For estimating the appropriate model of sequence evolution for all analyses in the DNA approach, the Akaike criterion (AIC) was carried out as implemented in the program MrModeltest 2.2.

4.8.6 Maximum Likelihood (ML)

The ML (Felsenstein 1981) method is based on statistical calculations and besides MP, a method that uses an 'optimality criterion'. The topology that has the highest likelihood to produce the data is preferred over all other possible trees. The formal description for ML is

$$LD = Pr(D|M, \theta, \tau, \nu)$$

which is the probability of obtaining the data (observed sequences) given a model of evolution (M), a vector of k model parameters (θ), a tree topology (τ) and a vector of branch length (ν). The overall likelihood of obtaining the data is the sum of the log likelihoods of each individual site. As described above, ML incorporates a number of explicit models of sequence evolution, which determine the relative rate of substitutions and give different degrees of freedom to shape parameters.

In ML, heuristic searches with random sequence addition were performed. As a result of the high computing time replicate number was restricted to 500. Additionally, to decrease the computing time, the option 'reconlimit=7' was used and parameters were set according the suggestion of MrModeltest.

4.8.7 Bayesian inference (BI)

This relatively new approach for tree reconstruction simultaneously estimates trees and obtains measurements of uncertainty for each resulting clade. It combines the advantage of using a user

defined substitution model and a fast tree-topology algorithm (Felsenstein 2004). In Bayesian statistics the goal is to obtain a full probability distribution over all possible parameter values. The optimal topology is the one that maximizes the posterior probability. The formula of Bayesian theorem for phylogenetic analyses is as following:

$$Posterior\ probability = \frac{Likelihood \times prior\ probability}{Sum\ of\ all\ hypotheses}$$

The denominator in Bayesian inference involves summing up all possible trees and in addition to each tree, all possible combinations of branch lengths and substitution model parameter values must be calculated. This is hardly possible to compute and impossible to perform analytically (Huelsenbeck et al. 2002).

Basically the analysis works as follows: MrBayes is using **M**arkov **C**hain **M**onte **C**arlo simulation (MCMC) (Metropolis et al. 1953, Hastings 1970) to search for possible trees. MCMC uses stochastic simulation to obtain a sample from the posterior distribution of trees. In every generation the algorithm generates a new topology with slightly different prior probabilities. As default in MrBayes two independent Markov chains were run simultaneously, each starting with different random trees. First a random tree topology with random prior probabilities of the parameters, like branch lengths, base frequencies, substitution model parameters, and shape parameter of the gamma distribution is proposed as starting point for each Markov chain. Every Markov chain is run over several thousand generations; in each generation a new tree is proposed by changing the parameters, which is evaluated with respect to the former topologies according to the posterior probability and is performed with the Metropolis-Hasting algorithm. To avoid the local optimum problem, MrBayes uses the method of Metropolis coupled MCMC. Beside the two Markov chains referred as cold chains, it includes three heated chains for each Markov chain. These heated chains can more readily find new hills. Sometimes the cold chain exchanges states with one heated chain and therefore can jump to a new hill. The first trees produced in the two runs are referred as start-up-phase. This so-called burn-in phase produces only sub optimal trees with low likelihood values and must be excluded.

In this study MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003) was used applying the GTR+I+Γ model of nucleotide substitution. Model parameter values were treated as unknown and were estimated and optimized separately in each analysis.In the present version, MrBayes is not able to treat gaps as fifth character state and therefore they were treated as missing data. All priors were set according to the GTR+I+Γ model:

lset nst=6, rates=invgamma,nucmodel=4by4

prset revmatpr=dirichlet, statefreqpr=dirichlet, shapepr=uniform, pinvarpr=uniform

The parameters for the mcmc analyses: Each Markov chain, three heated (temp=0.2) and one cold, was started from a random tree and all four chains (nchains=4) ran simultaneously for one million generations (ngen=1000000). Trees were sampled every 1000^{th} generation (samplefreq=1000). The burn-in phase was determined after visualization of the log likelihood values against the generations in Microsoft Excel (Windows). All trees inside the burn-in phase were discarded. The remaining trees were summarized in a majority rule consensus tree.

4.9 Phylogenetic tree reconstruction with RNA model

4.9.1 Bayesian inference

Data partition

As mentioned above, treating covariant characters as single and independent characters may lead to wrong inferences in the tree reconstruction. Mixed models were applied to avoid this problem. For Bayesian inference with mixed DNA/RNA models the sequence data was partitioned into unpaired (loops) and paired (stem) positions. Therefore, loop regions can be treated under DNA models and stem regions under RNA models to consider co-variation. The PERL script `'RNARecode.pl'` was used to generate the two data partitions (developed by Bernhard Misof). This tree reconstruction approach was applied only on the full character set for the manual alignment.

Bayes factor test

A Bayes factor test was applied to choose the best evolutionary model for the mixed model approach (Kaas & Raftery 1995, Nylander et al. 2004). Different models were compared to each other: GTR and HKY85 substitution models were tested for the unpaired positions and corresponding GTR and HKY85 for doublet models. The latter represent two of the Schöninger & von Haeseler (1994) DNA/RNA models, which are implemented in MrBayes. Gamma-distributed among-site rate variation (Γ) was assumed in all models.

Testing two different models for the unpaired and paired sites results in four different combinations:

1. Test1: HKY+Γ/HKY+Γ (loop/stem)
2. Test2: GTR+Γ/HKY+Γ
3. Test3: HKY+Γ/GTR+Γ
4. Test4: GTR+Γ/GTR+Γ

For each of the four combinations two Markov chains (one cold chain and three heated chains) were run for one million generations. Trees were sampled every 100^{th} generation. The burn-in phase was determined after visualization of the log likelihood values of the cold chains against the generations in Microsoft Excel (Windows), and all trees inside this burn-in phase were discarded. For each model the harmonic mean was calculated using the likelihood values of both cold chains. All four resulting harmonic means were pairwise compared following the indication of Kaas & Raftery (1995):

$$2lnB10 = 2(lnL1 - lnLo)$$

Lo = model with the lowest ln likelihood (=harmonic mean)
L1 = comparative model.

All Bayesian analyses were executed with MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003), which copes with DNA/RNA, mixed models of substitution simultaneously.

Tree reconstruction with mixed DNA/RNA model

After choosing the best model of sequence evolution, the main analysis was run for five million generations with a sampling frequency of every 100^{th} tree. After the analysis, trees inside the burn-in phase were excluded, and the remaining was summarized to a consensus tree.

Table 4. *Performed analyses.*

Full character set		**Reduced character set**	
Manual approach	Automated approach	Manual approach	Automated approach
MP	**MP**	**MP**	**MP**
(DNA model)	(DNA model)	(DNA model)	(DNA model)
Modeltest	**Modeltest**	**Modeltest**	**Modeltest**
ML	**ML**	-	-
BI	**BI**	**BI**	**BI**
(DNA model)	(DNA model)	(DNA model)	(DNA model)
Bayes Factor Test	-	-	-
BI	-	-	-
(mixed model)			

4.10 Nucleotide frequency estimation

All tree reconstruction methods, like MP, Bayesian inference, and ML assume homogeneity of nucleotide frequencies across the data set. When the data set is heterogeneous in base composition, it can lead to inaccurate trees (Omilian & Taylor 2001). To evaluate if the produced data sets showed significant heterogeneity of base composition, χ^2-tests were performed. In all data sets not alignable sites, missing, constant, and ambiguous characters were excluded and subsequently a χ^2-tests were performed in PAUP*4.0B10.

4.11 Visualization of the trees

Trees were imported in MrEnt v1.2.1 (Zuccon & Zuccon 2006) and subsequently edited.

5 Results

5.1 Sequences

In this study fragments of seven myriapod species are newly sequenced, additionally sequences from 35 myriapods and other arthropod sequences are obtained from NCBI. The alignment starts directly after stem H183 (55bp after the hypothetical beginning) and ends after stem H2630 (8bp before D12). During the manual alignment it became evident that the sequence of *G. hexasticha* shows a highly divergent sequence compared to the remaining taxa. Due to this reason the sequence of *G. hexasticha* is discarded a priori from the phylogenetic reconstruction. Apart from the sequence of *G. hexasticha* all remaining taxa exhibit high variability in sequence length (Table 5), spanning from 5049bp from *C. flavidus* to 3331bp from Cyclopidae sp. Insertion sites are not restricted exclusively to the Divergent Domains but two large insertion events are found in the bulge region of stem H1457, and between stem H1405b and H1405a both in the two sequences of *Symphylella* sp. The large size of the 28S rRNA sequence of *C. flavidus* could not be traced back to a certain region; in *C. flavidus* insertion events seem to have occurred nearly in all Divergent Domains.

Table 5. *Sequence length in base pairs. For a better comparison of the length variation all sequences span from stem H183 till the second hemi helix of stem H2630. Notice that, the sequences of M. tardigradum and Argulus sp. are shorter and end by stem H2640 and H2735, respectively. Abbreviations: M=Myriapoda; CH=Chelicerata; C=Crustacea, H=Hexapoda.*

Taxon	**Sites**	**Taxon**	**Sites**
C. flavidus (M)	5049	*P. lagurus* (M)	3652
P. argus (CR)	4241	*Monographis* sp. (M)	3647
Symphylella sp. 1 (M)	4179	*S. coleoptrata* (M)	3621
Symphylella sp. (M)	4178	*C. tasmanianus* 1 (M)	3604
Hanseniella sp. (M)	4173	*C. longicaudata* (HE)	3604
Scutigerella sp. 1 (M)	4168	*L. forficatus* (M)	3603
Pauropoda sp. (M)	4084	Cyprididae sp. (CR)	3596
Scutigerella sp. (M)	4020	*S. empusa* (CR)	3596
Allopauropus sp. (M)	4003	*C. tasmanianus* (M)	3587
Cryptops sp. (M)	3927	*Argulus sp.* (CR)	3584
Paradoxosomatidae sp. (M)	3908	*C. ferrugineus* (HE)	3573
P. complanatus (M)	3902	*M. tardigradum*	3566
C. georgiana (M)	3871	*S. viridis* (HE)	3539
O. politus (M)	3844	*Colossendeis* sp. (CH)	3534
S. mutilans (M)	3840	*Eremobates* sp. (CH)	3523
C. hartmeyeri(M)	3801	*L. polyphemus* (CH)	3481
Orthoporus sp. (M)	3780	*P. imperator* (CH)	3479
C. caeruleocinctus (M)	3724	*S. rubens* (CH)	3477
S. coleoptrata 1 (M)	3717	*M. hrabei* (HE)	3452
M. religiosa (HE)	3673	*Triops* sp. (CR)	3448
Polyxendidae sp. (M)	3657	Cyclopidae sp. (CR)	3331

5.2 Nucleotide frequency and χ^2 test

Tables in Appendix show the proportion of A, C, G, T for the full and reduced character set as well as the χ^2 test of homogeneity of base frequencies across taxa for the MA and AA, respectively. When comparing the two alignments and final data sets (manual alignment versus `RNAsalsa` and

full character sets versus reduced character sets) no significant differences can be seen in base frequencies. The relation between A/T and G/C is about 0.37:0.63 (in *Scutigerella* sp. 1 for the full character set MA) and 0.49:0.51 (in Cyprididae sp. for the reduced character set MA and AA). All character sets are checked for heterogeneity of base compositions in PAUP. The results lead to the acceptance of the null hypothesis (H_0), homogeneous base composition among taxa in all character sets (Table 6). For these tests not alignable positions, missing, ambiguous, and constant sites are excluded.

Table 6. *χ^2 test. Statistical outcome for the χ^2 test of homogeneity of base composition across taxa for all four data sets.*

	Full character set		Reduced character set	
Data sets	Manual approach (MA)	Automated approach (AA)	Manual approach (MA)	Automated approach (AA)
Incl. characters	952	841	764	648
χ^2	130.94	97.33	87.26	70.69
p	0.30	0.96	0.99	1.00
df	123	123	123	123

5.3 Data sets

5.3.1 Full character set (A)

The manual alignment comprises 5917 characters; of which 3348 (56.6%) ambiguously aligned characters are manually excluded. The alignment from `RNAsalsa` comprises 5872 positions, from which `ALISCORE` scores 3248 (55.3%) sites as randomly similar. After the cleaning of the alignments the full character set 'manual approach' consists of 2569 positions, whereas the full character set 'automated approach' retains 2624 positions for the subsequent phylogenetic tree reconstructions. (Figure 2)

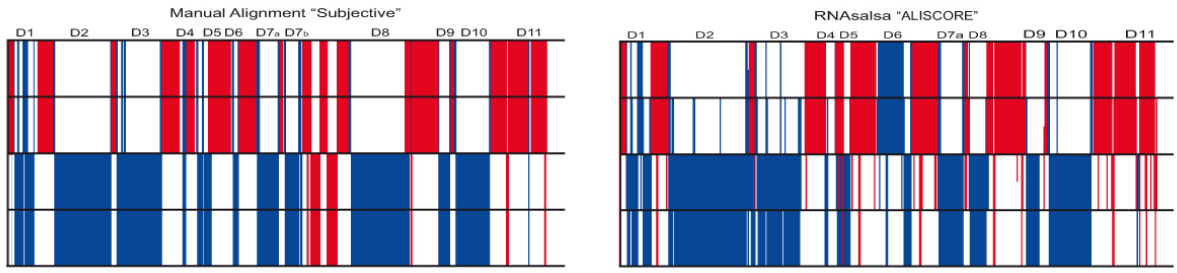

Figure 2. *Character exclusion for the full character set. Visualization of the results of randomly similar aligned positions, according to the two different full character sets. Positions with negative score are excluded in subsequent tree reconstruction.*
The full manual alignment consists of 5917 positions, whereas the alignment generated with `RNAsalsa` consists of 5872 positions (x-axis).
left: By the 'subjective' exclusion of characters 3348 positions are not alignable (negative values on the y-axis)
right: For the fully automated approach `ALISCORE` detects 3248 randomly similar sites (y-axix).
D1 till D12 mark the position of the Divergent Domains, which are coloured in blue. Core Regions, which are nested inside the variable regions are coloured in red.

5.3.2 Reduced character set (B)

To test for the influence of a restricted character choice, from both full character sets the Divergent Domains are discarded, resulting in two reduced character sets composed of 2287 and 2126 positions for the manual approach and for the automated approach, respectively.

5.4 Tree reconstruction

5.4.1 Maximum Parsimony

After exclusion of missing, ambiguous and parsimony uninformative characters the matrices include 706 and 610 positions for the manual approach (MA) and for automated approach (AA), respectively.

MP trees for the full character set
For the MA 32 most parsimonious trees with a tree length of 3195 steps (CI=0.40; RI=0.57) are summarized in a 50% majority rule consensus tree. The MP analysis for AA yielded in 3 most parsimonious trees (tree length=2476; CI=0.42; RI=0.57) that are combined in a 50% majority rule consensus tree (see Figure 3).

MP trees for the reduced character sets
540 and 447 parsimony informative characters for the MA and AA are included in the MP analyses. The MP analyses found nine most parsimonious trees with 2318 steps (CI=0.40; RI=0.59) with the MA, and five most parsimonious trees for AA (tree length=1716 steps; CI=0.42; RI=0.59) (see Figure 9).

