Bonner zoologische Beiträge	Band 55 (2006)	Heft 3/4	Seiten 191–202	Bonn, November 2007

Utility of H3-Genesequences for phylogenetic reconstruction – a case study of heterobranch Gastropoda –*

Angela DINAPOLI[1)], Ceyhun TAMER[1)], Susanne FRANSSEN[1)],
Lisha NADUVILEZHATH[1)] & Annette KLUSSMANN-KOLB[1)]
[1)]Department of Ecology, Evolution and Diversity – Phylogeny and Systematics, J. W. Goethe-University, Frankfurt am Main, Germany

*Paper presented to the 2nd International Workshop on Opisthobranchia, ZFMK, Bonn, Germany, September 20th to 22nd, 2006

Abstract. In the present study we assessed the utility of H3-Genesequences for phylogenetic reconstruction of the Heterobranchia (Mollusca, Gastropoda). Therefore histone H3 data were collected for 49 species including most of the major groups. The sequence alignment provided a total of 246 sites of which 105 were variable and 96 parsimony informative. Twenty-four (of 82) first base positions were variable as were 78 of the third base positions but only 3 of the second base positions.

H3 analyses showed a high codon usage bias. The consistency index was low (0,210) and a substitution saturation was observed in the 3rd codon position. The alignment with the translation of the H3 DNA sequences to amino-acid sequences had no sites that were parsimony-informative within the Heterobranchia.

Phylogenetic trees were reconstructed using maximum parsimony, maximum likelihood and Bayesian methodologies. *Nodilittorina unifasciata* was used as outgroup.

The resolution of the deeper nodes was limited in this molecular study. The data themselves were not sufficient to clarify phylogenetic relationships within Heterobranchia. Neither the monophyly of the Euthyneura nor a step-by-step evolution by the "basal" groups was supported. A conclusion about the monophyly of Opisthobranchia and Pulmonata could not be extracted from our data because we did not have any resolution at this point.

We believe histone H3 alone provides no new marker for studying deep molecular evolution of the Heterobranchia due to the high grade of conservation and the low phylogenetic signal.

Surprisingly there was a good resolution on the genera level. Analyses conducted with maximum parsimony and Bayesian inference (using all data) recovered all (or nearly all) genera mostly with statistically significantly supported nodes. Further studies focusing on the possible utility of histone H3 for the resolution of recent splits will be necessary.

Keywords. Heterobranchia, Opisthobranchia, histone H3, molecular phylogeny.

1. INTRODUCTION

Many questions regarding gastropod phylogeny have not yet been answered such as the molecular confirmation of the Heterobranchia concept based on morphological studies from HASZPRUNAR (1985, 1988). This taxon contains the Pentaganglionata HASZPRUNAR, 1985 also known as Euthyneura SPENGEL, 1881 (with the Opisthobranchia and Pulmonata) and several mostly little known "basal" groups (e.g. Valvatoidea, Omalogyroidea, Architectonicoidea, Rissoelloidea and Pyramidelloidea) which present a step-by-step evolution towards the euthyneuran level of organisation (HASZPRUNAR 1988). The hyperstrophy of the protoconch is the most important autapomorphous character of the Heterobranchia. The Euthyneura are characterised by the presence of two additional (so-called parietal) ganglia. However, the monophyly of the Euthyneura has not been clarified by molecular studies, yet. In some studies they are recovered monophyletic (COLGAN et al. 2000, 2003; KNUDSEN et al. 2006) in others not (THOLLESSON 1999). The Pulmonata and Opisthobranchia can be separated by characters respective of the nervous system (presence of a procerebrum and cerebral bodies in pulmonates and presence of a rhinophoral nerve in Opisthobranchia and Pyramidelloidea). However the molecular confirmation regarding the monophyly of the Opisthobranchia (VONNEMANN et al. 2005; GRANDE et al. 2004a) and the Pulmonata (TILLIER et al. 1996, DAYRAT et al. 2001) is still a matter of debate. There is no comprehensive investigation concerning the "basal" groups. Only a few representative taxa (e.g. Valvatoidea – *Cornirostra pellucida*, Architectonicoidea – *Philippea lutea*, Pyramidelloidea – *Pyramidella dolabrata*) have been included in current molecular studies (COLGAN et al. 2000; GRANDE et al. 2004a, 2004b).

