SOIL ORGANISMS Volume 83 (3) 2011 pp. 347–358

ISSN: 1864-6417

Protura and molecular phylogenetics: status quo of a young love

Emiliano Dell'Ampio[1, 3], Nikolaus Urban Szucsich[2] and Günther Pass[1]

[1] *Department of Evolutionary Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria*

[2] *Department of Entomology, Biocenter Grindel and Zoological Museum, University of Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany*

[3] *Corresponding author: Emiliano Dell'Ampio (e-mail: emiliano.dell.ampio@univie.ac.at)*

Abstract

Protura are among the latecomers to molecular phylogenetics. The first sequences were published about a decade ago; since then relatively little additional data has been collected. Nonetheless, the available molecular phylogenetic analyses have provided valuable contributions towards the elucidation of the phylogenetic position of the Protura. All studies endorse Protura as closely related to hexapods. To detail, remarkable results were obtained with respect to the potential hexapod sister group. Only few molecular studies support the traditional Ellipura hypothesis (Protura plus Collembola). Almost all corroborate a sister group relationship between Protura and Diplura, which previously was never considered by morphologists. Besides the absence of eyes (hence the name Nonoculata proposed by Luan et al. 2005), few further substantial morphological synapomorphies have yet been mentioned to support this taxon. With respect to the internal relationships among the different groups of Protura, molecular data indicates monophyly of Acerentomata and Eosentomata, while Sinentomata seem to be para- or polyphyletic; a result also obtained in some morphological analyses. However, since the molecular taxon sampling remains meagre, any conclusions must be treated as preliminary. Requirements for further molecular investigations are pointed out.

Keywords: Hexapoda, Nonoculata, Ellipura, mitochondrial genomes, EST

1. Historical background

The phylogenetic position of the Protura has been a point of controversy *ab initio*. The first to describe these unusual arthropods, Silvestri (1907), classified them as an order of Insecta Apterygota. Berlese (1909), on the contrary, concluded from his comprehensive investigations that Protura are a separate arthropod clade related to myriapods and insects. This controversial debate was of great concern to the scientific community in the first years after the discovery of these animals. However, the view that Protura are true insects becomes subsequently widely accepted (for details on the history of research in Protura see the contribution by Pass & Szucsich in the same issue). Börner (1910) considered the Protura as the sister group of Collembola and coined the resulting clade Ellipura. This taxon can be found in the cladograms of almost all zoological and entomological textbooks up to the present day. However, there are current reports which again question the placement of Protura within Hexapoda, drawing attention to certain idiosyncratic morphological characters and some embryological features which deviate from ordinary hexapod development (e.g. Dallai 1991, Machida 2006, Dallai et al. 2010).

In light of this controversial background the demand for molecular investigations appears imperative to clarify the phylogenetic position of Protura. Nevertheless, this taxon is among the latecomers to molecular phylogenetics. The first published molecular analyses that included a representative of Protura (*Nipponentomon* sp.) appeared in 1998 (Colgan et al. 1998, Wheeler 1998), and data collection up to now remains meagre (Table 1). Despite this, the published molecular studies have provided valuable contributions toward the elucidation of the phylogenetic position of Protura. The present review gives an overview on the status quo. Special attention is drawn to (i) describe and discuss conflicting points between molecular and traditional hypotheses based on morphology and (ii) to briefly mention emerging areas that should be considered for future investigations.

Abbreviations: EST = Expressed Sequence Tag; EF-1α = Elongation factor-1α; U2 snRNA = small nuclear ribonucleic acid U2; ITS2 = Internal Transcribed Spacer 2; H3 = histone H3; 18S and 28S = nuclear ribosomal 18S and 28S genes; 12S and 16S = mitochondrial ribosomal 12S and 16S genes; GC-bias = guanine-cytosine bias; BI = Bayesian inference; ML = Maximum likelihood; ME = Minimum evolution; MP = Maximum parsimony; GTR+I+G = general time-reversible plus invariant plus gamma model of sequence evolution; Mk model = Markov k-state (number of states observed) parameter model for discrete morphological data.