5.4.2 Maximum Likelihood and Bayesian inference

For both tree reconstruction methods only missing and ambiguous characters are excluded and results in 2155 and 2008 characters for the data sets MA and AA for the full character set, whereas the reduced character set from the MA consists of 1905 and from AA of 1747 included positions.

The estimation of the appropriate model of sequence evolution with the AIC criterion in MrModeltest reveals the GTR+I+Γ as best fitting model for all tested character sets. (Table 7) The ML analyses was performed only for the full character set and resulted in one most likely tree with a log-likelihood value of 19194.89 for the data set manual approach (Figure 4), and a log-likelihood value of 16051.57 for the data set based on the automated approach (Figure 5).

5.4.3 Bayesian inference (mixed DNA/RNA model)

This analysis was performed only on the full character set with the manual approach (Figure 8), in which the consensus secondary structure consists of 1374 paired and 1198 unpaired positions. This information is required for the implementation of mixed models in Bayesian analysis. In this analysis the unpaired regions are treated with DNA substitution models, whereas paired regions are treated with RNA substitution models. A Bayes factor test was used for the identification of the optimal combination of substitution models. As default in MrBayes v3.1.2 two independent Markov chains run simultaneously and the harmonic means of the negative log-likelihood values of these 2 runs for each of the 4 different evolutionary models was calculated (Table 8). The best

Table 7. *Estimated parameters and model selection with the AIC criterion for ML.*

Parameters according to the model selection from MrModeltest2.2		
	Manual approach	Automated approach
Model	GTR+I+Γ	GTR+I+Γ
Basefrequency	A=0.25	A=0.26
	C=0.24	C=0.23
	G=0.30	G=0.30
	T=0.21	T=0.20
Relative rate matrix	1.14	0.91
	3.02	2.16
	1.41	1.16
	0.77	0.58
	6.15	5.50
Rates	gamma	gamma
Shape (α)	0.6969	0.7094
Pinvar	0.3998	0.4004
Nr. of rate categories	4	4
Score of best tree	19194.89	16051.57

fitting model was chosen applying the formula of Kaas & Raftery (1995). Under this data set the Bayes factor test identified the GTR+Γ for both substitution models as optimal model.

Table 8. *Harmonic means of the neg. ln-Likelihood values for the four different tested combinations.*

Model	**DNA model** unpaired	**RNA model** paired	**Harmonic mean**
Model1	HKY+Γ	HKY+Γ	22723.877
Model2	GTR+Γ	HKY+Γ	22652.047
Model3	HKY+Γ	GTR+Γ	22725.534
Model4	GTR+Γ	GTR+Γ	22652.075

5.5 Resulting topologies

5.5.1 Myriapoda

For both strategies of character choice, all performed phylogenetic tree reconstruction methods are congruent in the discovery for the monophyletic status of Pancrustacea, Eucheliceratа, Symphyla, Pauropoda as well as Scolopendromorpha, Helminthomorpha and Pentazonia, with high to moderate statistical support (Table 9 and 10). Remarkably, in all performed analyses the monophyly of Myriapoda is only revealed in topologies based on the data set manual approach (except for the Bayesian analysis, mixed models) and never appears by the Automated approach. However, the monophyly of Myriapoda receives in the two MP analyses (full and reduced character sets) and in the Bayesian analysis (reduced character set) low support (bootstrap $\leq$50%, pp=0.69), whereas the Bayesian analysis with the full character set receives maximal support pp=1.00.

The monophyletic status of the Chilopoda is achieved from data sets (MA and AA) under the full character sets. However, three reduced character sets do not support this hypothesis. In all three topologies, either *C. flavidus* is outside the Chilopoda and builds a sister group relationship to a clade composed of Symphyla and Pauropoda (MA with MP, Figure 9A), or Pauropoda is nested inside the Chilopoda (AA with MP and Bayesian inference, Figure 9B and 11). Only one analyses (full character set MA with MP) recovers the basal split between Geophilomorpha and Heteroterga (Ax 1999). However, the competing hypothesis, which assumes the basal split between Scutigeromorpha (Notostigmomorpha) and Pleurostigmomorpha, is substantiated in all of the performed analyses with the data sets based on Automoted approaoch (note, that in both reduced character sets the two Pauropoda sequences are nested inside the Chilopoda).
Trees, which are generated with the manual approach, show varying results. Only one analyses support this division (reduced character set, but without *C. flavidus*, Figure 9 A), whereas all other analyses suggest either *C. tasmanianus* (full character set for ML, Figure 4 and both Bayesian analyses, Figure 6 and Figure 8), or *C. flavidus* as basal group Ax (hypothesis of 1999); Figure 3A) within the Chilopoda, or this splitting is not resolved (Figure 10). Inside the Chilopoda the four sequences of Scolopendromorpha always cluster together and in the majority of the analyses *L. forficatus* forms the sister group. As mentioned before, *C. tasmanianus* as sole representative of the Craterostigmomorpha, and *C. flavidus* as single representative of Geophilomorpha appear in various positions inside the Chilopoda.

The widely morphological accepted monophyletic taxon Diplopoda is not found in most of the performed analyses. Exclusively, in the Bayesian inference 'mixed models' the Diplopoda is considered as monophylum (pp=0.96) as well as in the topology derived from MP for the reduced character set with `RNAsalsa` and `ALISCORE` (9B), but the bootstrap support is below 50%. Though, the clade Diplopoda is clearly divided into two clades, Helminthomorpha and Penicillata, each strongly supported as monophylum (bootstrap values $\leq$99% and pp=1.00). However, the position of these two groups is instable and cannot be located unambiguously.

Throughout all topologies, the placement of Pauropoda and Symphyla is instable. The Symphyla clusters with Pauropoda in all but one MP analyses as well as in the Bayesian analysis, 'mixed models'. By data sets of the MA in all model based methods (ML and Bayesian inference) Pauropoda is recovered as sister group to Penicillata and the Symphyla appears as sister group to the Helminthomorpha. This is partly true for analyses with the AA (Table 9 and Table 10).
It should also be mentioned, that the widely morphological accepted clade Dignatha, grouping Diplopoda and Pauropoda, is not found as monophyletic assemblage across all produced topologies. The Progoneata clade occurs in all but one model-based analyses with the data set MA (full character set, ML and Bayesian inference with DNA model and reduced character set, Bayesian inference) (Table 9 and Table 10).

Another noteworthy point is the sister group relationship of the sequence *Scutigerella* sp. 1 (Gai et al. 2006) with the sequence of *Hanseniella* sp. (with high statistical support; bootstrap=100% and pp=1.00). Never, the two sequences of *Scutigerella* sp. cluster together. In comparison the two sequences of *Symphylella* sp. as well as the two sequences of Pauropoda always cluster together with high nodal support (bootstrap values=100% and pp=1.00). In all topologies, the *Scutigerella* sp. is grouped with both sequences of *Symphylella* sp., and is separated from the remaining Scutigerellidae

(*Hanseniella* sp. and *Scutigerella* sp.)

5.5.2 Euarthropod relationships

The taxon Hexapoda appears only in MP analyses as paraphylum (full character set and reduced character set with both approaches, Table 3 and Table 9), due to the fact that *S. viridis* shows in all these topologies a sister group relationship to the maxillopod Cyclopidae sp. In all other analyses the monophyly of Hexapoda receives strong to moderate statistical support (pp=0.92 till pp=1.00). In no analyses the Crustacea appears monophyletic, unlike the Pancrustacea clade, which units the hexapods and crustaceans, are highly supported in all analyses (pp=1.00 and bootstrap values $\geq$95%, except for the MP search from the reduced character set from AA, in which this clade receives only a bootstrap value of 75%). The clade Chelicerata, which unites the pycnogonide *Colossendeis* sp. with the Euchelicerata, is supported under the two model-based methods (ML and Bayesian inference) with the data set AA as well as with MA performed Bayesian analysis under 'mixed models', whereas by all other methods *Colossendeis* sp. shows various placements on the trees. In contrast all analyses support a monophyletic Euchelicerata with moderate to high statistical support (Table 9 and 10).

All tree-building methods are not able to give some conclusions about deep nodes of the Euarthropoda tree, like favouring either the Mandibulata- or Paradoxopoda-hypothesis. The deep nodes are poorly resolved, which can be seen in all topologies with the automated approach (AA), all exhibits either polytomies or short branches, and receive no statistical support values.

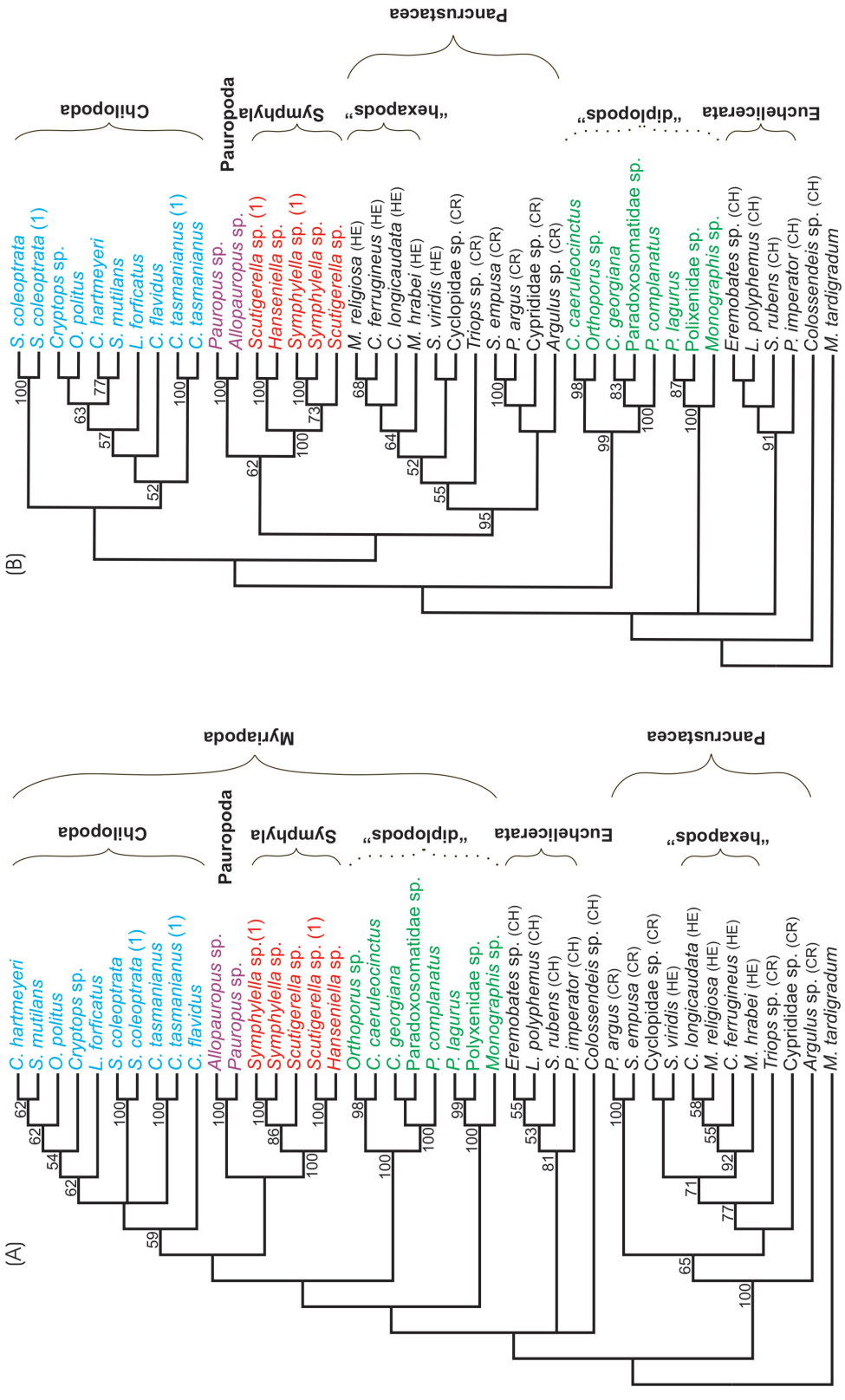

Figure 3. *50% majority rule consensus of most parsimonious trees for the two full character sets with bootstrap support above 50% indicated at the corresponding nodes. Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; clades under " indicates a paraphyletic assamblage. (A) manual approach: consensus cladogram of 32 most parsimonious trees with a tree length of 3195 steps; CI:0.40; RI:0.57; (B) Automated approach: consensus cladogramm of 3 most parsimoniuous trees with a tree length of 2476 steps; CI:0.42, RI:0.57.*

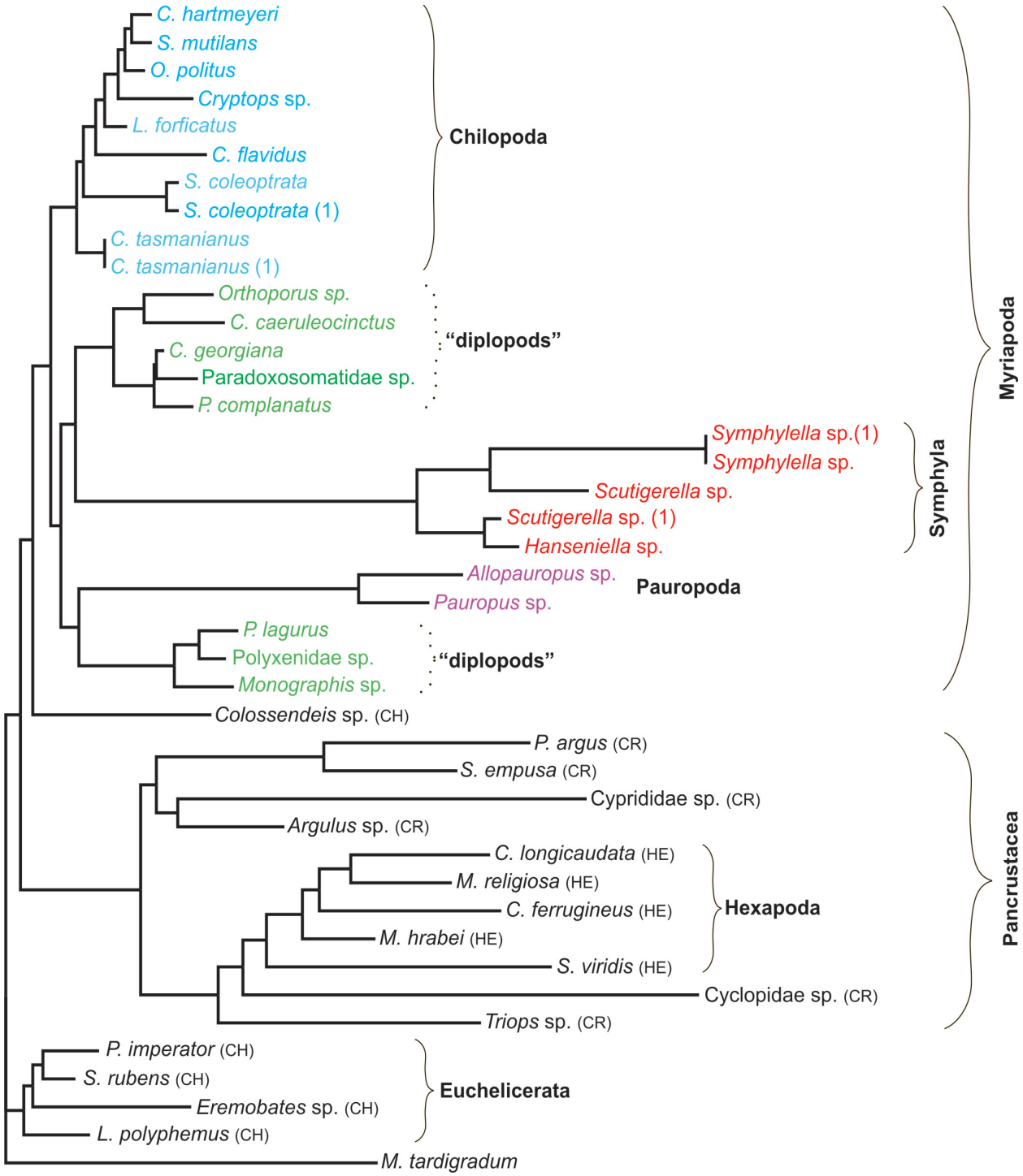

Figure 4. *Maximum Likelihood tree (2155 included characters) based on the manual approach. Tree and branch lengths were calculated by the ML search. Parameters were set according to the suggestion of MrModeltest (Table 7); −lnL of the best tree: −19194.89.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; "diplopods" indicate paraphyletic diplopods.