In recent years molecular systematic analyses in gastropods have utilised a variety of genetic markers, e.g. nuclear 28S ribosomal RNA and/or 18S ribosomal RNA or mitochondrial 16S ribosomal RNA and/or cytochrome oxidase subunit I (Tillier et al. 1994, 1996; Dayrat et al. 2001; Vonnemann et al. 2005; Thollesson et al. 1999; Remigio & Hebert 2003). Nevertheless, new genetic markers are needed for the resolution of certain phylogenetic relationships (especially regarding deeper nodes). Partial fragments of the gene coding for the extremely conservative H3 protein (Maxson et al. 1983) were first used to clarify arthropod molecular evolution (Colgan et al. 1998) and later polychaete (Brown et al. 1999), gastropod (Colgan et al. 2000, 2003), polyplacophoran (Okusu et al. 2003), cephalopod (Lindgren et al. 2004) and hexapod (Kjer et al. 2006) phylogeny. All studies used a combined dataset in their approaches. In their study of gastropod phylogeny, Colgan et al. (2000) did not find a monophyletic Heterobranchia while within the Euthyneura, the Opisthobranchia are paraphyletic with respect to the pulmonates. Very similar phylogenetic relationships were shown in Colgan et al. (2003). The Heterobranchia as well as the Opisthobranchia and Pulmonata are rarely recovered as monophyletic in these studies.

In the present study we wanted to test the utility of H3 gene sequences for phylogenetic reconstruction within the Heterobranchia (focusing primarily on the Opisthobranchia). We were especially interested in testing whether H3 is suitable to resolve deeper nodes within heterobranch phylogeny. Therefore, partial histone H3 data were collected for 49 species including most of the major groups (Euthyneura with Opisthobranchia and Pulmonata and "basal" groups with Valvatoidea, Architectonicoidea, Omalogyroidea, Rissoelloidea and Pyramidelloidea).

2. MATERIALS AND METHODS

2.1. Specimens and DNA extraction

The studied taxa and the accession numbers are listed in Table 1. Twenty of the 49 sequences are taken from GenBank. Opisthobranchia are represented by 26 species (including 11 suborders). *Nodilittorina unifasciata* (Caenogastropoda Cox, 1960) was used as an outgroup.

DNA was extracted from ethanol-preserved individuals using the DNeasy Tissue Kit from Qiagen (Hilden, Germany).

Table 1 . Taxonomic positions and collecting locations of the sampled taxa. Accession numbers of sequences included in the analyses (ZSM = Zoologische Staatssammlung); published sequences taken from GenBank are marked with an asterisk.