2. Proturan sequences: A sparse and patchy data collection

Despite the exponential growth of molecular phylogenetic investigations in arthropods, a scan for publications which include at least a single sequence of Protura reveals that the available documentation, when compared to other hexapod taxa, is currently still scarce: Colgan et al. (1998), Wheeler (1998), Carpenter & Wheeler (1999), Shao et al. (1999), Carapelli et al. (2000), Edgecombe et al. (2000), Giribet & Ribera (2000), Shao et al. (2000), Giribet & Wheeler (2001), Giribet et al. (2001), Wheeler et al. (2001), Zhang et al. (2001), D'Haese (2002), Luan et al. (2003), Giribet et al. (2004), Kjer (2004), Luan et al. (2004), Giribet et al. (2005), Luan et al. (2005), Kjer et al. (2006), Mallatt & Giribet (2006), Misof et al. (2007), Gao et al. (2008), Dell'Ampio et al. (2009), von Reumont et al. (2009), Xie et al. (2009), Koenemann et al. (2010), Mallatt et al. (2010), Meusemann et al. (2010), Andrew (2011). However, a closer look to the above listed literature reveals that only few of these studies generated new sequence data. Moreover not all of them are openly accessible via GenBank (http://www.ncbi.nlm.nih.gov/genbank). For example, no data is present in this database from the studies of Wheeler (1998), Carpenter & Wheeler (1999), Shao et al. (1999, 2000) and Zhang et al. (2001). Wheeler (1998) and Carpenter & Wheeler (1999) cited publications 'in press' as sources of their sequences, which in their published form deviated from the anticipated taxon sampling (e.g. different proturan species or none). Scanning of the database (Tab. 1) reveals a relatively low number of sequences which moreover exhibit a strong bias toward the nuclear ribosomal RNA genes 18S and 28S. Alternative markers are represented by few and generally short sequence entries only. Finally, no complete mitochondrial genome is available so far (please see Addendum by chapter 7). The largest genomic data collection (both nuclear and mitochondrial) is thus restricted to a single entry, the Expressed Sequence Tag (EST) data of *Acerentomon franzi* Nosek, 1965 (Meusemann et al. 2010).

Tab. 1 List of GenBank entries on proturan sequences (17.05.11).

Species		Gene	Accession no.
ACERENTOMATA			
Acerella muscorum	(Ionesco, 1930)	rRNA 28S	AF483354
			AF483412
Acerentomon franzi	Nosek, 1965	rRNA 18S	EU368597
		rRNA 28S	EF199976
		histone H3	FJ664243
		EST data	from FN186135 to FN190445
Acerentomon sp.		EF-1α	AH009876
Acerentomon sp.		rRNA 12S	AF252395
Acerentulus traegardhi	Ionesco, 1937	rRNA 18S	AF005453
			AF173233
			AF372424
		rRNA 28S	AF005469
			AF354687
Nipponentomon sp.		histone H3	AF110861
		U2 snRNA	AF110890
Baculentulus densus	(Imadaté, 1960)	rRNA 28S	EU376049
Baculentulus tienmushanensis	(Yin, 1963)	rRNA 18S	AY037169
		rRNA 5.8S	AY859554
		ITS2	AY859554-55
		rRNA 28S	AY859555-56
			AF416872
			DQ016591
			EF192433
Gracilentulus maijiawensis	Yin & Imadaté, 1979	rRNA 28S	EF192435
Gracilentulus shipingensis	Yin, 1984	rRNA 18S	AY596354
		rRNA 28S	AY596374
Kenyentulus ciliciocalyci	Yin, 1987	rRNA 18S	AY145139
		rRNA 28S	AY596375
*Hesperentomon hwashanense**	Yin, 1982	rRNA 18S	AY596353
		rRNA 28S	AY596373
Neocondeellum dolichotarsum	(Yin, 1977)	rRNA 18S	AY037170
		rRNA 28S	AF416873
EOSENTOMATA			
*Eosentomon commune***	Yin, 1965	rRNA 18S	AY596356
		rRNA 28S	AY596377
Eosentomon sakura	Imadaté & Yosii, 1959	rRNA 18S	AY596355
		rRNA 28S	AY596376
			EF192434
Eosentomon sp.		rRNA 18S	AF497743
Eosentomon sp.			EU368598
Eosentomon sp.		rRNA 28S	EU376047
Paranisentomon triglobulum	(Yin & Zhang, 1982)	rRNA 18S	AY596357
		rRNA 28S	AY596378
'SINENTOMATA'			
Fujientomon dicestum	Yin, 1977	rRNA 18S	AY596359
		rRNA 28S	AY596380
Sinentomon erythranum	Yin, 1965	rRNA 18S	AY596358
		rRNA 28S	AY596379
			EF192442

Notes: Accession no. = accession numbers; EST = Expressed Sequence Tag; EF-1α=Elongation factor-1α; U2 snRNA = U2 small nuclear ribonucleic acid U2; ITS2 = Internal Transcribed Spacer 2; * = listed as *Hesperentomon huashanensis*; ** = listed as *Eosentomon communis*.

Protura are considered to encompass three main groups: Acerentomata, Eosentomata and Sinentomata (Yin 1996, Szeptycki 2007). Eosentomata can be easily distinguished from Acerentomata by the presence of a tracheal system and two-segmented abdominal appendages on segments one to three. On the contrary, the erection of Sinentomata is mainly based on ultrastructural details of sperm and pseudoculi. Therefore this group cannot be easily understood by characters of the habitus alone. Only very few of the present molecular phylogenetic studies on hexapods include representatives (at least one species) of all three subgroups. All investigations are based on the complete (or almost complete) rRNA 18S gene sequence, which was used either alone (Xie et al. 2009) or in combination with a partial or almost complete sequence of the rRNA 28S gene (Luan et al. 2005, Gao et al. 2008, von Reumont et al. 2009).