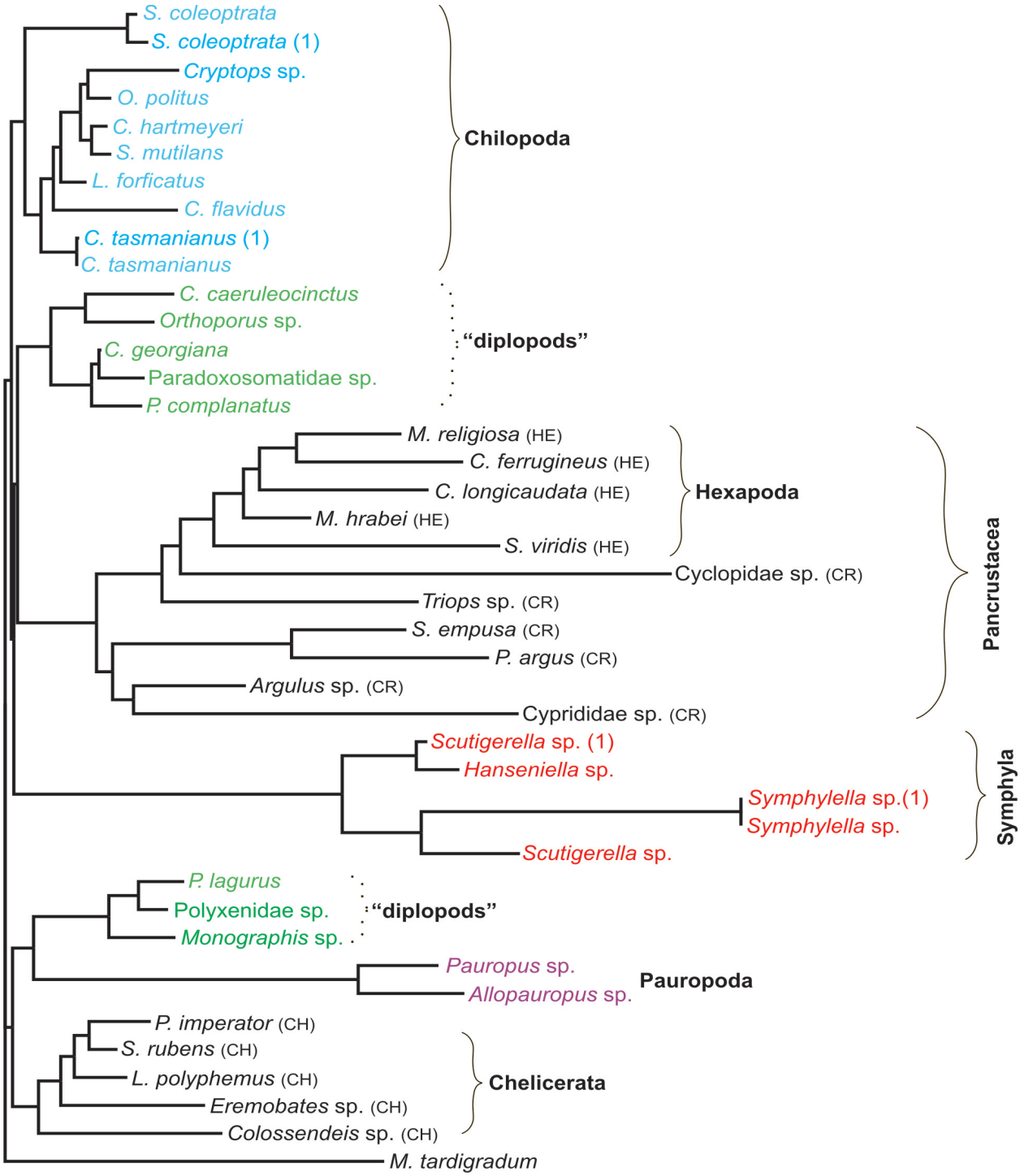

Figure 5. *Maximum Likelihood tree from the full character set (2008 included characters) based on the automated approach. Tree and branch lengths were calculated by the ML search. Parameters were set according to the suggestion of MrModeltest (Table 7); −lnL of the best tree: −16051.57.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; clades under " indicates a paraphyletic assamblage.

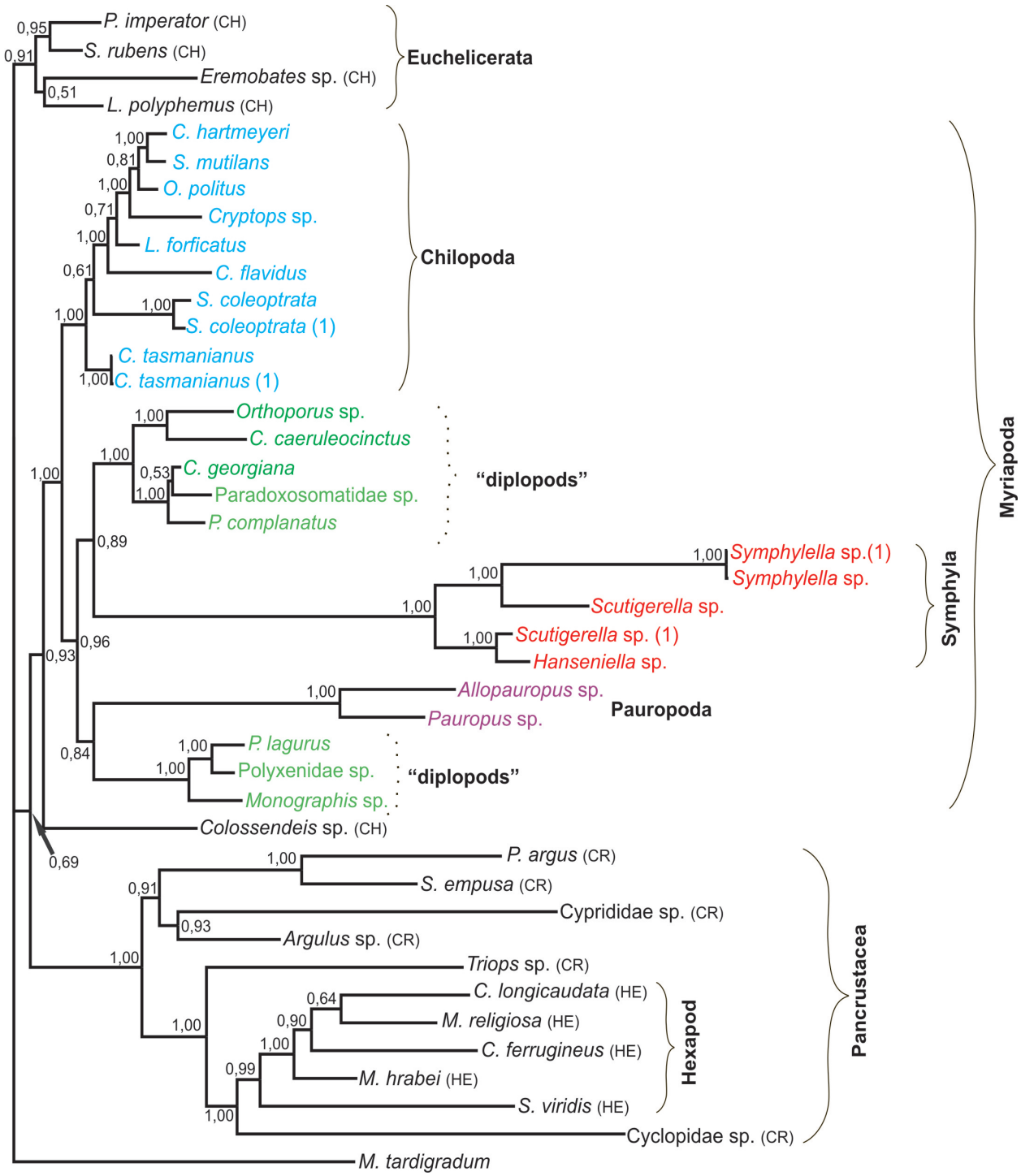

Figure 6. *Bayesian inference (DNA model, GTR+I+Γ). 50% consensus trees for the full character set (2155 included characters) based on the manual approach. Numbers at the nodes indicate the posterior probability values. Tree, branch lengths, and parameters were estimated during the analysis.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; note that "diplopods" indicate paraphyletic diplopods.

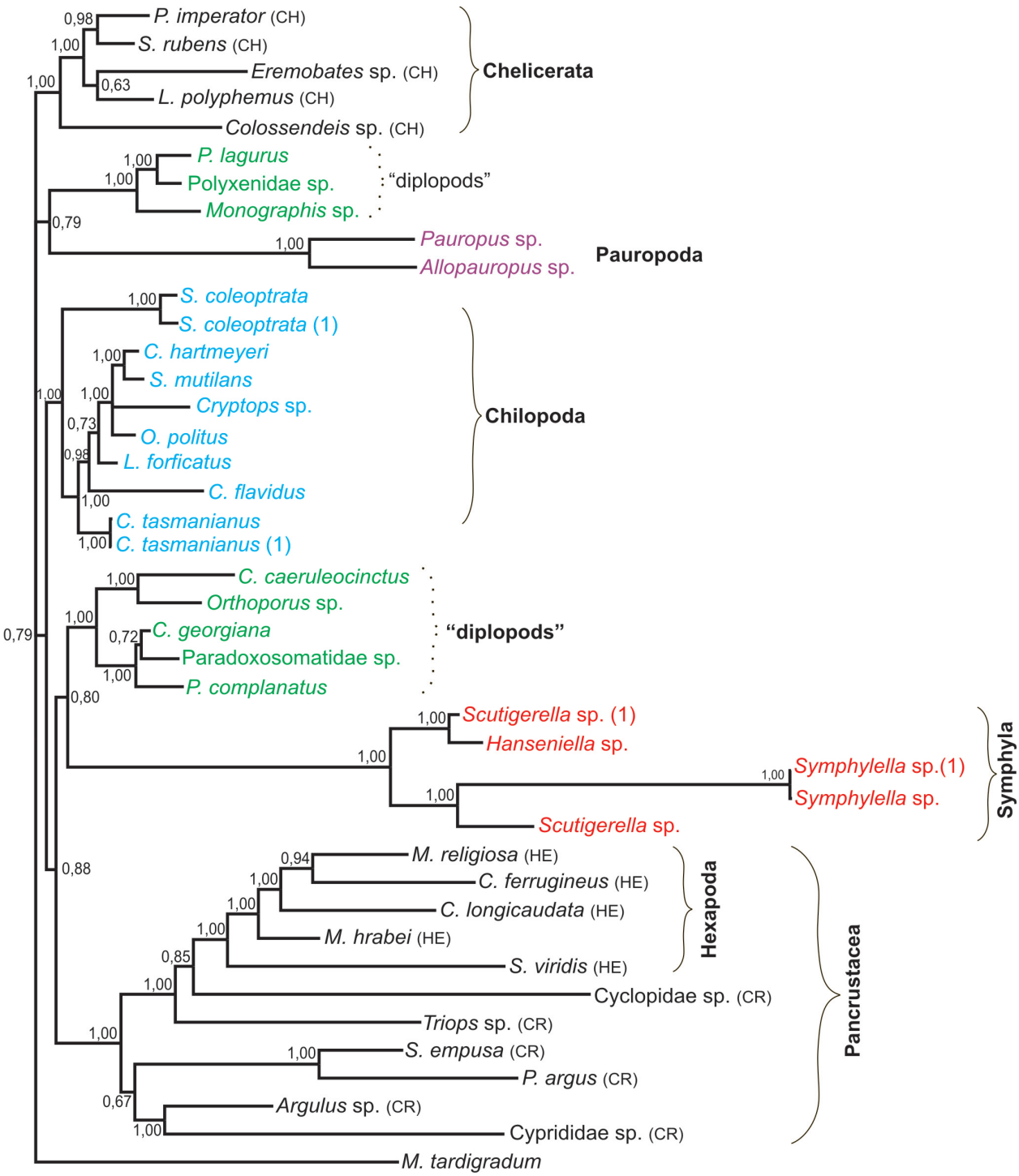

Figure 7. *Bayesian inference (DNA model, GTR+I+Γ). 50% consensus trees for the full character set (2008 included characters) based on the automated approach. Numbers at the nodes indicate the posterior probability values. Tree, branch lengths, and parameters were estimated during the analysis.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; "diplopods" indicate paraphyletic diplopods.