Major Taxon	Species	Locality	GenBank Accession Number
Caenogastropoda			
Littorinoidea			
Littorinidae	*Nodilittorina unifasciata* (Gray, 1826)	Genbank	AF033705*
Conidae	*Conus miles* Linnaeus, 1758	Genbank	AF033684*
Campanilidae	*Campanile symbolicum* Iredale, 1917	Genbank	AF033683*
Opisthobranchia			
Nudibranchia			
Tethydidae	*Tethys fimbria* Linné, 1767	Blanes, Spain	EF133468
Discodorididae	*Discodoris atromaculata* (Bergh, 1880)	Genbank	DQ280013*
Arminidae	*Armina neapolitana* (Delle Chiaje, 1824)	Banyuls-sur-Mer, France	EF133469
Pleurobranchoidea			
Pleurobranchidae	*Pleurobranchaea meckeli* Leue, 1813	Blanes, Spain	EF133470
Tylodinoidea			
Umbraculidae	*Umbraculum umbraculum* (Lightfoot, 1786)	Atlantic Ocean, Meteor Bank	EF133471
Cephalaspidea			
Scaphandridae	*Scaphander lignarius* (Linné, 1758)	Blanes, Spain	EF133472
Philinidae	*Philine aperta* (Linnaeus, 1767)	Genbank	DQ093508*
Gastropteridae	*Gastropteron meckeli* Kosse, 1813	Blanes, Spain	EF133473
Anaspidea			
Akeridae	*Akera bullata* Müller, 1776	Kattegat, Denmark	EF133474
Aplysiidae	*Aplysia californica* Cooper, 1863	Miami, USA	EF133475
Aplysiidae	*Aplysia cf. juliana* Quoy & Gaimard, 1832	Genbank	AF033675*
Aplysiidae	*Bursatella leachii* de Blainville, 1817	Dingo Beach, Australia	EF133476
Thecosomata			
Cavoliniidae	*Clio pyramidata* Linné, 1767	Canary Islands; Spain	EF133477
Creseidae	*Creseis* sp.	Genbank	DQ280012*
Gymnosomata			
Pneumodermatidae	*Pneumoderma cf. atlantica* (Oken, 1815)	USA, Atlantic	EF133478

Major Taxon	Species	Locality	GenBank Accession Number
Sacoglossa			
Placobranchidae	*Elysia timida* (Risso, 1818)	Blanes, Spain	EF133479
Placobranchidae	*Elysia pusilla* (Bergh, 1872)	Genbank	DQ534792*
Placobranchidae	*Elysia crispata* Morch, 1863	Genbank	DQ534790*
Placobranchidae	*Elysia viridis* (Montagu, 1804)	Genbank	DQ534790*
Cylindrobullidae	*Cylindrobulla beauii* Fischer, 1857	Florida, USA	EF133480
Acochlidia			
Hedylopsidae	*Hedylopsis spiculifera* (Kowalewsky, 1901)	Rovinj, Croatia	EF133481
Microhedylidae	*Unela glandulifera* (Kowalewsky, 1901)	Rovinj, Croatia	EF133482
Architectibranchia			
Hydatinidae	*Micromelo undatus* (Bruguiere, 1792)	Genbank	DQ093513*
Acteonoidea			
Acteonidae	*Pupa solidula* (Linné, 1758)	Dingo Beach, Australia	EF133483
Acteonidae	*Rictaxis punctocaelatus* (Carpenter, 1864)	Cayucos, California, USA	EF133484
Bullinidae	*Bullina lineata* (Gray, 1825)	Genbank	AF033680*
Pulmonata			
Systellommatophora			
Onchidiidae	*Onchidium sp.*	Genbank	AF033706*
Onchidiidae	*Onchidella floridana* (Dall, 1885)	Florida, USA	EF133485
Onchidiidae	*Onchidella* sp.	Genbank	DQ093511*
Stylommatophora			
Charopidae	*Hedleyoconcha delta* (Pfeiffer, 1857)	Genbank	AF033693*
Siphonariidae	*Siphonaria serrata* (Fischer, 1807)	South Africa	EF133486
Siphonariidae	*Siphonaria concinna* Sowerby,1824	South Africa	EF133487
Siphonariidae	*Siphonaria zelandica* (Quoy & Gaimard, 1832)	Genbank	AF033713*
Amphibolidae	*Salinator solida* (Schacko, 1878)	Genbank	AF033712*
Eupulmonata			
Ellobiidae	*Ophicardelus ornatus* (Ferussac, 1821)	Genbank	AF033707*
"basal" Heterobranchia/Triganglionata			
Pyramidelloidea			
Pyramidellidae	*Turbonilla lactea* (Linné, 1758)	Roscoff, France	EF133488
Pyramidellidae	*Turbonilla sp.*	Wellington, New Zealand	EF133489
Architectonicoidea			
Architectonicidae	*Heliacus variegatus* (Gmelin, 1791)	Tropical aquarium (ZSM 20012193)	EF133490
Architectonicidae	*Philippea lutea* (Lamarck, 1822)	Genbank	AF033708*
Valvatoidea			
Cornirostridae	*Cornirostra pellucida* (Laseron, 1954)	Genbank	AF033685*
Orbitestellidae	*Orbitestella vera* Powell, 1940	Wellington, New Zealand	EF561623
Orbitestellidae	*Orbitestella* sp.	Leigh, New Zealand	EF561624
Omalogyroidea			
Omalogyridae	*Omalogyra burdwoodiana* Strebel, 1908	Antarctic (ZSM Mol-20021228)	EF133491
Rissoelloidea			
Rissoelidae	*Rissoella elongatospira* Ponder, 1966	Wellington, New Zealand	EF561622
Rissoelidae	*Rissoella micra* Finlay, 1924	Wellington, New Zealand	EF561620
Rissoelidae	*Rissoella cystophora* Finlay, 1924	Wellington, New Zealand	EF561621