3. The sister group of Protura: Ellipura versus Nonoculata

Almost all studies discussed in the present review address either arthropod or at least pancrustacean phylogeny. All of them favour a placement of Protura within Hexapoda. Whenever Protura cluster outside hexapods in single molecular analyses, either the statistical support is very low, or factors are mentioned which may explain the erroneous placement (e.g. Colgan et al. 1998, Giribet & Ribera 2000, Dell'Ampio et al. 2009, Koenemann et al. 2010).

Most morphologists favour a sister group relationship between Protura and Collembola, within the taxon Ellipura (e.g. Hennig 1953, Kristensen 1981, 1997, Bitsch & Bitsch 2000, 2004) or Parainsecta (Kukalová-Peck 1987). Ellipura is obtained by only few molecular studies. Weak support for this taxon is reported from three single gene analyses based on mitochondrial genes (Shao et al. 1999, Carapelli et al. 2000, Zhang et al. 2001). Additional support for Ellipura is found in some combined analyses, but only under precise parameter sets (for details, see Wheeler 1998, Carpenter & Wheeler 1999, Edgecombe et al. 2000, Wheeler et al. 2001). The exiguous molecular support mirrors the weak morphological support for Ellipura (Ax 1999). Such weakly supported taxa usually are only accepted in the absence of any support for an alternative relationship.

However, analyses of most molecular datasets mentioned here gave rise to a new phylogenetic hypothesis, i.e. the sister group relationship between Protura and Diplura. The resulting clade was named 'Nonoculata' by Luan et al. (2005), stressing the absence of eyes in the proposed sister groups. Nonoculata is supported in almost all analyses based on one or both nuclear ribosomal genes (Giribet & Wheeler 2001, D'Haese 2002, Luan et al. 2003, Giribet et al. 2004, Kjer 2004, Luan et al. 2004, Giribet et al. 2005, Luan et al. 2005, Kjer et al. 2006, Mallatt & Giribet 2006, Misof et al. 2007, Gao et al. 2008, Dell'Ampio et al. 2009, von Reumont et al. 2009, Xie et al. 2009, Koenemann et al. 2010, Mallatt et al. 2010). Only few studies based on these markers do not support Nonoculata. All these studies used direct optimization method (Wheeler 1996) and are based on parsimony analyses alone (Wheeler 1998, Carpenter & Wheeler 1999, Giribet & Ribera 2000, Giribet et al. 2001, Wheeler et al. 2001). In these analyses Nonoculata is retrieved only in few instances when nucleotide data alone are considered (Giribet & Ribera 2000, Wheeler et al. 2001).

More recent analyses of nuclear ribosomal rRNA genes found support for Nonoculata, irrespective of methods and software used for alignment and phylogenetic reconstruction. Giribet et al. (2004, 2005) used direct optimization as implemented in the software POY (Wheeler et al. 2002) and conducted sensitivity analyses. The analyses are based on a pool of markers (including ribosomal genes) built up by a total of five or nine loci, respectively.

The most parsimonious cladograms of both the total evidence analysis (Giribet et al. 2004, 2005) and the molecular combined analysis (five loci of Giribet et al. 2004) result in a sister group relationship of Protura and Diplura. Although the statistical support is quite weak in all instances (jackknife frequencies above 50% of the replicates only in one analysis, see Fig. 3 of Giribet et al. 2004), the clade is stable irrespective of tested parameter sets (Giribet et al. 2004, 2005). In a completely different approach, Kjer et al. (2006) likewise retrieved Nonoculata with maximum statistical support. In that study a combined super matrix (18S + 28S + H3 + EF-1α + 12S + 16S + intervening tRNA + 170 morphological characters) was reconstructed under partitioned Bayesian inference (BI) by specifying distinct models of evolution for nucleotides (GTR+I+G) and morphological characters (Mk model). Nonoculata occurred in the majority-rule consensus tree and in trees inferred by: (i) weighted parsimony (ribosomal genes plus amino acid sequences from protein coding genes), (ii) BI with a site-specific rate model (all molecular loci) and (iii) BI restricted to nuclear rRNA genes with a GTR+I+G substitution model (for details, see Kjer et al. 2006).

Luan et al. (2005) not only retrieved support for Nonoculata, but even found that Ellipura is statistically rejected by the ML-based parametric-bootstrap test. Despite these pieces of evidence, the 'Nonoculata clade' has long been suspected to be an artefact caused by a shared nucleotide bias (a GC-bias) in combination with a long-branch attraction effect (Luan et al. 2005, Mallatt & Giribet 2006, Gao et al. 2008, Dell'Ampio et al. 2009). In all analyses both Diplura and Protura are characterised by extremely long branches and a high proportions of guanine and cytosine in their ribosomal sequences (e.g. Luan et al. 2005, Kjer et al. 2006, Mallatt & Giribet 2006, Gao et al. 2008, Dell'Ampio et al. 2009). This is even more important, since Collembola have neither long branches (e.g. Kjer et al. 2006, von Reumont et al. 2009), nor biased gene sequences (Luan et al. 2005 [Tab. 3], Mallatt & Giribet 2006 [Suppl. Tab. S2], Misof et al. 2007 [Fig. 1], Dell'Ampio et al. 2009 [Suppl. Tab. S5], Mallatt et al. 2010 [Suppl. Mat. S1]). Inhomogeneity of base frequencies across taxa violates the assumption of stationarity, common to all tree reconstruction methods. If not properly modeled it can mislead all commonly used methods, including Maximum likelihood and Bayesian inference (e.g. Foster & Hickey 1999, Omilian & Taylor 2001, Jermin et al. 2004, Gowri-Shankar & Rattray 2006).