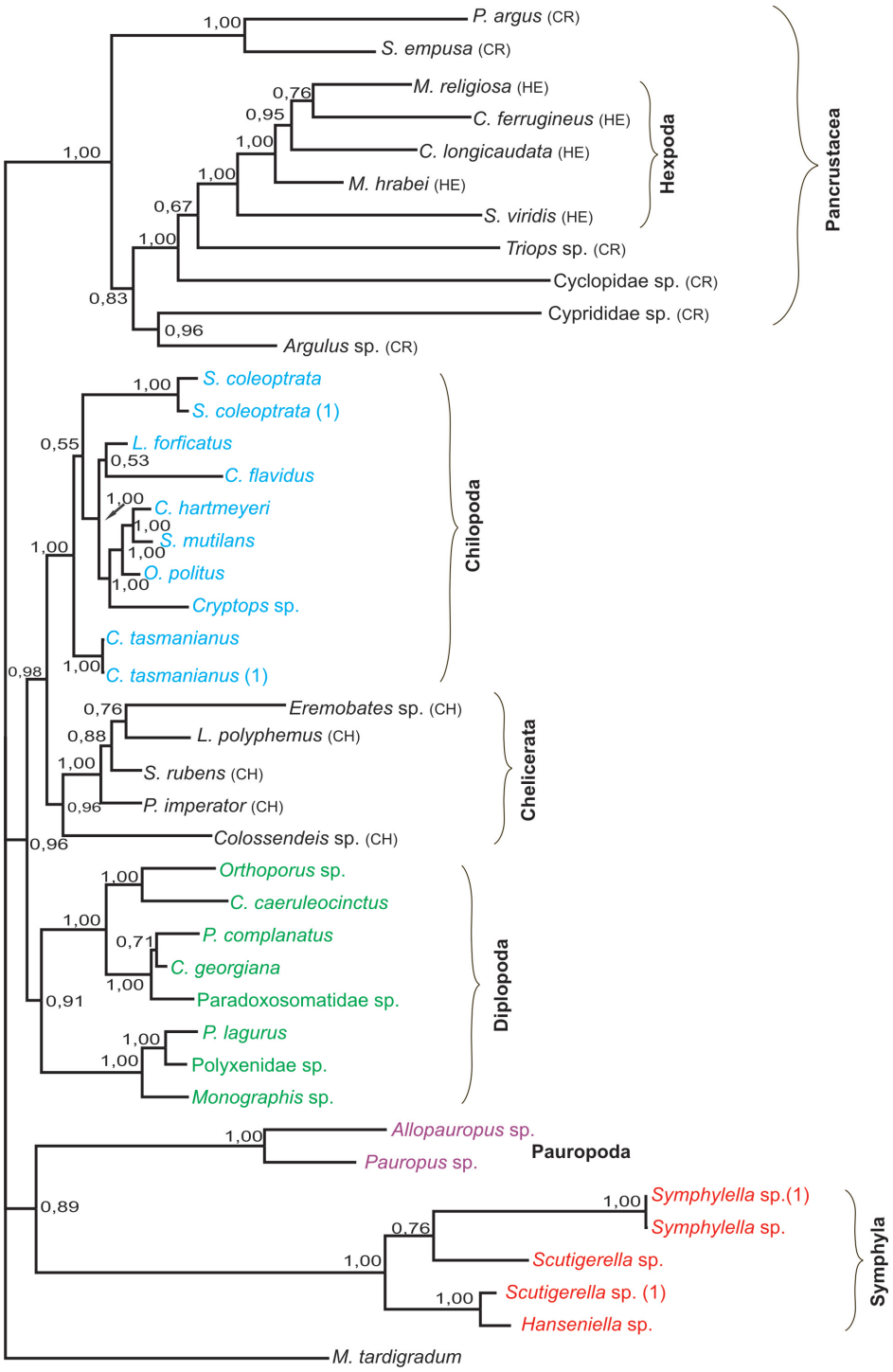

Figure 8. *Reconstruction of phylogenetic relationships employing the Bayesian method with the mixed DNA/RNA model (GTR+Γ/GTR+Γ) for the full character set based on the manual approach. Numbers at the nodes indicate the posterior probability values. Tree, branch lengths, and parameters were estimated during the analysis.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda.

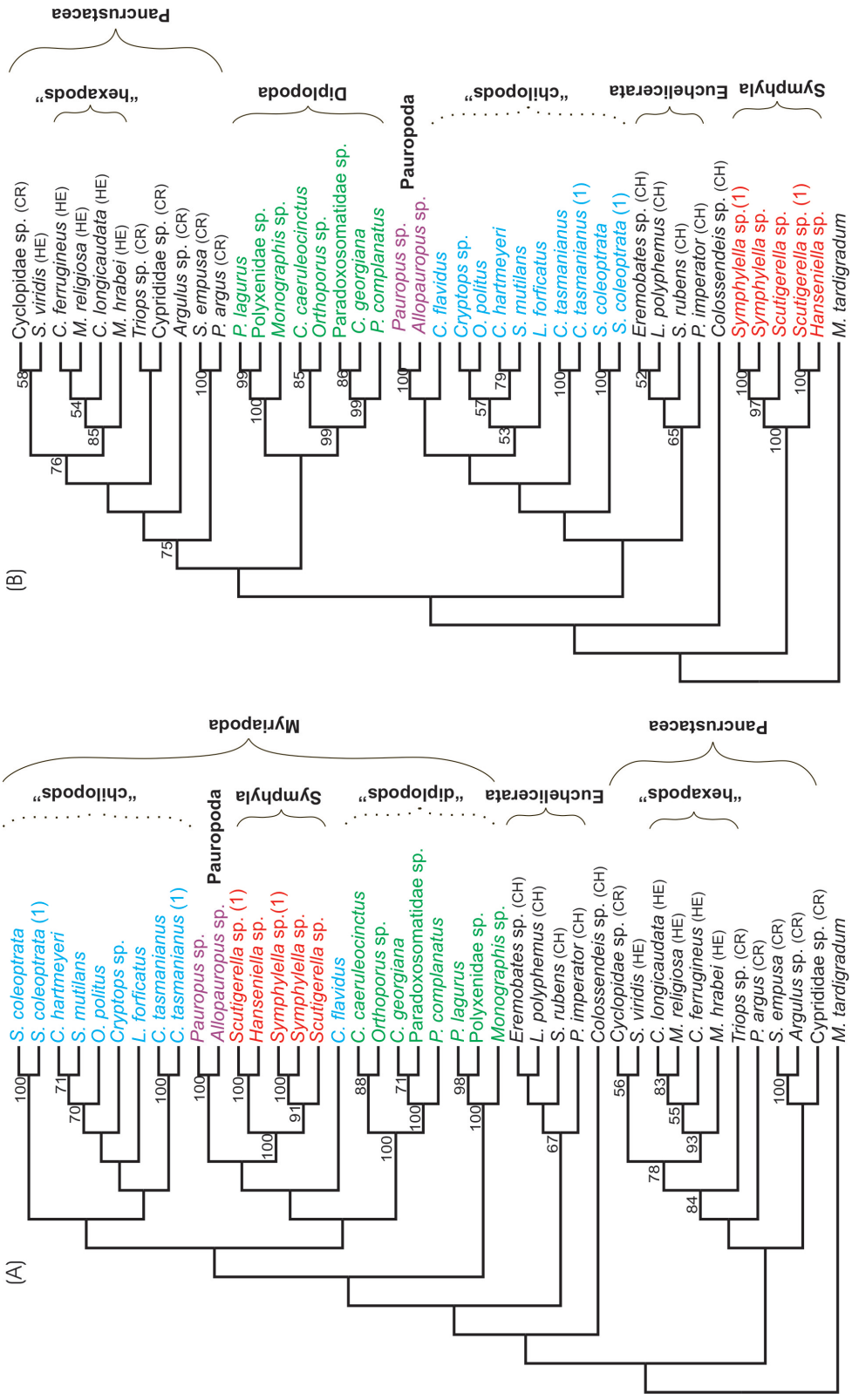

Figure 9. *50% majority rule consensus of most parsimonious trees for the two reduced character sets with bootstrap support above 50% indicated at the corresponding nodes.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; clades under " refers to a non-monophyletic group. (A) manual approach: consensus cladogram of 9 most parsimonious trees with a tree length of 2318 steps; CI:0.40; RI:0.59; (B) Automated approach: consensus cladogram of 5 most parsimonious trees with a tree length of 1716 steps; CI:0.42, RI:0.59.

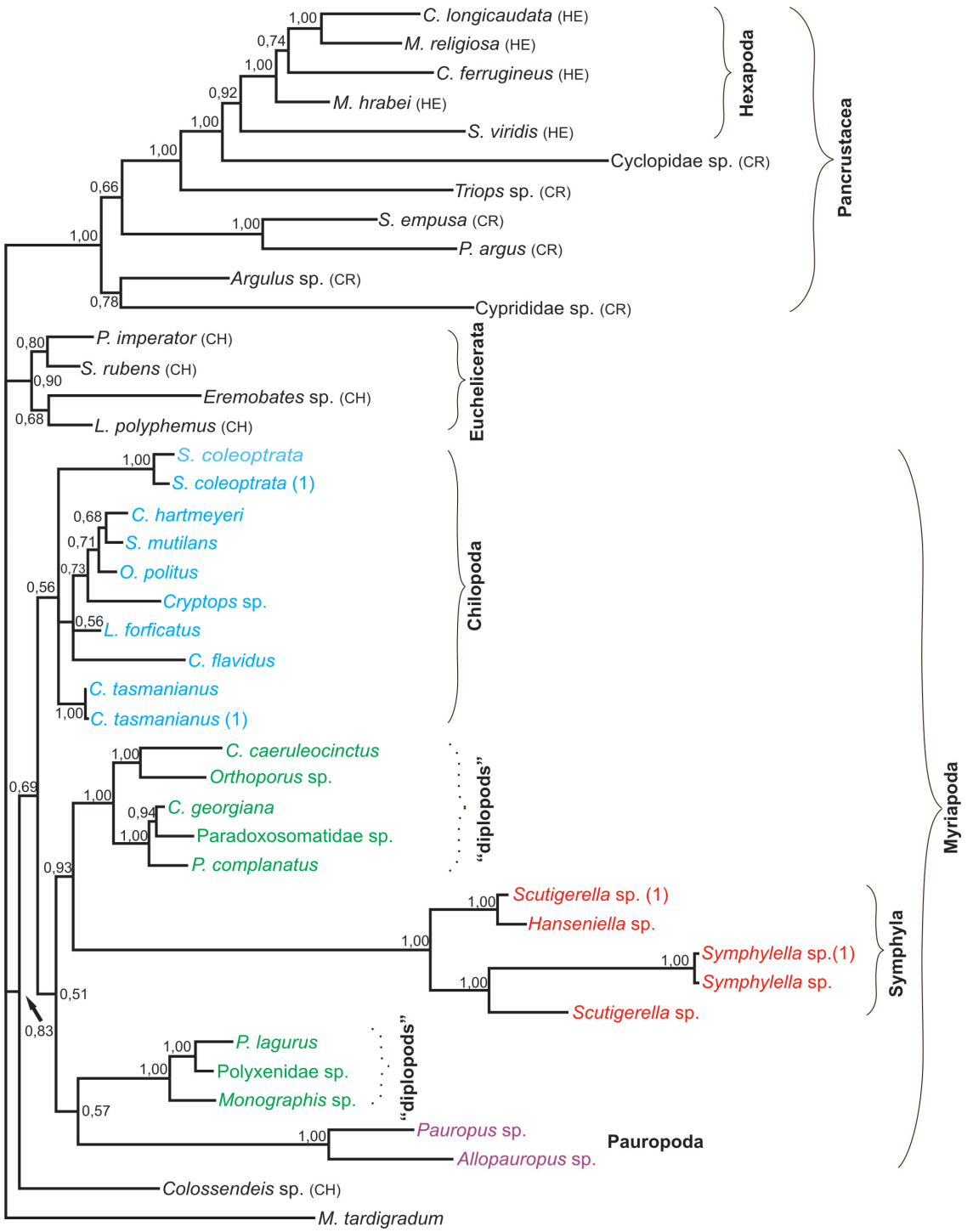

Figure 10. *Bayesian inference (DNA model, GTR+I+Γ). 50% consensus trees for the reduced character set based on the manual approach. Numbers at the nodes indicate the posterior probability values. Tree, branch lengths, and parameters were estimated during the analysis.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; "diplopods" indicate paraphyletic diplopods.

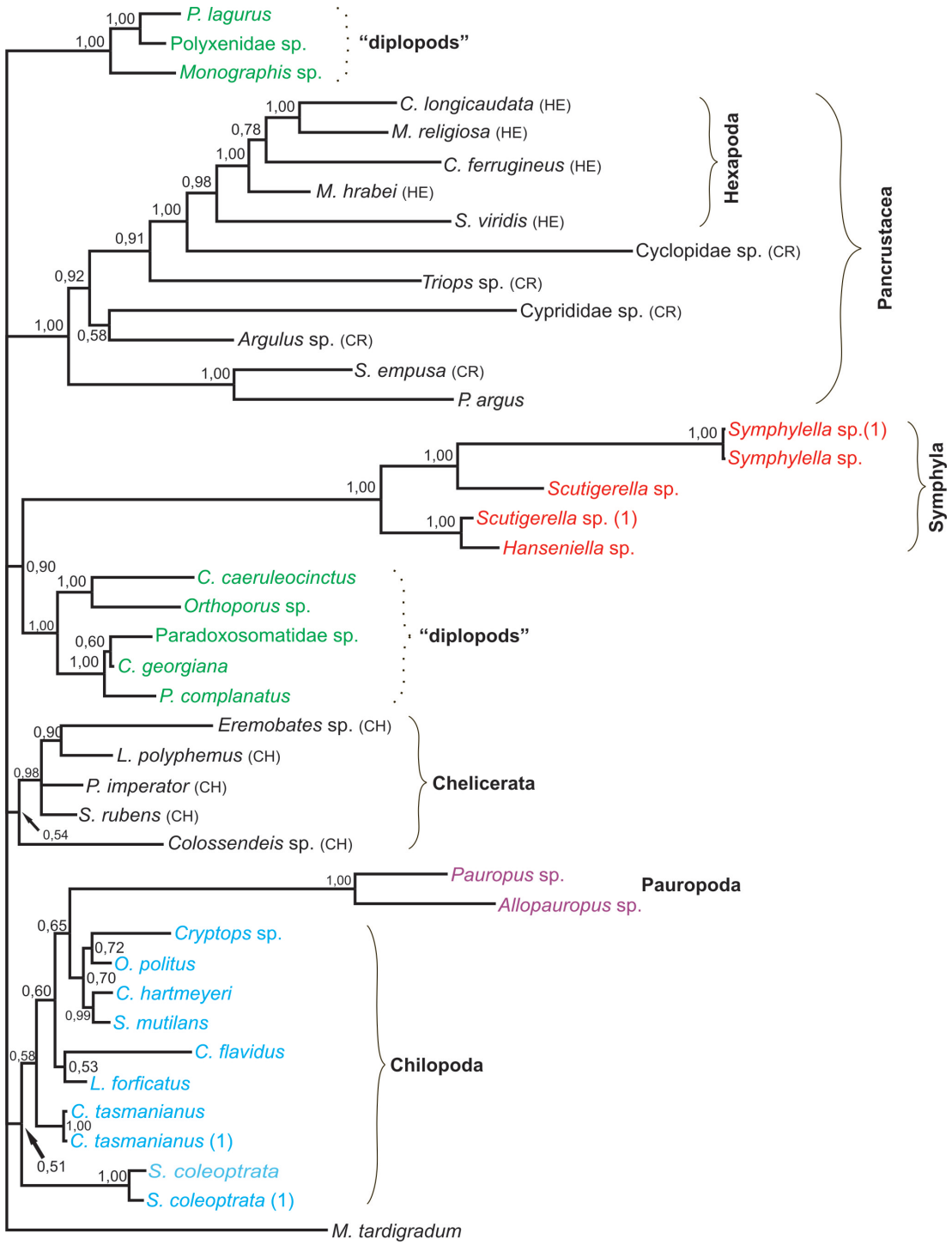

Figure 11. *Bayesian inference (DNA model, GTR+I+Γ). 50% consensus trees for the reduced character set based on the automated approach. Numbers at the nodes indicate the posterior probability values. Tree, branch lengths, and parameters were estimated during the analysis.*
Abbreviations: CH=Chelicerata, CR=Crustacea, HE=Hexapoda; "diplopods" indicate paraphyletic diplopods.