2.2. DNA amplification and sequencing

The following degenerated primers were used: H3-F: 5'-ATG GCT CGT ACC AAG CAG AC(ACG) GC-3' and H3-R: 5'-ATA TCC TT(AG) GGC AT(AG) AT(AG) GTG AC-3' (Colgan et al. 1998) and produced a 246 bp product. The PCR profile was as follows: 95 °C for 5 min, followed by 35 cycles of 30 s at 95 °C, 25 s at 52 °C, 45 s at 72 °C and a final extension at 72 °C for 5 min and Taq Polymerase, recombinant from Invitrogen (Karlsruhe, Germany) was used. All products were purified using the QIAquick Gel Extraction Kit from Qiagen (Hilden, Germany) and sequenced in both directions with a CEQ 2000 Beckmann Coulter using the CEQ DTCS Quick Start Kit (Krefeld, Germany). First we sequenced only one fragment per specimen. Moreover, to avoid mistakes according to the conservative character of the H3 gene, we sequenced a random sample of 5 species a second time whereby no varieties could be detected.

2.3. Sequence alignment

Sequences were aligned manually using the software package BioEdit version 7.0.5 (Hall 1999). The H3 DNA sequences were translated into the amino acid sequences in GeneDoc version 2.6.002 (Nicholas & Nicholas 1997). The alignment is available from the authors upon request.

2.4. Statistical tests

Codon usage statistics were calculated using GCUA version 1.2 (MCINERNEY 1997). The purpose of this function is to calculate the Number (N) of times a particular codon is observed in an alignment and also to calculate the Relative Synonymous Codon Usage (RSCU) values for the dataset. RSCU values define the number of times a particular codon is observed relative to the number of times that the codon would be observed in the absence of any codon usage bias. Without any codon usage bias, the RSCU value would be 1.00. A codon that is used less frequently than expected will have a value of less than 1.00 and a codon that is used more frequently than expected will have a volume of more than 1.00 (MCINERNEY 1997).

The degree of bias ($\sum\chi^2/n$) which is the sum of the χ^2 values for the individual amino acids divided by the total number of inferred residues (n) for the combination of data from all species (SHIELDS et al. 1988) was determined.

The substitution saturation was calculated for all 3 codon positions using the method developed by XIA et al. (2003) implemented in the software package DAMBE version 4.2.13 (XIA & XIE 2001).

2.5. Phylogenetic reconstruction

Appropriate models for the analyses were selected after running Modeltest version 3.4 (POSADA & CRANDALL 1998) and using the Akaike information criterion (AIC) (see tab. 2).