Among hexapods the GC-bias is especially accentuated in Diplura (e.g. Luan et al. 2005, Mallatt & Giribet 2006, Misof et al. 2007, von Reumont et al. 2009). Luan et al. (2005) noticed that selective exclusion of all diplurans from their taxon sampling delivers a 'ribosomal data matrix' in which stationarity of nucleotide frequencies was restored. This fact is also evident in three of the seven datasets tested by von Reumont et al. (2009). Accordingly, Dell'Ampio et al. (2009) demonstrated that among the seven 'nonoculatan species' (two proturans and five diplurans) included in their study, only the Diplura failed the χ^2 homogeneity-test of base frequencies available in the software TREE-PUZZLE (Schmidt et al. 2002). Interestingly, the analyses based on the resulting data matrices (with and without hypervariable regions) show that the Ellipura are only reconstructed (with low statistical support) when the most conserved fragments of the gene ('core regions') are considered. In the study of Mallatt & Giribet (2006) and in the remaining four datasets considered by von Reumont et al. (2009), stationarity is restored by the simultaneous exclusion of Diplura with few additional 'non-dipluran' taxa (e.g. *Peripatus* sp., some Diptera and in one case *Sinentomon erythranum* Yin, 1965).

Support for Nonoculata as a valid taxon comes additionally from three recent studies focused on arthropod phylogeny (von Reumont et al. 2009, Meusemann et al. 2010, Andrew 2011). To avoid the above mentioned-misleading effects due to non stationary processes, von Reumont et al. (2009) used simultaneous application of a mixed DNA/RNA model coupled with a time-heterogeneous approach and new heuristics to exclude randomly aligned sections of the data matrix (ALISCORE [Misof & Misof 2009]). Using this analytical method, the evolution of nuclear rRNA genes should be modeled more realistically. While some unorthodox nodes are corrected by the application of a time-heterogeneous model, Protura remain as the sister group of Diplura with maximum support.

Meusemann et al. (2010) and Andrew (2011) provide the only analyses independent of rRNA genes that support the taxon Nonoculata. Both studies have a phylogenomic approach based on EST data with an overlapping strategy for orthologous gene prediction by the usage of the software HaMStR (Ebersberger et al. 2009). Meusemann et al. (2010) show a maximum support for Nonoculata. Andrew (2011) contributed no new EST's to the dataset but used alternative alignment masking and gene/taxa-matrix reduction strategies. Although the resulting tree is generally less reliable than the original one of Meusemann et al. (2010), it does not differ with respect to the entognathous taxa.

4. The internal relationships of Protura

Of the three main subgroups of Protura: Acerentomata, Eosentomata and Sinentomata, only the monophyly of the Sinentomata has provoked controversial discussion among proturologists. On the one hand, ultrastructural characters of sperm and morphological details of the pseudoculi are mentioned in favour of the monophyly of Sinentomata (Dallai & Yin 1983, Xue & Yin 1990, Yin et al. 1990, Yin & Xue 1993). On the other hand, the genus *Fujientomon* Imadaté, 1964 did not cluster with *Sinentomon* Yin, 1965 in a cladistic analysis based on morphological characters of representatives from 71 genera of Protura, but clustered with Protentomidae inside Acerentomata (François 2003).

Interestingly, none of these two competing hypotheses, Fujientomidae + Sinentomidae or Fujientomidae + Protentomidae, is supported in any molecular study. However, all studies give strong support to the monophyly of both Acerentomata and Eosentomata (Tab. 2).

Molecular analyses, which permit conclusions concerning the phylogenetic relationships of Acerentomata, Eosentomata, Fujientomidae and Sinentomidae (the latter two often subsumed as Sinentomata), are restricted to four publications (Luan et al. 2005, Gao et al. 2008, von Reumont et al. 2009, Xie et al. 2009). Of these only the molecular analyses of Luan et al. (2005) and Xie et al. (2009) included representatives of both Fujientomidae and Sinentomidae (*Fujientomon dicestum* Yin, 1977 and *Sinentomon erythranum*). These two species never clustered together, leaving Sinentomata either paraphyletic (Luan et al. 2005) or polyphyletic (Xie et al. 2009). However, the relative positions of *Fujientomon* and *Sinentomon* could not be assessed with any confidence. Without directly mentioning it, the rRNA 18S secondary structure data of Xie et al. (2009), however, reveal a possible correspondence in length variation between *Sinentomon* and *Fujientomon*.