Table 9. *Summary of the performed analyses for the full character set.*

Full character set

	MP (DNA model)		ML (DNA model)		BI (DNA model)		BI (mixed models)
	MA	**AA**	**MA**	**AA**	**MA**	**AA**	**MA**
Mandibulata	–	–	+	–	0.69 (+ *Collosendeis*)	–	–
Paradoxopoda	+	–	–	–	–	–	–
Atelocerata	–	–	–	–	–	–	–
Pancrustacea	100	95	+	+	1.00	1.00	1.00
Crustacea	–	–	–	–	–	–	–
Hexapoda	92 (- *S. viridis*)	64 (- *S. viridis*)	+	+	0.99	1.00	1.00
Chelicerata	–	–	–	+	–	1.00	0.96
Euchelicerata	81	91	+	+	0.91	1.00	1.00
Myriapoda	+	–	+	–	1.00	–	–
Chilopoda	59	+	+	+	1.00	1.00	1.00
Diplopoda	–	–	–	–	–	–	0.91
Pauropoda	100	100	+	+	1.00	1.00	1.00
Symphyla	100	100	+	+	1.00	1.00	1.00
Dignatha	–	–	–	–	–	–	–
Progoneata	–	–	+	–	0.96	–	–
Helminthomorpha	100	99	+	+	1.00	1.00	1.00
Penicillata	100	100	+	+	1.00	1.00	1.00
Pleurostigmomorpha	–	52	–	+	–	1.00	–
Scolopendromorpha	54	63	+	+	1.00	1.00	1.00
Symphyla & Pauropoda	+	62	–	–	–	–	0.89

Table 10. *Summary of the performed analyses for the reduced character set.*

Reduced character set

	MP (DNA model)		BI (DNA)	
	MA	**AA**	**MA**	**AA**
Mandibulata	–	–	–	–
Paradoxopoda	+	–	–	–
Atelocerata	–	–	–	–
Pancrustacea	99	75	1.00	1.00
Crustacea	–	–	–	–
Hexapoda	93 (- *S. viridis*)	85 (- *S. viridis*)	0.92	0.98
Chelicerata	–	–	–	0.54
Euchelicerata	67	65	0.90	0.98
Myriapoda	+	–	0.69	–
Chilopoda	+ (- *C. flavidus*)	+ (+ Pauropoda)	0.56	0.51
Diplopoda	–	+	–	–
Pauropoda	100	100	1.00	1.00
Symphyla	100	100	1.00	1.00
Dignatha	–	–	–	–
Progoneata	–	–	0.51	–
Helminthomorpha	100	99	1.00	1.00
Penicillata	100	100	1.00	1.00
Pleurostigmomorpha	+ (+ *C. flavidus*)	+ (+ Pauropoda)	–	0.58
Scolopendromorpha	+	57	0.73	0.70
Symphyla& Pauropoda	+	–	–	–

6 Discussion

Nuclear RNA genes were shown to be very informative for reconstructions of deep nodes within Arthropoda (Mallatt et al. 2004, Mallatt & Giribet 2006, von Reumont et al. in review). This generalization of the suitability has to be adapted with the inclusion of myriapod sequences. The present results indicate that in this group the phylogenetic signal is superimposed by noise. This is surprising, given that Gai et al. (2006), based on 18S and 28S sequences, mentioned strong support for many nodes, as well as high congruence regarding clades with strong morphological support. The unexpected difficulties of the present study to receive unambiguous hypotheses of relationships among Myriapoda demand to search for causes of misleading effects in the course of the analyses. The possibility to use test-versions of `RNAsalsa` and `ALISCORE`, software constructed for automated structure-based alignments of RNA genes and detection of randomly similar sections across the alignment, promised to avoid the repeatedly mentioned critic of errors caused by the manual alignment strategy. This approach, however, did not decrease ambiguities. A comprehensive test for all possible causes for errors is beyond the scope of the presented study; potential explanations will be given wherever indicated by the presented results.

6.1 Monophyly of Myriapoda- Dependence on the alignment and character exclusion strategy

Irrespective of the enlargement of the taxon sampling (including five Symphyla sequences to break long branches, increasing the amount of Chilopoda and Diplopoda sequences) in comparison to previous studies, no satisfying answer can be given with respect to the monophyly of Myriapoda. The presented trees indicate that results in this respect are highly dependent on the underlying alignment and character exclusion approach. Even in the study, where a part of the newly sequenced taxa were included into a bigger taxon sampling, covering all major groups of arthropods, myriapod relationships remained among the worst resolved groups (von Reumont et al. in review).
Of the results presented here, all five trees, which weakly support the monophyly of Myriapoda, are based on data sets that derived from the manual approach. In contrast, trees from data sets generated with the automated strategy never resulted in monophyletic myriapods (see Table 9 and 10). These findings are in contrast to the study of Gai et al. (2006), who, using similar alignment and character choice approaches (manual approach) and tree reconstruction methods (MP, ML, and Bayesian inference), received high statistical support for the monophyly of Myriapoda. This high support is rather surprising, since it was never received in other studies on RNA genes before (Mallatt & Giribet 2006, von Reumont et al. in review). Unfortunately, Gai et al. (2006) only provides the cleaned alignment and does not allocate the structural alignment with notation of ambiguously aligned and subsequently excluded sites. Hence it is impossible to compare their alignment with the alignments of the present study, with respect to underlying hypotheses of positional homology.

This disagreement illustrates one of the main problems in phylogenetics, the alignment. Therefore, the choice of the alignment method and character evaluation has a great impact on the resulting hypotheses (Kjer 1995, Hickson et al. 1996, Kjer et al. 2007). This is particularly true, when the phylogenetic signal is at the edge of resolution, which is obviously the case within and among the myriapod subgroups. It is difficult to objectively evaluate alignment approaches and it was beyond the scope of the present study to compare both alignment strategies and exclusion

modes in detail, but some remarks should be mentioned in this context. In Figure 2 the results of included/excluded sites in the two approaches are visualized. The automated approach retained more sites as not randomly similar (AA: 2624, MA: 2569), however, subsequent data sets generated with the manual strategy include more characters (MA: 2155, AA: 2008; note missing, gaps and ambiguous characters are excluded from final data sets). It can only be speculated if the monophyly of Myriapoda is based on either these additional included characters of the data set MA or if these results emerged from different assumptions about positional homology hypotheses across the alignment.

Another remarkable point between these two approaches concerns the basal split inside the Chilopoda. The inclusion of fragments of the Divergent Domains (full character set, Figure 4-9) results in the monophyly of Chilopoda. In contrast, topologies from data sets with only Core Regions (reduced character set; Figure 10-12) reject the Chilopoda as natural clade. This is true for both approaches, although within the Chilopoda the two approaches differ considerably. To date the preferred hypothesis regarding the basal split inside the Chilopoda is the Notostigmomorpha/Pleurostigmomorpha concept (for a detailed discussion of competing hypotheses see below in section *'Relationships within Chilopoda and Diplopoda'*). On the one hand, both approaches are consistent in their inability to solve this phylogenetic level with only conserved regions inside the gene (reduced character sets, see Figure 9-11). On the other hand, based on the full character sets, this hypothesis is supported by the automated strategy, but never achieved with data sets from the manual approach.

Again, it can only speculated whether these results are dependent on different character choices inside the Divergent Domains, or on different homology hypotheses between the alignments. Like shown in Figure 2, the manual approach includes only characters of the Divergent Domains, which are on the border to the Core Regions, in contrast to the alignment with `RNAsalsa` and evaluation of the alignment with `ALISCORE`, in which characters are included within the Divergent Domains in more variable regions). A thorough comparison of both approaches (manual and automated approach) is highly demanded since both are consistent in the finding of several clades like Chilopoda, Symphyla, Pauropoda, Helminthomorpha, Penicillata, Pancrustacea, and Euchelicerata, but are inconsistent about the monophyly of Myriapoda and relationships among the Chilopoda.

According to von Reumont et al. (in review), who used both new software, `RNAsalsa` and `ALISCORE` appear as promising alignment and alignment masking tools, and are therefore a step in the right direction to handle a broad taxon sampling even with highly divergent sequences. Interestingly, application of a more biologically realistic model of tree reconstruction (MA, Bayesian mixed DNA/RNA model, Figure 8) resulted in polyphyletic myriapods, in which Symphyla is sister group to Pauropoda, and the remaining myriapod groups (with monophyletic Diplopoda) cluster together with Chelicerata. While the doublet model corrects for violations of assumptions of independence across sites, it also creates a collapse in deep nodes of the tree. Although the incorporation of secondary structure information allows for improved and more biologically realistic estimates of phylogeny, relationships among the main euarthropod clades, in particular the position of the four subgroups of myriapods remain poorly resolved (von Reumont et al. in review).

When the Pancrustacea is considered as natural group, to date no compelling alternative to the monophyly of Myriapoda is suggested. Despite absent characters, (rejected as synapomorphies for a monophyletic group from Dohle 2001) most cited morphological synapomorphies of Myriapoda are correlated with head appendages in particular the structure of the mandible as well as details of

the tentorial endoskeleton (Edgecombe 2004). Although, regarding the existing literature and with the inclusion of molecular data, the picture becomes more complicated. In studies which are based on larger character matrices, the monophyletic status of Myriapoda appears to be dependent on the underlying data (morphological or molecular), alignment strategy, and tree reconstruction method: It is strongly supported by protein-coding genes (Regier et al. 2005b) and purely morphological studies (Ax 1999, Edgecombe et al. 2000, Koch 2003, Bitsch & Bitsch 2004), weakly supported in total evidence analyses (Giribet et al. 2001, 2005, Edgecombe 2004) and ambiguously resolved by ribosomal genes with both the strategy of direct character optimization (instable position: Giribet & Ribera 2000) and with the inclusion of secondary structure information to infer positional homology;(monophyly of Myriapoda: Mallatt et al. 2004, Gai et al. 2006); (instable position: Mallatt & Giribet 2006, von Reumont et al. in review).

In particular these ambiguous results, derived from ribosomal genes, could be partly explained by a low myriapod taxon sampling and misleading effect caused by heterogeneous base composition across taxa (Mallatt & Giribet 2006, von Reumont et al. in review). Since in this study these misleading effects were eliminated from the analyses by inclusion of more myriapod taxa as well as homogenous base frequency across taxa (see Table 6), the myriapod subgroups remain an enigma among the euarthropods.

The present study indicates that the almost complete 28S rRNA gene alone does not provide enough resolution to answer unequivocally the question about the monophyletic status of Myriapoda. However, despite the increased taxon sampling among myriapods, some key species, in particular for Diplopoda and Pauropoda are still missing, and it cannot be ruled out if a denser taxon sampling would increase the phylogenetic signal.

6.2 Relationships among myriapod subgroups - Differences in tree reconstruction methods

Throughout all character matrices and tree reconstruction methods, Chilopoda, Symphyla, Pauropoda, Helminthomorpha, and Penicillata build high statistically supported monophyletic clades. Nevertheless, the relationships among those clades remain contested.

According to the data set based on the manual approach, in which the Myriapoda are monophyletic, there are incongruences between trees derived from model based methods and those from MP analyses. Chilopoda as sister group to Progoneata finds support in the present study by ML and Bayesian inference. This result is upheld by a variety of molecular and morphological data (Ax 1999, Edgecombe et al. 2000, Giribet et al. 2001, Edgecombe 2004, Giribet et al. 2005). In contrast, in both MP topologies the Penicillata appears basal inside the Myriapoda. This finding is surprising since no study so far placed the Penicillata as sister group to the remaining myriapods. The relationship inside the Progoneata shows little agreement with morphological data and total evidence analyses. Most of these studies suggest a sister group relationship between Symphyla and Dignatha (Kraus & Kraus 1994, Dohle 1998, Ax 1999, Edgecombe 2004). Received topologies show a tendency of Symphyla to cluster with Helminthomorpha, and this group is sister group to a clade composed of Pauropoda and Penicillata. These pairings were never proposed before.

However, the relationship among the Progoneata or even among the myriapods are far from

being resolved in phylogenetic analyses based on molecular data.

For example in the study of Giribet et al. (2005) the morphological analysis alone reveals Progoneata in its traditional form: (Dignatha & Symphyla). However, with inclusion of 9 molecular genes the analysis indicates a different relationship inside the Progoneata (Diplopoda (Symphyla & Pauropoda)). In the study of Gai et al. (2006), based on the 18S and 28S rRNA genes, the Chilopoda are considered as sister to Pauropoda and Symphyla, and Diplopoda at the base of the Myriapoda. In the ribosomal gene study of Giribet & Ribera (2000) the inability to resolve relationships of myriapods was mostly traced back to the instable position of Symphyla and Pauropoda. Even the nuclear protein coding-genes analyses of Regier et al. (2005b) yield in topologies of unstable relationships and low node support among the myriapods. In most studies the positions of Symphyla and Pauropoda have not yet been determined unequivocally (Giribet & Ribera 2000, Regier et al. 2005b, Mallatt & Giribet 2006, von Reumont et al. in review). This is also obvious in the present study, where the relationships proposed for symphylans and pauropods are extremely fragile. In general, these two taxa appear in different positions, which are presumably dependent on the underlying tree reconstruction method.

Contrariwise, all but one MP analyses unite the Pauropoda with the Symphyla. This MP derived combination finds confirmation in recent molecular studies (Giribet et al. 2001, Gai et al. 2006, von Reumont et al. in review). This finding seems unusual, because morphological characters do not support this clade. Screening rRNA sequences of Symphyla and Pauropoda, they appear to be highly derived and lead to very long branches in phylogenetic analyses. Major causes for these long branches and assumed are e.g. variation of evolutionary rate among sites and lineages. However, at the moment these variations of evolutionary rates can only be incorporated in the algorithm of the tree reconstruction when base compositional bias for certain taxa is present within the given taxa set. The very latest study of (von Reumont et al. in review) conducted the most comprehensive analyses to infer arthropod phylogeny based on two nuclear ribosomal RNA genes (18S and 28S). Based on the underlying taxa set, that indicated heterogeneity of base composition across taxa, the authors incorporated non-stationary substitution processes combined with a mixed DNA/RNA model in a Bayesian framework. In all constructed trees, the Symphyla and Pauropoda appeared near the root.

This study demonstrated that even with the incorporation of more realistic models of sequence evolution (dependence of characters inside stem regions), and the incorporation of models of non-stationary substitution processes (heterogeneity in base composition), the phylogenetic position of Symphyla and Pauropoda still remained unresolved. The authors concluded that the use of more complex models minimizes misplacement of some taxa, but unusual positions of some long-branch taxa in particular Symphyla and Pauropoda remain and are resistant to these corrections.

As mentioned above, the placement of Symphyla and Pauropoda being dependent on the usage of MP or model based methods, is most remarkable. The sister group relationship between Pauropoda and Symphyla as found in most MP analyses, may be an artificial group due to Long Branch Attraction (LBA). Causes for LBA are hard to determine. For all nucleotide sites of all species among all included data sets, differences in base composition were minimal and should not have interfered with the recovery of the phylogenetic signal. Therefore, shared nucleotide bias can be excluded as a possible explanation. Apparently, model based methods disrupt these assumed "artificial" clade of Symphyla and Pauropoda (Huelsenbeck & Crandall 1997), but clades revealed

in these methods (Symphyla & Helminthomorpha and Pauropoda & Penicillata) were unlikely. This grouping could be traced back to the low number of included diplopod sequences, in which several key taxa are still missing. The present study suggests that Symphyla and Pauropoda sequences were subject to different evolutionary rates in time, however, this did not result in a bias of base composition. Such a bias is usually used to identify high divergent taxa within a given data set. To date, unfortunately there is no approach available to test for this special case when no bias is present. Furthermore, no recent phylogenetic analysis is able to correct for this misleading effect.