The following analyses were conducted using PAUP* version 4.0 b10 (SWOFFORD, 2002) (settings: heuristic search strategy; tbr; gaps were treated as fifth bases): a) Maximum parsimony for all data and b) Maximum likelihood for all data.

Bootstrapping (FELSENSTEIN 1985) was performed for maximum parsimony with 1000 replicates and for maximum likelihood with 100 replicates.

The following analyses were conducted using MrBayes version 3.1.2 (RONQUIST & HUELSENBECK 2003): Bayesian inference: a) all data (with one model for all three codon positions), b) all data (with codon specific models) and c) only codon position one and two (third codon position excluded; with one model for codon position one and two).

For Bayesian inference a Metropolis Chain Monte Carlo analysis with four chains and 1 000 000 generations was performed with the first 1000 trees ignored as burn-in.

3. RESULTS

3.1. Statistical tests

The sequences provided a total of 246 sites of which 105 were variable and 96 parsimony informative. Twenty-four (of 82) first base positions were variable as were 78 of the third base positions but only 3 of the second base positions. Insertion/deletion events (indels) were not observed in any of the groups. The amino acid alignment had no sites that were parsimony-informative within the Heterobranchia.

H3 analyses showed a high codon usage bias (Tab. 3). The bias was principally against the use of A and U in the third codon position. χ^2 tests were performed for all amino acids and revealed that the null hypothesis which is the expected equal usage of the codons can be rejected for all amino acids with a significance level of 0,001 excepting histidine ($p<0,05$). For aspartic acid the null hypothesis can not be rejected ($p=0,21$). The degree of bias ($\sum\chi^2/n$) showed a high value of 0,617.

Table 2. Information on used models.

Codon-Position	Model	Gamma distribution shape parameter	Proportion of invariable sites
1st Position	GTR+I	α=equal	Pinvar=0.5692
2nd Position	TVMef+I	α=equal	Pinvar=0.8647
3rd Position	TVM+G	α=0.8835	Pinvar=equal
1st and 2nd Position	GTR+I+G	α=0.9167	Pinvar=0.6909
1st, 2nd and 3rd Position	GTR+I+G	α=1.0265	Pinvar=0.5408

Table 3. Codon Usage Bias. N = number of times a particular codon is observed in a dataset (alignment). RSCU values = number of times a particular codon is observed, relative to the number of times that the codon would be observed in the absence of any codon usage bias. Amino acids (AA) are indicated by the three letter abbreviations.

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	6	(0.08)	Ser	UCU	45	(1.38)
Phe	UUC	142	(1.92)	Ser	UCC	34	(1.04)
Leu	UUA	3	(0.05)	Ser	UCA	13	(0.40)
Leu	UUG	43	(0.66)	Ser	UCG	8	(0.24)
Tyr	UAU	2	(0.04)	Cys	UGU	0	(0.00)
Tyr	UAC	96	(1.96)	Cys	UGC	0	(0.00)
Ter	UAA	0	(0.00)	Trp	UGG	0	(1.00)
Ter	UAG	0	(0.00)	Pro	CCU	83	(1.36)
Ter	UGA	0	(0.00)	Pro	CCC	107	(1.75)
Leu	CUU	48	(0.74)	Pro	CCA	42	(0.69)
Leu	CUC	81	(1.24)	Pro	CCG	13	(0.21)
Leu	CUA	2	(0.03)	Arg	CGU	233	(2.59)
Leu	CUG	214	(3.28)	Arg	CGC	127	(1.41)
His	CAU	31	(1.27)	Arg	CGA	18	(0.20)
His	CAC	18	(0.73)	Arg	CGG	9	(0.10)
Gln	CAA	41	(0.28)	Thr	ACU	53	(0.87)
Gln	CAG	252	(1.72)	Thr	ACC	146	(2.38)
Ile	AUU	21	(0.43)	Thr	ACA	45	(0.73)
Ile	AUC	126	(2.57)	Thr	ACG	1	(0.02)
Ile	AUA	0	(0.00)	Ser	AGU	8	(0.24)
Asn	AAU	0	(0.00)	Ser	AGC	88	(2.69)
Asn	AAC	0	(0.00)	Arg	AGA	74	(0.82)
Lys	AAA	135	(0.61)	Arg	AGG	78	(0.87)
Lys	AAG	306	(1.39)	Ala	GCU	201	(1.49)
Val	GUU	9	(0.18)	Ala	GCC	279	(2.07)
Val	GUC	89	(1.82)	Ala	GCA	47	(0.35)
Val	GUA	4	(0.08)	Ala	GCG	12	(0.09)
Val	GUG	94	(1.92)	Gly	GGU	32	(0.87)
Asp	GAU	38	(0.78)	Gly	GGC	38	(1.03)
Asp	GAC	60	(1.22)	Gly	GGA	72	(1.96)
Glu	GAA	69	(0.70)	Gly	GGG	5	(0.14)
Glu	GAG	128	(1.30)	Met	AUG	49	(1.00)