In particular, *Sinentomon erythranum* emerged as the sister taxon of a clade Acerentomata + Eosentomata in the molecular analysis with the most comprehensive taxon sampling including five representatives of Acerentomata, three Eosentomata and one each of Fujientomidae and Sinentomidae (Luan et al. 2005). Furthermore, *Fujientomon* became the sister to all the remaining Protura (leaving Sinentomata paraphyletic). An alternative clade Sinentomidae + Acerentomata, however, is clearly reconstructed by Gao et al. (2008), von Reumont et al. (2009) and Xie et al. (2009). The statistical support for both competing scenarios was relatively high (Tab. 2).

Tab. 2 Support values for main subgroups of Protura. Only studies are included which contain at least a single representative of Acerentomata, Eosentomata and Fujientomidae or Sinentomidae (see text).

Author	**ACE**	**EOS**	**SIN**	**EOS+ACE**	**fuj+EOS**	**sin+ACE**
Luan et al. 2005						
45 taxa dataset ML	YES	YES	NO	YES	NO	NO
45 taxa dataset BI (p.a.)	1.00	1.00	NO	1.00	NO	NO
45 taxa dataset BI (s.m.a.)	1.00	1.00	NO	1.00	NO	NO
45 taxa dataset ME-LogDet	97%	100%	NO	<5%	NO	NO
45 taxa dataset MP	96%	100%	NO	64%	NO	NO
34 taxa dataset ML	97%	100%	NO	64%	NO	NO
Gao et al. 2008						
BI	1.00	1 species	1 species	NO	--*	1.00
ML	100%	1 species	1 species	NO	--*	80%
ME-LogDet	100%	1 species	1 species	NO	--*	100%
von Reumont et al. 2009						
BI time heterogeneous	1.00	0.99	1 species	NO	--*	<0.70
BI time homogeneous	1.00	0.99	1 species	NO	--*	0.80
Xie et al. 2009						
BI	1 species	1 species	NO	NO	YES	YES

Notes: ACE = Acerentomata; EOS = Eosentomata; SIN = Sinentomata; fuj = Fujientomidae; sin = Sinentomidae; BI = Bayesian inference; ML = Maximum likelihood; ME = Minimum evolution; MP = Maximum parsimony; p.a. = partitioned analysis; s.m.a. = single model analysis; YES = node present in given topology; NO = node not present in given topology; * = no species of Fujientomidae included.

Nonetheless, based on all results no final conclusions can be made, since none of the studies directly focused on internal proturan phylogeny. Moreover, the results are based on nuclear ribosomal genes only, and we stress that support from independent molecular datasets is urgently required.

5. Conclusions and outlook

In sum, the molecular information concerning Protura is currently still scarce, and some of the phylogenetic conclusions must be regarded as preliminary. One crucial reason for caution is the fact that most of the studies are based on the same markers: the nuclear ribosomal 18S and 28S genes. Moreover the current poor taxon sampling is strongly biased towards representatives of Acerentomata (Tab. 1), and therefore permits only limited conclusions with respect to internal relationships among the main subtaxa of Protura. A tree summarizing the present knowledge describes the situation best as unresolved, with a basal polytomy among Eosentomata, Acerentomata, Sinentomidae and Fujientomidae (Fig. 1). Thus, testing the validity of Sinentomata and revealing the reciprocal relationships among the major subgroups are two major tasks of future analyses on internal phylogeny of Protura.

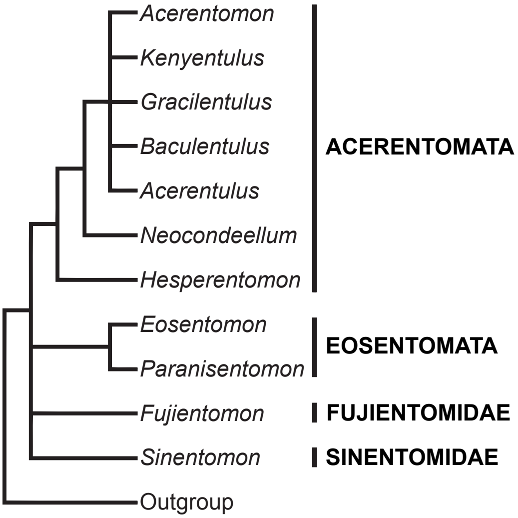

Fig. 1 Tree summarizing the relationships among Protura. The topology is in agreement with most of the studies that include at least two proturan species. Supported nodes include monophyly of Protura, Eosentomata and Acerentomata.

In retrospect, molecular phylogenetics have rightly stimulated the discussion about the phylogenetic position of Protura. With respect to the potential sister group, few analyses support the Ellipura hypothesis, while most of them favour Nonoculata. Determining if Nonoculata is a valid taxon must await corroboration from data other than ribosomal rRNA sequences. Some more recent contributions seem to resolve doubts that the taxon might be simply an artefact due to a shared GC-bias (von Reumont et al. 2009, Meusemann et al. 2010, Andrew 2011). Possible contributions from alternative molecular datasets include analyses of structural motifs in rRNA genes which may give support to phylogenetic hypotheses. By plotting such information on the Bayesian tree, Xie et al. (2009) found some structural motifs in the secondary structure of the 18S rRNA shared exclusively by Protura and Diplura. Phylogenomic analyses based on EST data certainly represent a new frontier in molecular phylogenetics. Nonetheless, this approach is still in its infancy, especially concerning basal hexapods. Additional EST data of all primarily wingless insect groups (but especially from Protura and Diplura) are therefore highly demanded to test for the validity of Nonoculata and to verify the complete lack of alternative signals.