6.3 Relationships within Chilopoda and Diplopoda

6.3.1 Chilopoda

Today, regarding the monophyletic status of Chilopoda, there are no real competing hypotheses that would favour a non-monophyly. This also sustain in the present study based on the 28S rRNA. However, minor differences can be observed between the two data matrices (full character set and reduced character set). Topologies based on the full character sets support the Chilopoda as monophyletic group regardless of the alignment and character choice strategy and analytical method. Exclusion of variable characters (Divergent Domains) yields in a low resolution, since in some topologies the Pauropoda are nested inside the Chilopoda (Figure 9B and Figure 12) or *C. flavidus* (Geophilomorpha) clusters together with the Symphyla and Pauropoda (Figure 9A). Nevertheless, the results based on the full character sets are in agreement with all recently published molecular and morphological studies, which defend the monophyly of Chilopoda (Ax 1999, Giribet et al. 2001, Edgecombe & Giribet 2002, 2004, Bitsch & Bitsch 2004, Giribet et al. 2005, Regier et al. 2005b, Mallatt et al. 2004, Mallatt & Giribet 2006, Gai et al. 2006, von Reumont et al. in review). Inside the Chilopoda, some ambiguities arise between studies based on nuclear protein-coding genes and ribosomal genes. The Notostigmomorpha/Pleurostigmomorpha hypothesis is regarded as well supported by studies based on ribosomal genes and morphology (Giribet & Wheeler 1999, Giribet & Ribera 2000, Edgecombe & Giribet 2002). In contrast, protein-coding genes do not support this Pleurostigmomorpha/Notostigmomorpha but rather place the Craterostigmomorpha basal among the Chilopoda groups (Regier et al. 2005a) In the present study, as mentioned above (see section *'Monophyly of Myriapoda - Differences in the alignment character exclusion strategy'*) a considerable difference between the two approaches affects the basal split inside the Chilopoda. All topologies received with `RNAsalsa` are congruent in the discovery of the split between Notostigmomorpha and Pleurostigmomorpha. On the contrary, this basal split is ambiguously recovered with data sets based on the Manual alignment, in which either *C. flavidus* or *C. tasmanianus* branch off first. The basal position of *C. flavidus* inside the Chilopoda was once proposed before by Ax (1999). But neither the proposed clades of Gonopodophora (Lithobiomorpha&Scutigeromorpha) nor the Triakontapoda (Gonopodophora&Heteroterga) nor the Heteroterga clade (Triakontapoda&Scolopendromorpha) was resolved under all data sets. The view of Ax (1999) is not held by modern experts and is rejected in favour of the before mentioned Notostigmomorpha/Pleurostigmomorpha hypotheses. In the present study this hypothesis is, though strategy dependent, preferred over the competing hypotheses.

Screening all data sets and tree reconstruction methods the relationship among the Pleurostigmomorpha is not ambiguously resolved. The four Scolopendromorpha taxa always cluster together,

but in most analysis L. forficatus appeared as sister group and *C. flavidus*, as representative of the Geophilomorpha, as sister to this clade. Specifically, the analyses never recovered the clade Epimorpha s. str. or Phylactometria, which is proposed by morphological and combined analyses (Giribet & Wheeler 1999, Edgecombe & Giribet 2002, Edgecombe 2004).

However, the inability of the present study to resolve the relationships inside the Pleurostigmomorpha as well as inside the Chilopoda is mostly caused by a low taxon sampling. I assume with a broader taxon sampling this ambiguity can be eliminated.

6.3.2 Diplopoda

Due to problems in amplification and sequencing of *G. hexasticha* it should be mentioned that to date there is no trustworthy sequence of Pentazonia available, so the discussion about the monophyly of Diplopoda should be considered carefully. Therefore, the Chilognatha clade (consisting of Pentazonia and Helminthomorpha) can neither be rejected nor supported in this study. Almost none of the analyses recovered monophyly of Diplopoda, however, Helminthomorpha and Penicillata each appeared as a monophyletic unit with high statistical support (bootstrap support$\geq$99; pp$\geq$0.99). The choice of the taxon sampling was rather based on the availability of specimens than on an even distribution of diplopod groups. Therefore, the Penicillata are represented with three sequences, all five Helminthomorpha sequences belong to the Eugnatha. Unfortunately, to date like the before mentioned Pentazonia sequence no Colobognatha sequence is disposable. Phylogenetic analyses of the Diplopoda based on total evidence analysis are scarce. The most recent molecular phylogenetic analysis is based on nuclear protein-coding genes and comprises a broad taxonomic diplopod sampling (Regier et al. 2005b). However, conclusions about deep nodes are limited due to the lack of resolution in these nodes. The authors agree with some well accepted clades like the Helminthomorpha, Pentazonia, and Colobognatha, but fail to completely resolve the relationships among the diplopods. Sierwald & Bond (2007) combine morphological data (Sierwald et al. 2003) with the molecular data of Regier et al. (2005b), and provide the first combined analyses to infer diplopod phylogeny. One of the most striking results is the high nodal support values, which are lacking when the two data sets are analyzed separately. Additionally, the Penicillata are inferred basal inside the Diplopoda, followed by the Pentazonia and the Helminthomorpha. These two analyses, Sierwald & Bond (2007) and Regier et al. (2005b) are to date the only studies comprising a comprehensive diplopod taxon sampling. To resolve the question about diplopod relationships in detail, it is highly demanded to infer diplopod phylogeny using the ribosomal genes encoding for the 28S rRNA, for which to date only eight sequences are available, and subsequently combine all data in a total evidence analyses. This is also true for a broader frame, and should be considered for all myriapod subgroups.

6.4 Mandibulata/Pancrustacea, Mandibulata/Atelocerata, or Paradoxopoda?

One striking result of the present study based on ribosomal genes is the high supported Pancrustacea and Euchelicerata (regardless of character choice and alignment strategy). The Pancrustacea concept finds support in all recent studies based on both molecular and morphological evidences (Turbeville et al. 1991, Friedrich & Tautz 1995, Zrzavý & Štys 1997, Shultz & Regier 2000, Dohle 2001, Giribet et al. 2001, 2005, Regier et al. 2005a, Harzsch et al. 2005, Mallatt & Giribet 2006, Gai et al. 2006, Bourlat et al. 2008, Ungerer & Scholtz 2008).

However, the relationship between Euchelicerata and Pycnogonida (sea spiders) is not fully resolved. In the present study the position of the pycnogonide *Colossendeis* sp. varies among the data sets. The position of this group is debated in recent studies. The Cormogonida concept (Giribet et al. 2001) suggests that the Pycnogonida are placed as sister group to all extant arthropods. In contrast the study of Giribet et al. (2005) places this clade to the Euchelicerata by most parameter sets. The analyses of Mallatt & Giribet (2006) joins the pycnogonids with euchelicerates and myriapods. A recent neuroanatomical study from Maxmen et al. (2005) suggests that the pycnogonid chelifores are not homolog to the euchelicerate cheliceres. The position of pycnogonids whether this group is sister to the Euchelicerata or to be placed at the base of all arthropods is not resolved ambiguously.

In the present study almost all data sets support the idea of the monophyly of Hexapoda. When the hexapods are not monophyletic *S. viridis* are outside this clade and cluster together with Cyclopidae sp. (Crustacea). Most studies, which are based on ribosomal genes, indicate the Hexapoda as monophyletic group (e.g. Mallatt et al. 2004, Mallatt & Giribet 2006, Dell'Ampio et al. 2009), which is in agreement with many morphological studies (Kraus & Kraus 1994, Dohle 2001). Unlike studies based on mitochondrial genes, that rather indicate the hexapods as paraphyletic unit. (e.g. Cook et al. 2005) Interestingly, as for the question of the monophyly of Myriapoda, the present study gives no clear signal about the ancient split events among euarthropods. Neither Mandibulata or Atelocerata nor the competing hypothesis, Paradoxopoda, can be favoured among all performed trees. The present study indicates that data sets derived from the Manual strategy (full character set) possess weak phylogenetic signal at this phylogenetic level, and is dependent on the tree reconstruction method. MP analyses favour the Paradoxopoda, whereas the topologies based on the two model-based methods received indication for the Mandibulata. Note, that with Bayesian inference analyses (mixed models), all data sets based on the automated approach, and even with a restricted character set (MA) the respective trees show little resolution in this particular node. Recently, the Paradoxopoda as natural unit is now contested, since it derives only from molecular studies, and no compelling synapomorphies are found so far. Wägele & Mayer (2007) suggest that topologies, which obtain Paradoxopoda, can be partly explained as signal erosion. Rota-Stabelli & Telford (2008) indicate in their mitochondrial gene analyses, that this argumentative hypothesis is dependent on outgroup choice, and suggest a more careful and objective choice of outgroup(s), when dealing with derived sequences. The two studies of Mallatt (Mallatt et al. 2004, Mallatt & Giribet 2006) even indicate that it is highly dependent on the taxon sampling: Mallatt et al. (2004) inferred high statistical support for the Paradoxopoda hypothesis, but Mallatt and Giribet (2006) arrived at the conclusion, that with a more expanded taxon sampling the Mandibulata can not be rejected entirely. Therefore, more investigations in both molecular and morphological considerations must be performed to resolve this particular node inside the Euarthropoda. Presumably the 28S rRNA gene is not adequate to resolve this question and more slow evolving genes should be considered.

6.5 Myriapod 28S rDNA sequences

Heterogeneity of sequence length is one characteristic of the present study. Among the data set the shortest sequence derives from the crustacean Cyclopidae sp. with 3331nt. In comparison the longest sequence originates from *C. flavidus* (Chilopoda, Geophilomorpha) and is more than 1700bp longer. Table 7 shows the length variation among the data set, and indicates that the longest

sequences are mostly found in myriapods. *C. flavidus* possesses several insertions in almost all Divergent Domains. Sequences of Pauropoda and Symphyla are longer than 4000bp, and insertions occur also in conserved regions. Additionally, the amplified sequence of *G. hexasticha* (Diplopoda) was unexpectedly long, about 5571bp. Currently, no other full 28S rRNA sequence of a Glomeridae or another genus of Pentazonia available at NCBI to control the amplified sequence, except the 28S rRNA sequence of Sphaerotheriidae sp.Mallatt & Giribet (2006) assumed that this sequence might be a pseudo gene. Hence, this sequence was discarded a priori from their phylogenetic analysis. The newly amplified glomerid sequence of *G. hexasticha* is therefore the second unusual sequence of a Glomeridae and casts some doubt on the suggestion of Mallatt & Giribet (2006), that rather peculiar changes might have occurred in the evolution of pentazonian sequences. Two factors suggest the *G. hexasticha* sequence represents a functional 28S rRNA gene instead of a pseudo gene or an experimental artefact. (i) The presence of several insertions in areas amplified with different primer pairs, and (ii) the maintenance of base pairing in conserved regions. To test if all Pentazonia shares these unusual long sequences it is necessary to amplify and sequence more species of this family. These exceptionally long sequences are not only a property of the 28S rRNA molecule but also sustain in the rRNA of the small subunit (18S). Giribet et al. (2001) point out that among Metazoa, several myriapod sequences are longer than the expected sequences. All families of Geophilomorpha (Chilopoda), except the basal Mecistocephalidae, display insertions of 300bp inside one distinct variable region. Pauropoda sequences are longer than 2200bp (Gai et al. 2006), species of the family Polyzonidae (Diplopoda, Helminthomorpha) possess sequences longer than 2700bp. Symphyla are extraordinary. On the one hand Giribet et al. (2001) claimed that the three investigated sequences of Symphyla revealed a deletion of about 500bp in the central region of the molecule, leading to 18S rRNA sequences of about 1350bp (results based upon two species of Scutigerella and one species of Hanseniella). On the contrary, Mallatt et al. (2004) and Gai et al. (2006) who amplified further species, received sequences of about 1970bp. It is not clear, which of these obtained results display the functional "symphylan" 18S rRNA molecule. According to Giribet & Ribera (2000), such abnormalities in primary sequence in many myriapod species may affect their phylogenetic position. However, it can be argued that with the inclusion of only conserved regions in the final data set (realized in the manual approach), in which homology hypothesis are well corroborated, this length heterogeneity in primary sequences can be corrected, and has no effect in subsequent tree reconstruction studies. It is highly necessary to amplify and sequence more 28S rRNA genes of myriapods and compare each other for shared structural elements. Consequently, it is possible that the secondary structure of myriapods diverged from the 'normal' euarthropod structure and this would lead to 'wrong' phylogenetic results. Additionally, to investigate the secondary structure of these highly interesting Euarthropoda groups in detail would give some conclusions of presumably curious evolutionary changes inside ribosomal genes of myriapods.

6.6 Conclusions

This study was the first attempt to use new alignment and character choice approaches on an enhanced taxon sampling of myriapods covering all subgroups. The data of the study suggest that while the 28S rRNA gene contain enough signal to reconstruct relationships among most euarthropods like Pancrustacea, Hexapoda, Euchelicerata, Chilopoda, Symphyla, Pauropoda, Helmintho-

morpha, and Penicillata, it failed to place the myriapods among the Euarthropoda and even their monophyletic status remains unresolved. This might indicate that among myriapods the rRNA genes underwent evolutionary changes, which deviate from the "standard euarthropod model". This putative deviation asks for a special treatment, in particular with regard to their secondary structures prediction and the choice of tree reconstruction methods. Regarding the 28S rRNA gene and the phylogenetic position of the myriapods inside the Euarthropoda, only a low signal could be detected at this phylogenetic level, and neither the Mandibulata nor the Paradoxopoda hypothesis can be favoured. Furthermore, for a given data set, the choice of the alignment strategy is mostly based on preferences of investigators for a certain alignment method. The present study provides clear evidence that the choice of alignment strategy is very important for all subsequent analyses and often leads to conflicts in some constructed clades. It shows, that instable outcomes of alignment and character choice approaches stress the necessity of the development of tools (benchmark tests) for the comparison of structural alignment approaches, including long ribosomal genes. Despite these problems in the choice of the appropriate alignment method, regarding the Myriapoda, the low resolution by ribosomal genes sustains in almost all molecular studies, and may be due to rapid radiation, with the result that only few informative sites could have accumulated, supporting the Myriapoda as natural clade. (Regier et al. 2005b) This present study clearly demonstrates a high need for detailed investigations of the evolution of secondary structure of myriapod sequences. Further developments in tree reconstruction methods, and the search for new adequate markers, which are not affected by peculiar evolutionary rates among the taxon sampling, are also of paramount importance.

References

Ax, P. (1999). *Das System der Metazoa II. Ein Lehrbuch der phylogenetischen Systematik*. Gustav Fischer Verlag, Stuttgart.

Bäcker, H., Fanenbruck, M., & Wägele, J. W. (2008). A forgotten homology supporting the monophyly of Tracheata: The subcoxa of insects and myriapods re-visited. *Zoologischer Anzeiger*, *247*(3), 185–207.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, *289*(5481), 905–920.