The consistency index in maximum parsimony analyses was low (0,210) as well as the retention index (0.444). Using the method developed by Xia et al. (2003) a substitution saturation was observed in the third codon position (I_{ss} 0,543 > $I_{ss\cdot c}$ 0,298). To support this observation the entropy for each position in the alignment was calculated using BioEdit Version 7.0.5.2. Codon position one showed an average entropy value of 0.09 whereas the value for the second codon position was 0,009 and 0,69 for the third codon position. If the nucleotides occur more or less equally at a certain position within the alignment the entropy is highest with the value of 1,36.

3.2. Phylogenetic analyses

The maximum parsimony 50 % majority-rule consensus tree (Fig. 1) showed all genera (*Aplysia, Elysia, Onchidella, Siphonaria, Orbitestella, Turbonilla* and *Rissoella*) recovered as monophyletic. However, some of the bootstrap supports were low and there was no bootstrap support for a monophyletic *Rissoella*. Beyond the genera level, only Architectonicoidea were detected as monophyletic. All other nodes lacked support.

The genera *Orbitestella*, *Turbonilla*, *Onchidella* and *Aplysia* and again, the Architectonoicoidea were found to be monophyletic in the maximum likelihood analyses. The

Fig. 1. 50 % majority-rule consensus tree of maximum parsimony of 14 most parsimonious trees based on histone H3 data set (nucleotides), number of parsimony informative characters = 96, consistency index (CI) = 0,210.

Fig. 2. 50 % majority rule consensus Bayesian inference cladogram for the histone H3 dataset (based on nucleotides); Bayesian posterior probabilities provided at the branches.