From the other 'genomic side', the publication of complete mitochondrial genomes of two species of Acerentomidae and one of Sinentomidae (Carapelli et al. in prep., Chen et al. 2011 please see Addendum), presented at the VIII International Seminar on Apterygota in Siena (September 2010) will certainly improve the present situation. With these data it should be possible to test whether mitochondrial genes generally tend to support Ellipura (Shao et al. 1999, Carapelli et al. 2000, Zhang et al. 2001). To achieve a more balanced taxon sampling, mitochondrial genome data from Eosentomidae would be beneficial.

Finally it should be noted that the molecular approach has challenged morphologists to search for possible synapomorphies of the taxon Nonoculata (for first attempts, see Luan et al. 2005, Szucsich & Pass 2008, Dell'Ampio et al. 2009). The late love of molecular phylogenetics and Protura thus yielded unexpected fruits which may also enrich other frontiers of zoological research.

6. Acknowledgements

We would like to thank Karen Meusemann for providing valuable comments that helped to improve the manuscript. Daniela Bartel and Gerald Timelthaler helped the authors in the process of GenBank sequence research. Thanks are due to John Plant for checking the English of the manuscript. The authors were supported by grant P 20497-B17 of the Austrian Science Fund (FWF).

7. References

Andrew, D. R. (2011): A new view of insect-crustacean relationships II. Inferences from expressed sequence tags and comparisons with neural cladistics. – Arthropod Structure & Development, in Press, doi:10.1016/j.asd.2011.02.001.

Ax, P. (1999): Das System der Metazoa II. Ein Lehrbuch der Phylogenetischen Systematik. – Fischer, Stuttgart. 384 pp.

Berlese, A. (1909): Monografia dei Myrientomata. – Redia 6: 1–182.

Bitsch, C. & J. Bitsch (2000): The phylogenetic interrelationships of the higher taxa of apterygote hexapods. – Zoologica Scripta **29**:131–156.

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

Börner, C. (1910): Die phylogenetische Bedeutung der Protura. – Biologisches Centralblatt **30**: 633–641.

Carapelli, A., F. Frati, F. Nardi, R. Dallai & C. Simon (2000): Molecular phylogeny of the apterygotan insects based on nuclear and mitochondrial genes. Proceedings of Vth International Seminar on Apterygota (Cordoba 1998) – Pedobiologia **44**: 361–373.

Carapelli, A., Y. Bu, W. J. Chen, Y. X. Luan, F. Nardi, F. Frati, & R. Dallai (2010): New complete mtDNA sequences of Protura: genome evolution and phylogenetic significance. – Oral presentation contribution (Talk SSP3) at the VIII International Seminar on Apterygota (ISA), 12-16 September 2010 (Siena, Italy).

Carpenter, J. M. & W. C. Wheeler (1999): Cladística numérica, análisis simultáneo y filogenia de hexápodos; 'Numerical Cladistics, Simultaneous Analysis and Hexapod Phylogeny.' – In: Melic, A., J. J. de Haro, M. Méndez & I. Ribera (eds.): Filogenia y Evolución de Arthropoda. – Boletin Sociedad Entomológica Aragonesa, **26**: 333–346.

Chen, W. J., Y. Bu, A. Carapelli, R. Dallai, W. Y. Yin & Y. X. Luan (2010): First Complete Mitochondrial Genome of Protura *Sinentomon erythranum*: new evidence on phylogenetic position of the taxa. – Oral presentation contribution (Talk SSP2) at the VIII International Seminar on Apterygota (ISA), 12-16 September 2010 (Siena, Italy).

Colgan, D. J., A. McLauchlan, G. D. F. Wilson, S. P. Livingston, G. D. Edgecombe, J. Macaranas, G. Cassis & M. R. Gray (1998): Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. – Australian Journal of Zoology **46**: 419–437.

Dallai, R. & W. Y. Yin (1983): Sperm structure of *Sinentomon* (Protura) and phylogenetic considerations. – Pedobiologia **25**: 313–316.

Dallai, R. (1991): Are Protura really Insects? – In: Simonetta, A. M. & S. C. Morris (eds): The early evolution of metazoa and the significance of problematic Taxa. – Cambridge University Press, Cambridge: 263–269.

Dallai, R., D. Mercati, Y. Bu, W. Yin, G. Callaini & M. G. Riparbelli (2010): The spermatogenesis and sperm structure of *Acerentomon microrhinus* (Protura, Hexapoda) with considerations on the phylogenetic position of the taxon. – Zoomorphology **129**: 61–80.

Dell'Ampio, E., N. U. Szucsich, A. Carapelli, F. Frati, G. Steiner, A. Steinacher & G. Pass (2009): Testing for misleading effects in the phylogenetic reconstruction of ancient lineages of hexapods: influence of character dependence and character choice in analyses of 28S rRNA sequences. – Zoologica Scripta **38**: 155–170.