Bitsch, C., & Bitsch, J. (2004). Phylogenetic relationships of basal hexapods among the mandibulate arthropods: a cladistic analysis based on comparative morphological characters. *Zoologica Scripta*, *33*, 511–550.

Bourlat, S. J., Nielsen, C., Economou, A. D., & Telford, M. J. (2008). Testing the new animal phylogeny: A phylum level molecular analysis of the animal kingdom. *Molecular Phylogenetics and Evolution*, *49*, 23–31.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.

Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., & Akam, M. (2001). Hox genes and the phylogeny of the arthropods. *Current Biology*, *11*, 759–763.

Cook, C. E., Yue, Q., & Akam, M. (2005). Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proceedings. Biological sciences*, *272*(1569), 1295–1304.

Dell'Ampio, E., Szucsich, N. U., Carapelli, A., Frati, F., Steiner, G., Steinacher, A., & Pass, G. (2009). Testing for misleading effects in phylogenetic reconstruction of ancient lineages of hexapods: influence of character dependence and character choice in analyses of 28S rRNA sequences. *Zoologica Scripta*, *38*(2), 155–170. In pending revisions.
URL `http://www3.interscience.wiley.com/cgi-bin/fulltext/121578670/HTMLSTART`

Dixon, M. T., & Hillis, D. M. (1993). Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution*, *10*(1), 256–267.
URL `http://mbe.oxfordjournals.org/cgi/content/abstract/10/1/256`

Dohle, W. (1985). Phylogenetic pathways in the chilopoda. *Bijdr Dierkunde*, *55*, 55–66.

Dohle, W. (1998). *Myriapod-insect relationships as opposed to an insect-crustacean sister group relationship*, vol. Arthropod relationships. Fortey, R. A. and Thomas, R. H.

Dohle, W. (2001). Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name "Tetraconata" for the monophyletic unit Crustacea + Hexapoda. *Annales de la Societe Entomologique de France (New Series)*, *37*(3), 85–103.

Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*(1), 113.

Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edgecombe, G. D. (2004). Morphological data, extant myriapoda, and the myriapod stem-group. *Contributions to Zoology*, *73*, 207–252.

Edgecombe, G. D., & Giribet, G. (2002). Myriapod phylogeny and the relationships of Chilopoda. In J. L. Bousquets, & J. Morrone (Eds.) *Biodiversidad, Taxonomía y Biogeografia de Artrópodos de México: Hacia una Síntesis de su Conocimiento*, vol. III, (pp. 143–168). Aula-Verlag.

Edgecombe, G. D., & Giribet, G. (2004). Adding mitochondrial sequence data (16s rrna and cytochrome c oxidase subunit i) to the phylogeny of centipedes (myriapoda: Chilopoda): an analysis of morphology and four molecular loci. *Journal of Zoological Systematics and Evolutionary Reseach*, *42*, 89–134.

Edgecombe, G. D., Wilson, G. D. F., Colgan, D. J., Gray, M. R., & Cassis, G. (2000). Arthropod Cladistics: Combined analysis of histone H3 and U2 snRNA sequences and morphology. *Cladistics*, *16*(2), 155–203.

Enghoff, H. (1984). Phylogeny of millipedes: a cladistic analysis. *Zool Z Syst Evolforsch*, *22*, 8.

Farris, J. S. (1989). The retention index and the rescaled consistency index. *Cladistics*, *5*, 417–419.

Farris, J. S., Kluge, A. G., , & Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. *Systematic Zoology*, *19*(2), 172–189.

Felsenstein, J. (1981). Evolutionary trees from dna sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376.

Felsenstein, J. (1985). Confidence limits on phylogenies; an approach using the bootstrap. *Evolution*, *39*, 783–791.

Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer, Sunderland.

Fontaine, A., de Monte, A., & Touzet, H. (2008). Magnolia: Multiple alignment of protein coding and structural rna sequences. *Nucleic Acids Research*, *36*.

Friedrich, M., & Tautz, D. (1995). Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature*, *376*(6536), 165–167.

Gai, Y.-H., Song, D.-X., Sun, H.-Y., & Zhou, K.-Y. (2006). Myriapod monophyly and relationships among myriapod classes based on nearly complete 28S and 18S rDNA sequences. *Zoological Science*, *23*(12), 1101–1108.

Gillespie, J. J., Johnston, J. S., Cannone, J. J., & Gutell, R. R. (2006). Characteristics of the nuclear (18s, 5.8s, 28s and 5s) and mitochondrial (12s and 16s) rrna genes of apis mellifera (insecta: Hymenoptera): structure, organization, and retrotransposable elements. *Insect Molecular Biology*, *15*, 657–686.

Giribet, G. (2001). Exploring the behavior of poy, a program for direct optimization of molecular data. *Cladistics*, *17*(1 Pt 2), S60–S70.

Giribet, G., Edgecombe, G. D., & Wheeler, W. C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. *Nature*, *413*(6852), 157–161.

Giribet, G., & Ribera, C. (2000). A review of arthropod phylogeny: New data based on ribosomal DNA sequences and direct character optimization. *Cladistics*, *16*(2), 204–231.

Giribet, G., Richter, S., Edgecombe, G. D., & Wheeler, W. C. (2005). *The position of crustaceans within Arthropoda - Evidence from nine molecular loci and morphology*, vol. Crustacean Issues 16, Crustacea and Arthropod Relationships, Festschrift for Frederick R. Schram. Taylor & Francis, Boca Raton.

Giribet, G., & Wheeler, W. C. (1999). Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal DNA sequences and morphology. *Molecular Phylogenetics and Evolution*, *11*(2), 296–307.

Gutell, J. C., Robin R .and Lee, & Cannone, J. J. (2002). The accuracy of ribosomal RNA comparative structure models. *Current Opinion in Structural Biology*, *12*(3), 301–310.

Gutell, R., Weiser, B., Woese, C. R., & Noller, H. (1985). Comparative anatomy of 16s-like ribosomal rna. *Progress in Nucleic Acid Research and Molecular Biology*, *32*, 155–216.

Hall, T. A. (1999). BioEdit: a user-friendly biological alignment sequence EDITOR and analysis program for Windows95/98/NT. *Nucleic Acids Symposium Series*, *41*(2-3), 95–98.

Hancock, J. M., Tautz, D., & Dover, G. A. (1988). Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Molecular Biology and Evolution*, *5*(4), 393–414.

Harzsch, S., Müller, C. H. G., & Wolf, H. (2005). From variable to constant cell numbers: cellular characteristics of the arthropod nervous system argue against a sister-group relationship of Chelicerata and "Myriapoda" but favour the Mandibulata concept. *Development Genes and Evolution*, *215*(2), 53–68.

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*, 160–174.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*, 97–109.

Hickson, R., Simon, C., Cooper, A., Spicer, G., Sullivan, J., & Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution*, *13*(1), 150–169.
URL `http://mbe.oxfordjournals.org/cgi/content/abstract/13/1/150`

Hillis, D., Moritz, C., & Mable, B. (1996). *Nucleic Acids II: The Polymerase Chain Reaction, Chapter 7*, vol. Molecular systematics "Second edition". Sinauer Associates Inc Publishers Sunderland, Massachusets, U. S. A..

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology*, *66*(4), 411–453.

Hofacker, I. L. (2003). Vienna rna secondary structure server. *Nucleic Acids Research*, *31*(13), 3429–3431.
URL `http://nar.oxfordjournals.org/cgi/content/abstract/31/13/3429`

Hofacker, I. L., Fekete, M., & Stadler, P. F. (2002). Secondary structure prediction for aligned rna sequences. *Journal of Molecular Biology*, *319*, 1059–1066.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of rna secondary structures. *Monatshefte für Chemie*, *125*, 167–188.

Huelsenbeck, J. P., & Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, *28*(1), 437–466.
URL `http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.28.1.437`

Huelsenbeck, J. P., Larget, B., Miller, R. E., & Ronquist, F. (2002). Potential applications and pitfalls of bayesian inference of phylogeny. *Systematic Biology*, *51*(5), 673–688.

Hwang, U. W., Friedrich, M., Tautz, D., Park, C. J., & Kim, W. (2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, *413*(6852), 154–157.

Jow, H., Hudelot, C., Rattray, M., & Higgs, P. G. (2002). Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Molecular Biology and Evolution*, *19*(9), 1591–1601.

Jukes, T. H., & Cantor, C. R. (1969). *Evolution of protein molecules*. Mammalian protein metabolism, Academic Press, New York.

Kaas, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution*, *4*(3), 314–330.

Kjer, K. M., Gillespie, J. J., & Ober, K. A. (2007). Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between poy and structural alignment. *Systematic Biology*, *56*, 133–146.

Kluge, A., & Farris, J. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Zoology*, *18*, 1–32.

Koch, M. (2003). Monophyly of the Myriapoda? Reliability of current arguments. *African Invertebrates*, *44*(Proceedings of the 12th International Congress of Myriapodology), 137–153.

Kraus, O. (1998). Phylogenetic relationships between higher taxa of tracheate arthropods. *Arthropod Relationships*, *55*, 295–303.

Kraus, O., & Kraus, M. (1994). Phylogenetic system of the tracheata (mandibulata): on myriapoda-insecta interrelationships, phylogenetic age and primary ecological niches. *Verhandlungen des Naturwissenschaftlichen Vereins in Hamburg*, *34*, 5.

Mallatt, J., & Giribet, G. (2006). Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Molecular Phylogenetics and Evolution*, *40*(3), 772–794.

Mallatt, J. M., Garey, J. R., & Shultz, J. W. (2004). Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Molecular Phylogenetics and Evolution*, *31*(1), 178–191.

Maxmen, A., Browne, M. Q., W. E.and Martindale, & Giribet, G. (2005). Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. *Nature*, *437*.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1092.

Misof, B., & Misof, K. (in press). A monte carlo approach sucessfully identifies randomness of multiple sequence alignments. *Systematic Biology*.

Misof, B., Niehuis, O., Bischoff, I., Rickert, A., Erpenbeck, D., & Staniczek, A. (2007). Towards an 18S phylogeny of hexapods: Accounting for group-specific character covariance in optimized mixed nucleotide/doublet models. *Zoology (Jena)*, *110*(5), 409–429.

Moretti, S., Wilm, A., Higgins, D. G., Xenarios, I., & Notredame, C. (2008). R-coffee: a web server for accurately aligning noncoding rna sequences. *Nucleic Acids Research*, *36*(9), 52.

Morrison, D. A. (2006). Multiple sequence alignment for phylogenetic purposes. *Australian Systematic Botany*, *19*, 479–539.

Mullis, K. B., F., F., Scharf, S., Saiki, R. K., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of dna in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology*, (pp. 263–273).

Noller, H. F., & Woese, C. R. (1981). Secondary structure of 16s ribosomal rna. *Science*, *212*, 403–411.

Nylander, J. A. A. (2004). MrModeltest v2. *Department of Systematic Zoology, EBC, Uppsala University, Sweden*, (p. 2008).

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., & Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*(21), 47–67.

Omilian, A. R., & Taylor, D. J. (2001). Rate acceleration and long-branch attraction in a conserved gene of cryptic daphniid (crustacea) species. *Molecular Biology and Evolution*, *18*, 2201–2212.

Page, R. D. M., & Holmes, E. C. (1998). *Molecular evolution: A phylogenetic approach*. Blackwell Science Ltd. Oxford.

Pisani, D., Poling, L., Lyons-Weiler, M., & Hedges, S. B. (2004). The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biology*, *2*(1), 1.

Pocock, R. I. (1893). On the classification of the Tracheate Arthropoda. *Zoologischer Anzeiger*, *16*, 271–275.

Podsiadlowski, L., Kohlhagen, H., & Koch, M. (2007). The complete mitochondrial genome of *Scutigerella causeyae* (Myriapoda: Symphyla) and the phylogenetic position of Symphyla. *Molecular Phylogenetics and Evolution*, *45*(1), 251–260.

Regier, J. C., Shultz, J. W., Ganley, A. R. D., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J. E., Cummings, J. W., Michael P. andMartin, & Cunningham, C. W. (2008). Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, *57*(6), 920–938.
URL `http://dx.doi.org/10.1080/10635150802570791`

Regier, J. C., Shultz, J. W., & Kambic, R. E. (2005a). Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proceedings. Biological sciences*, *272*(1561), 395–401.

Regier, J. C., Wilson, H. M., & Shultz, J. W. (2005b). Phylogenetic analysis of myriapoda using three nuclear protein-coding genes. *Molecular Phylogenetics and Evolution*, *34*(1), 147–158.
URL `http://dx.doi.org/10.1016/j.ympev.2004.09.005`

Rodríguez, J. L., F. amd Oliver, Marín, A., & Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, *142*(4), 485–501.

Ronquist, F., & Huelsenbeck, J. P. (2003). Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572.

Rota-Stabelli, O., & Telford, M. L. (2008). A multi criterion approach for the selection of optimal outgroups in phylogeny: Recovering some support for Mandibulata over Myriochelata using mitogenomics. *Molecular Phylogenetics and Evolution*, *48*(1), 103–111.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.
URL `http://mbe.oxfordjournals.org/cgi/content/abstract/4/4/406`

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467.

Schluenzen, F., Tocilj, A., Zarivach, R., & Harms, J. (2000). Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*, *102*, 615–623.

Schöninger, M., & von Haeseler, A. (1994). A stochastic model for the evolution of autocorrelated DNA sequences. *Molecular Phylogenetics and Evolution*, *3*(3), 240–247.

Sharov, A. G. (1966). Basic arthropodan stock with special reference to insects. *Pergamon Press, Oxford*, (pp. 1–271).

Shultz, J. W., & Regier, J. C. (2000). Phylogenetic analysis of arthropods using two nuclear protein-encoding gene ssupports a crustacean + hexapod clade. *Proceedings of the Royal Society of London B Biologcial Science*, *267*(1447), 1011–1019.

Siebert, S., & Backofen, R. (2005). Marna: multiple alignment and consensus structure prediction of rnas based on sequence structure comparisons. *Bioinformatics*, *21*(16), 3352–3359.

Sierwald, P., & Bond, J. E. (2007). Current status of the myriapod class diplopoda (millipedes): Taxonomic diversity and phylogeny. *Annual Review of Entomology*, *52*(1), 401–420. PMID: 17163800.
URL http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.ento.52.111805.090210

Sierwald, P., Shear, W. A., Shelley, R. M., & Bond, J. E. (2003). Millipede phylogeny revisited in the light of the enigmatic order siphoniulida. *Journal of Zoological Systematics and Evolutionary Research*, *41*, 87.

Snodgrass, R. E. (1938). Evolution of the Annelida, Onychophora, and Arthropoda. *Smithsonian Miscellaneous Collections*, *97*, 1–159.

Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G., & Frank, J. (2001). Structure of the 80S ribosome from Saccharomyces cerevisiae tRNA ribosome and subunit subunit interactions. *Cell*, *107*, 373–386.

Stocsits, R. R., Letsch, H., Hertel, J., Misof, B., & Stadler, P. F. (2008). *RNAsalsa*. Zoologisches Forschungsmuseum A. Koenig, Bonn, version 0.7.3, available from http://rnasalsa.zfmk.de ed.

Swofford, D. L. (1998). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*. Sinauer Associates, Sunderland, Massachusetts, version 4 ed.