D'Haese, C. A. (2002): Phylogeny of apterygote hexapods. Abstracts of the XX Annual Meeting of the Willi Hennig Society. – Cladistics **18**: 220.

Ebersberger, I., S. Strauss & A. von Haeseler (2009): HaMStR: profile hidden markov model based search for orthologs in ESTs. – BMC Evolutionary Biology **9**: 157.

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

Foster, P. G. & D. A. Hickey (1999): Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. – Journal of Molecular Evolution **48**: 284–290.

François, J. (2003): Analyse cladistique de l'ordre des Protoures (Hexapoda, Protura). – Bulletin de la Société Entomologique de France **108**: 109–125.

Gao, Y., Y. Bu & Y. Luan (2008): Phylogenetic Relationships of Basal Hexapods Reconstructed from Nearly Complete 18S and 28S rRNA Gene Sequences. – Zoological Science **25**: 1139–1145.

Giribet, G. & C. Ribera (2000): A Review of Arthropod Phylogeny: New Data Based on Ribosomal DNA Sequences and Direct Character Optimization. – Cladistics **16**: 204–231.

Giribet, G. & W. Wheeler (2001): Some unusual small-subunit ribosomal DNA sequences of metazoans. – American Museum Novitates **3337**: 1–14.

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

Giribet, G., G. D. Edgecombe, J. M. Carpenter, C. A. D'Haese & W. C. Wheeler (2004): Is Ellipura monophyletic? A combined analysis of basal hexapod relationships with emphasis on the origin of insects. – Organisms, Diversity & Evolution **4**: 319–340.

Giribet, G., S. Richter, G. D. Edgecombe & W. C. Wheeler (2005): The position of crustaceans within Arthropoda - evidence from nine molecular loci and morphology. – In: Koenemann, S. & R. A. Jenner (eds): Crustacea and Arthropod Relationships: Crustacean Issues **16**. – Festschrift for F. R. Schram. Taylor & Francis, Boca Raton: 307–352.

Gowri-Shankar, V. & M. Rattray (2006): On the correlation between composition and site-specific evolutionary rate: Implications for phylogenetic inference. – Molecular Biology and Evolution **23**: 352–364.

Hennig, W. (1953): Kritische Bermerkungen zum phylogenetischen System der Insekten. – Beiträge zur Entomologie, Sonderheft 3, 85 pp.

Jermin, L. S., S. Y. W. Ho, F. Ababneh, J. Robinson & A. W. D. Larkum (2004): The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. – Systematic Biology **53**: 638–643.

Kjer, K. M. (2004): Aligned 18S and Insect Phylogeny. – Systematic Biology **53**: 506–514.

Kjer, K. M., F. L. Carle, J. Litman & J. Ware (2006): A molecular phylogeny of Hexapoda. – Arthropod Systematics & Phylogeny **64**: 3–44.

Kristensen, N. P. (1981): Phylogeny of insect orders. – Annual Review of Entomology **26**: 135–157.

Kristensen, N. P. (1997): The ground plan and basal diversification of the hexapods. – In: Fortey, R. A. & R. H. Thomas (eds): Arthropod relationships, systematic association, Ser 55. – Chapman and Hall, London: 281–293.

Koenemann, S., R. A. Jenner, M. Hoenemann, T. Stemme & B. M. von Reumont (2010): Arthropod phylogeny revisited, with a focus on crustacean relationships. – Arthropod Structure & Development **39**: 88–110.

Kukalová-Peck, J. (1987): New Carboniferous Diplura, Monura and Thysanura, the hexapod groundplan, and the role of thoracic side lobes in the origin of wings (Insecta). – Canadian Journal of Zoology **65**: 2327–2345.

Luan, Y., Y. Zhang, Q. Yue, J. Pang, R. Xie & W. Yin (2003): Ribosomal DNA gene and phylogenetic relationships of Diplura and lower Hexapods. – Science in China (Series C) **46**: 67–76.

Luan, Y., Y. Yao, R. Xie, Y. Yang, Y. Zhang & W. Yin (2004): Analysis of 18S rRNA gene of *Octostigma sinensis* (Projapygoidea:Octostigmatidae) supports the monophyly of Diplura. – Pedobiologia **48**: 453–459.

Luan, Y., J. M. Mallatt, R. Xie, Y. Yang & W. Yin (2005): The Phylogenetic Positions of Three Basal-Hexapod Groups (Protura, Diplura, and Collembola) Based on Ribosomal RNA Gene Sequences. – Molecular Biology and Evolution **22**:1579–1592.

Machida, R. (2006): Evidence from embryology for reconstructing the relationships of hexapod basal clades. – Arthropod Systematics & Phylogeny **64**: 95–104.

Mallatt, J. M. & G. Giribet (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**: 772–794.

Mallatt, J. M., C. W. Craig & M. J. Yoder (2010): Nearly complete rRNA genes assembled from across the metazoan animals: Effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction. – Molecular Phylogenetics and Evolution **55**: 1–17.

Meusemann, K., B. M. von Reumont, S. Simon, F. Roeding, S. Strauss, P. Kück, I. Ebersberger, M. Walzl, G. Pass, S. Breuers, V. Achter, A. von Haeseler, T. Burmester, H. Hadrys, J. W. Wägele & B. Misof (2010): A Phylogenomic Approach to Resolve the Arthropod Tree of Life. – Molecular Biology and Evolution **27**: 2451–2464.

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

Misof, B. & K. Misof (2009): A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. – Systematic Biology **58**: syp006.

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

Schmidt, H. A., K. Strimmer, M. Vingron & A. von Haeseler (2002): TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. – Bioinformatics **18**: 502–504.

Shao, H., Y. Zhang, R. Xie & W. Yin (1999): Mitochondrial cytochrome b sequences variation of Protura and molecular systematics of Apterygota. – Chinese Science Bulletin **44**: 2031–2036.

Shao, H., Y. Zhang, X. Ke, Q. Yue & W. Yin. (2000): Sequences of mitochondrial DNA cytochrome oxidase II in *Cryptopygus nanjiensis* and phylogeny of Apterygota. – Science in China (Series C) **43**: 589–596.

Silvestri, F. (1907): Descrizione di un nuovo genere di Insetti Apterigoti, rappresentante di un nuovo ordine. – Bollettino del Laboratorio di Zoologia generale e agraria della Facoltà di Agraria di Portici **1**: 296–311.

Szeptycki, A. (2007): Catalogue of the World Protura. – Acta Zoologica Cracoviensia **50**B: 1–210.

Szucsich, N. U. & G. Pass (2008): Incongruent phylogenetic hypotheses and character conflicts in morphology: The root and early branches of the hexapodan tree. – Mitteilungen der Deutschen Gesellschaft für Allgemeine und Angewandte Entomologie **16**: 415–430.

von Reumont, B. M., K. Meusemann, N. U. Szucsich, E. Dell'Ampio, V. Gowri-Shankar, D. Bartel, S. Simon, H. O. Letsch, R. R. Stocsits, Y. Luan, J. W. Wägele, G. Pass, H. Hadrys & B. Misof (2009): Can comprehensive background knowledge be incorporated into substitution models to improve phylogenetic analyses? A case study on major arthropod relationships. – BMC Evolutionary Biology **9**: 119.

Wheeler, W. C. (1996): Optimization Alignment: the end of multiple sequence alignment in phylogenetics? – Cladistics **12**: 1–9.

Wheeler, W. C. (1998): Sampling, groundplans, total evidence and the systematics of arthropods. - In: Fortey, R. A. & R. H. Thomas (eds.): Arthropod Relationships. - Chapman & Hall, London: 87-96.

Wheeler, W. C., M. Whiting, Q. D. Wheeler & J. M. Carpenter (2001): The Phylogeny of the Extant Hexapod Orders. - Cladistics **17**: 113–169.

Wheeler, W. C., D. Gladstein & J. DeLaet (2002): POY version 3.0. Program and documentation available at ftp.amnh.org/ pub/molecular/poy/.

Xie, Q., X. Tian, Y. Qin & W. Bu (2009): Phylogenetic comparison of local length plasticity of the small subunit of nuclear rDNAs among all Hexapoda orders and the impact of hyper-length-variation on alignment. – Molecular Phylogenetics and Evolution **50**: 310–316.

Xue, L. & W. Yin (1990): The ultrastructure of spermatozoa and spermatogenesis in *Fujientomon dicestum* (Insecta, Protura), with special reference to its significance on phylogeny. – Contributions of the Shanghai Institute of Entomology **8**: 125–130.

Yin, W., L. Xue & B. Tang (1990): On systematic positions of *Hesperentomon* and *Fujientomon* in view of the fine structure of spermatozoa and pseudoculi (Protura). – Contributions from Shanghai Institute of Entomology **9**: 121–125.

Yin, W. & L. Xue (1993): Comparative spermatology of Protura and its significance on proturan systematics. – Scientia Sinica (Series B) **36**: 575–586.

Yin, W. (1996): New considerations on systematics of Protura. In: 'Proceedings of XX International Congress of Entomology'. 25–31 August, 1996. Firenze, Italy: 1–60.

Zhang, Y., Y. Zhang, Y. Luan, Y. Chen & W. Yin (2001): Phylogeny of higher taxa of hexapoda according to 12sRNA sequences. – Chinese Science Bulletin **46**: 840–842.

8. Addendum

After the time (07.08.11) in which this manuscript has been accepted, the following paper containing the complete mitochondrial genome of *Sinentomon erythranum* was published (GenBank RefSeq number, NC_015982).

Chen, W. J., Y. Bu, A. Carapelli, R. Dallai, S. Li, W. Yin & Y. Luan (2011): The mitochondrial genome of *Sinentomon erythranum* (Arthropoda: Hexapoda: Protura): an example of highly divergent evolution. - BMC Evolutionary Biology **11**: 246.

Accepted 07 August 2011