Turbeville, J. M., Pfeifer, D. M., Field, K. G., & Raff, R. A. (1991). The phylogenetic status of arthropods, as inferred from 18s rrna sequences. *Molecular Biology and Evolution*, *8*(5), 669–686.

Ungerer, P., & Scholtz, G. (2008). Filling the gap between identified neuroblasts and neurons in crustaceans adds new support for Tetraconata. *Proceedings. Biological sciences*, *275*(1633), 369–376.

von Reumont, B. M., Meusemann, K., Szucsich, N. U., Dell'Ampio, E., Gowri-Shankar, V., Bartel, D., Simon, S., Letsch, H. O., Stocsits, R. R., Luan, Y., Wägele, J. W., Pass, G., Hadrys, H., & Misof, B. (in review). Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? a case study on major arthropod relationships. In review.

Wägele, J. W., & Mayer, C. (2007). Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*, *7*(1), 147.

Willmann, R. (2003). Die phylogenetischen Beziehungen der Insekten: Offene Fragen und Probleme. *Verhandlungen des Naturwissenschaftlichen Vereins in Hamburg*, (pp. 1–64).

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science*, *292*, 883–896.

Zrzavý, J., & Štys, P. (1997). The basic body plan of arthropods: insights from evolutionary morphology and developmental biology. *Journal of Evolutionary Biology*, *10*(3), 653–367.

Zuccon, A., & Zuccon, D. (2006). *MrEnt v1.2. [Computer software distributed by the authors]*. Stockholm: Swedish Museum of Natural History, Department of Vertebrate Zoology & Molecular Systematics Laboratory.

A Appendix

A.1 Base frequencies for the full character set

Full character set

Manual approach					Automated approach				
Taxon	**A**	**C**	**G**	**T**	**Taxon**	**A**	**C**	**G**	**T**
Scutigerella sp.(1)	0.20	0.32	0.31	0.17	*Scutigerella* sp.	0.20	0.31	0.30	0.18
Hanseniella sp.	0.20	0.31	0.30	0.18	*Scutigerella* sp.(1)	0.20	0.32	0.32	0.16
S. mutilans	0.20	0.29	0.32	0.18	*C. ferrugineus*	0.21	0.29	0.33	0.17
C. hartmeyeri	0.20	0.29	0.32	0.18	*Hanseniella* sp.	0.21	0.31	0.31	0.17
C. tasmanianus	0.20	0.30	0.32	0.18	Cyclopidae sp.	0.21	0.27	0.33	0.19
C. tasmanianus(1)	0.20	0.30	0.32	0.18	*Triops* sp.	0.21	0.28	0.32	0.19
Scutigerella sp.	0.20	0.30	0.30	0.19	S. mutilans	0.21	0.29	0.32	0.18
L. forficatus	0.21	0.29	0.32	0.18	*C. hartmeyeri*	0.21	0.29	0.32	0.18
Triops sp.	0.21	0.27	0.31	0.21	*C. tasmanianus*	0.21	0.30	0.32	0.18
O. politus	0.21	0.29	0.31	0.18	*C. tasmanianus*(1)	0.21	0.30	0.31	0.18
Colossendeis sp.	0.21	0.29	0.31	0.19	*Colossendeis* sp.	0.21	0.29	0.31	0.18
C. ferrugineus	0.21	0.29	0.31	0.18	*C. georgiana*	0.21	0.28	0.31	0.19
C. georgiana	0.21	0.28	0.31	0.19	*Monographis* sp.	0.21	0.27	0.31	0.20
Cyclopidae sp.	0.21	0.27	0.32	0.20	*P. imperator*	0.21	0.28	0.31	0.19
P. imperator	0.21	0.28	0.31	0.19	*L. forficatus*	0.22	0.30	0.31	0.18
Monographis sp.	0.21	0.27	0.31	0.21	*M. religiosa*	0.22	0.28	0.32	0.18
C. flavidus	0.22	0.29	0.31	0.18	Polyxendidae sp.	0.22	0.28	0.31	0.19
Polyxendidae sp.	0.22	0.28	0.31	0.20	*O. politus*	0.22	0.29	0.31	0.18
S. coleoptrata	0.22	0.29	0.30	0.19	*P. lagurus*	0.22	0.28	0.31	0.19
P. lagurus	0.22	0.28	0.30	0.20	*S. empusa*	0.22	0.29	0.31	0.19
S. coleoptrata(1)	0.22	0.29	0.30	0.19	**Mean**	**0.22**	**0.28**	**0.31**	**0.19**
Mean	**0.22**	**0.28**	**0.30**	**0.20**	*C. flavidus*	0.22	0.30	0.31	0.18
M. religiosa	0.22	0.27	0.31	0.20	*S. rubens*	0.22	0.28	0.32	0.18
P. complanatus	0.22	0.27	0.30	0.21	*Allopauropus* sp.	0.22	0.26	0.31	0.20
S. rubens	0.22	0.28	0.31	0.19	Paradoxosomatidae sp.	0.22	0.27	0.30	0.20
Orthoporus sp.	0.22	0.27	0.30	0.20	*M. hrabei*	0.22	0.28	0.31	0.18
Paradoxosomatidae sp.	0.22	0.27	0.30	0.21	*P. complanatus*	0.22	0.27	0.30	0.20
Cryptops sp.	0.22	0.28	0.30	0.20	Cyprididae sp.	0.23	0.25	0.30	0.22
S. empusa	0.22	0.27	0.30	0.21	*L. polyphemus*	0.23	0.28	0.31	0.19
L. polyphemus	0.23	0.27	0.31	0.20	*P. argus*	0.23	0.27	0.31	0.20
Allopauropus sp.	0.23	0.26	0.31	0.21	*S. coleoptrata*	0.23	0.29	0.30	0.18
C. caeruleocinctus	0.23	0.27	0.28	0.22	*Argulus* sp.	0.23	0.28	0.29	0.20
Cyprididae sp.	0.23	0.25	0.29	0.23	*Orthoporus* sp.	0.23	0.27	0.30	0.20
Argulus sp.	0.23	0.27	0.29	0.21	*S. coleoptrata*(1)	0.23	0.29	0.29	0.19
Eremobates sp.	0.23	0.27	0.30	0.20	*Cryptops* sp.	0.23	0.28	0.29	0.19
M. hrabei	0.23	0.27	0.30	0.20	*C. caeruleocinctus*	0.23	0.27	0.29	0.21
P. argus	0.24	0.26	0.29	0.21	*C. longicaudata*	0.23	0.27	0.29	0.21
Pauropoda sp.	0.24	0.26	0.29	0.21	*Eremobates* sp.	0.23	0.27	0.31	0.19
C. longicaudata	0.24	0.25	0.28	0.23	*M. tardigradum*	0.23	0.25	0.28	0.23
M. tardigradum	0.24	0.24	0.29	0.23	Pauropoda sp.	0.24	0.26	0.30	0.21
S. viridis	0.24	0.25	0.29	0.22	*Symphylella* sp.(1)	0.24	0.28	0.27	0.21
Symphylella sp.	0.24	0.28	0.28	0.20	*Symphylella* sp.	0.24	0.28	0.27	0.21
Symphylella sp.(1)	0.25	0.28	0.27	0.20	*S. viridis*	0.24	0.26	0.29	0.21
χ^2=130937.530; df=123; p=0.29; nt=952					χ^2=97334.994; df=123; p=0.96; nt=841				

A.2 Base frequencies for the reduced character set

Reduced character set

Manual approach					Automated approach				
Taxon	**A**	**C**	**G**	**T**	**Taxon**	**A**	**C**	**G**	**T**
Scutigerella sp.	0.21	0.3	0.28	0.21	*Scutigerella* sp.	0.22	0.3	0.28	0.2
Scutigerella sp.(1)	0.22	0.31	0.29	0.19	*Triops* sp.	0.22	0.27	0.31	0.2
Hanseniella sp.	0.22	0.3	0.28	0.19	*C. ferrugineus*	0.23	0.28	0.31	0.19
Triops sp.	0.22	0.26	0.3	0.22	*Scutigerella* sp.(1)	0.23	0.31	0.29	0.18
C. hartmeyeri	0.22	0.27	0.3	0.2	Cyclopidae sp.	0.23	0.26	0.3	0.21
L. forficatus	0.22	0.27	0.3	0.21	*Hanseniella* sp.	0.23	0.31	0.28	0.18
S. mutilans	0.22	0.27	0.3	0.2	*Colossendeis* sp.	0.23	0.28	0.29	0.2
C. ferrugineus	0.23	0.28	0.3	0.2	*L. forficatus*	0.23	0.28	0.29	0.2
C. tasmanianus	0.23	0.27	0.3	0.21	*M. religiosa*	0.23	0.27	0.3	0.2
C. tasmanianus(1)	0.23	0.27	0.3	0.21	*C. hartmeyeri*	0.23	0.27	0.29	0.2
Monographis sp.	0.23	0.26	0.29	0.23	*P. imperator*	0.23	0.27	0.29	0.2
O. politus	0.23	0.27	0.29	0.21	*S. mutilans*	0.23	0.27	0.3	0.2
P. imperator	0.23	0.27	0.3	0.21	*C. tasmanianus*	0.24	0.27	0.29	0.2
C. georgiana	0.23	0.27	0.29	0.22	*C. tasmanianus*(1)	0.24	0.28	0.29	0.2
Colossendeis sp.	0.23	0.27	0.3	0.2	*Monographis* sp.	0.24	0.26	0.28	0.22
Polyxendidae sp.	0.23	0.26	0.29	0.22	*C. georgiana*	0.24	0.27	0.28	0.21
Cyclopidae sp.	0.23	0.25	0.3	0.22	Polyxendidae sp.	0.24	0.27	0.28	0.21
C. flavidus	0.23	0.27	0.29	0.2	*M. hrabei*	0.24	0.27	0.3	0.19
Mean	**0.24**	**0.26**	**0.29**	**0.21**	*O. politus*	0.24	0.27	0.29	0.2
Allopauropus sp.	0.24	0.25	0.3	0.21	*S. empusa*	0.24	0.27	0.28	0.21
M. religiosa	0.24	0.26	0.3	0.21	**Mean**	**0.24**	**0.27**	**0.28**	**0.21**
Orthoporus sp.	0.24	0.26	0.29	0.22	*Allopauropus* sp.	0.24	0.25	0.29	0.21
P. lagurus	0.24	0.26	0.28	0.22	*P. lagurus*	0.24	0.27	0.28	0.21
P. complanatus	0.24	0.25	0.28	0.23	*S. rubens*	0.24	0.27	0.29	0.2
S. empusa	0.24	0.25	0.28	0.23	*C. longicaudata*	0.24	0.26	0.29	0.21
S. rubens	0.24	0.27	0.29	0.2	*M. tardigradum*	0.24	0.25	0.27	0.23
M. hrabei	0.24	0.26	0.29	0.21	*C. flavidus*	0.25	0.27	0.28	0.2
Paradoxosomatidae sp.	0.24	0.26	0.28	0.23	Paradoxosomatidae sp.	0.25	0.26	0.27	0.22
S. coleoptrata	0.24	0.26	0.28	0.21	Pauropoda sp.	0.25	0.25	0.29	0.21
C. caeruleocinctus	0.24	0.25	0.27	0.23	*P. complanatus*	0.25	0.25	0.27	0.23
Cryptops sp.	0.24	0.26	0.28	0.22	Cyprididae sp.	0.25	0.24	0.27	0.25
Cyprididae sp.	0.24	0.24	0.27	0.25	*Orthoporus* sp.	0.25	0.25	0.28	0.22
Pauropoda sp.	0.24	0.25	0.29	0.22	*P. argus*	0.25	0.25	0.28	0.22
Eremobates sp.	0.24	0.26	0.29	0.21	*Argulus* sp.	0.25	0.26	0.27	0.21
S. coleoptrata(1)	0.24	0.26	0.28	0.21	*C. caeruleocinctus*	0.25	0.25	0.27	0.23
Argulus sp.	0.25	0.25	0.27	0.23	*Cryptops* sp.	0.25	0.26	0.27	0.21
C. longicaudata	0.25	0.25	0.28	0.23	*L. polyphemus*	0.25	0.26	0.28	0.2
L. polyphemus	0.25	0.26	0.29	0.21	*S. viridis*	0.25	0.25	0.28	0.21
M. tardigradum	0.25	0.24	0.28	0.23	*Eremobates* sp.	0.25	0.26	0.28	0.2
P. argus	0.25	0.24	0.27	0.23	*S. coleoptrata*	0.26	0.27	0.27	0.2
S. viridis	0.25	0.24	0.27	0.23	*S. coleoptrata*(1)	0.26	0.27	0.27	0.2
Symphylella sp.	0.26	0.28	0.25	0.21	*Symphylella* sp.(1)	0.26	0.27	0.25	0.22
Symphylella sp.(1)	0.26	0.28	0.25	0.21	*Symphylella* sp.	0.26	0.27	0.25	0.22
χ^2=87256.208; df=123; p=0.99; nt=764					χ^2=70688.921; df=123; p=1.00; nt=648				

Curriculum vitae

Contact

Daniela Bartel
Schuhmeierplatz 3/7
1160 Vienna
Austria
Tel.: 0043 699 10957686
email: daniela.bartel1@chello.at

Personal data

Date and place of birth: 22^{nd} of Oct. 1978 in Vienna, Austria

Marital status: unmarried

Education

1984 - 1988	Public school, Vienna
1988 - 1992	Grammar school, Vienna
1992 - 1998	Secondary School of Economics and Tourism, Vienna

Study

1998 - 2003	1st period - Biology, University of Vienna
2003 - 2009	2nd period - Zoology, University of Vienna
February 2008	Research stay at the Forschungsmuseum A. Koenig - working group Prof. Dr. Bernhard Misof, Bonn, Germany

Topic of the diploma thesis

Did myriapods evolve from a single common ancestor?
Optimizing the phylogenetic signal of 28S rRNA sequences for myriapods.

Career history

July - Sept.1996	Trainee at the "*Grand Hotel*", Zell am See, Salzburg
July - Aug.1997	Trainee at the retirement home "*Auf der Schmelz*", Vienna
July - Aug.1998	Trainee in the company "*Multicopy*", Vienna
2000 - 2005	Asstistance in the parental company "*Riehs*" (Jewelry store)
Sept.2006	Assistance in the Organisation at the Congress of *IMST* (**I**nnovationen **M**achen **S**chulen **T**op), Vienna
Sept.2006 - Dec.2007	Company *AGO*
Oct. - Dec.2007	Sampling and determination of basal hexapods for an EST project in the frame of the FWF project *"17038 Phylogenie der basalen Hexapoden"*
since Jan.2008	Laboratory assistant at the Research group for "*Molecular Systematics*" at the National History Museum, Vienna

Scientific publications

von Reumont B. M., Meusemann K., Szucsich N. U., Dell'Ampio E., Gowri-Shankar V., **Bartel D.**, Simon S., Letsch H. O., Stocsits R. R., Luan Y., Wägele J. W., Pass G., Hadrys H., Misof B. (in press), Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. *BMC Evolutionary Biology*

Bartel D., Szucsich N. U., Dell'Ampio E., Pass G. (2009), Did myriapods evolve from a single common ancestor? Optimizing the phylogenetic signal of 28S rRNA sequences for myriapods. *Proceedings: Celebrating Darwin: From the origin of species to deep metazoan phylogeny*, pp 68