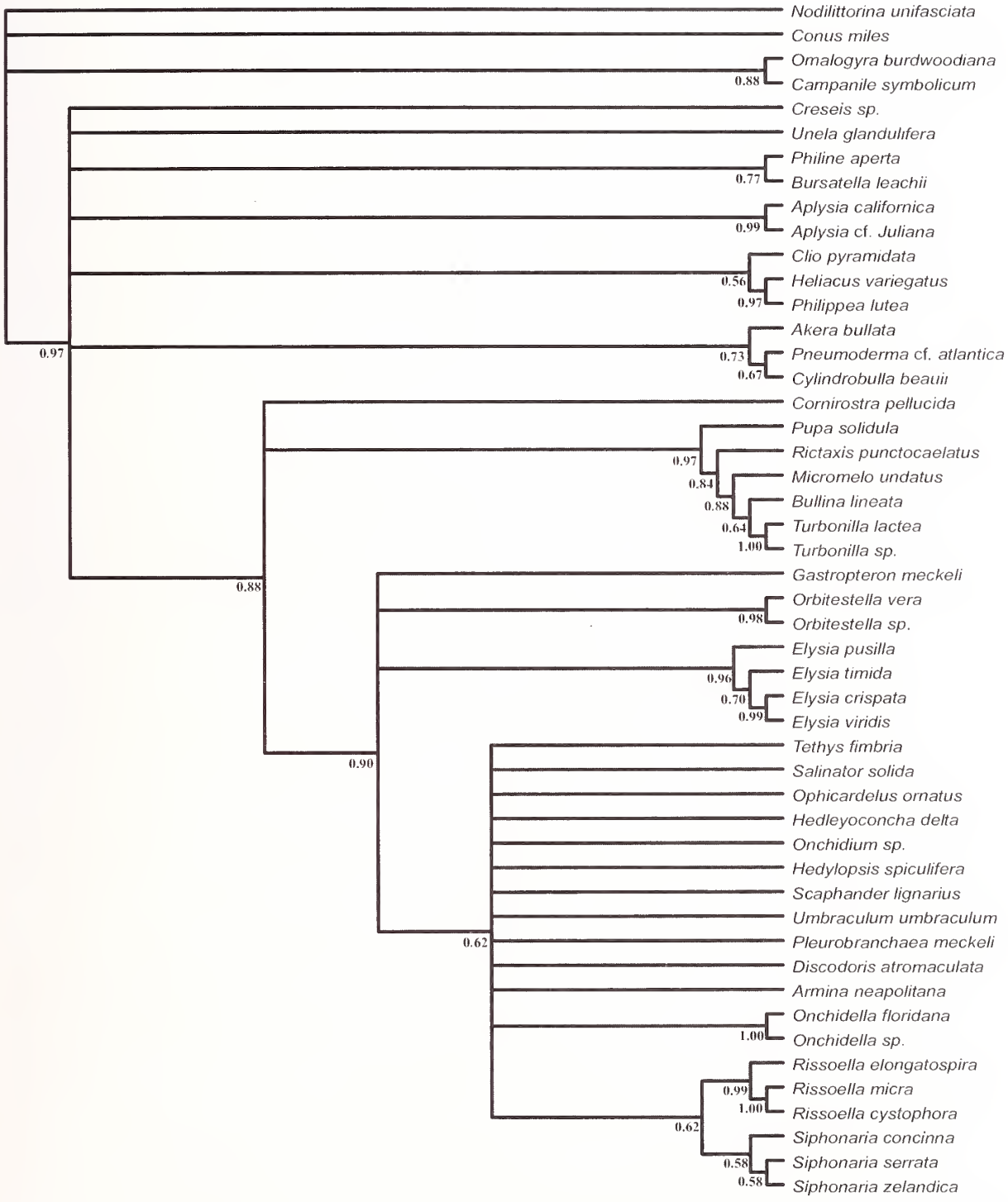

Fig. 3. 50 % majority rule consensus Bayesian inference phylogram for the histone H3 dataset (based on nucleotides, 3rd codon position excluded); Bayesian posterior probabilities provided at the branches.

basis of the 50 % majority-rule bootstrap tree (tree not shown) resembled a comb. There was no resolution of the deep nodes.

The 50 % majority-rule consensus Bayesian inference cladogram (with one model for all three codon positions) (Fig. 2) also recovered all genera as monophyletic. Only the Bayesian posterior probability for *Siphonaria* was low (0.58) because only values above 0.95 are statistically significant. Beside the genera level Architectonicoidea were monophyletic and the Caenogastropoda formed a clade together with *Omalogyra*. All other nodes had no statistically significant support.

The 50 % majority-rule consensus Bayesian inference cladogram (with codon specific models) (tree not shown) was quite similar to Figure. 2. All genera except *Rissoella* and Architectonicoidea were found to be monophyletic. The remaining nodes were supported by Bayesian posterior probabilities below 0.95.

The 50 % majority rule consensus Bayesian inference phylogram (with 3rd codon position excluded and with one model for codon position one and two) (Fig. 3) recovered only the genera *Rissoella* and *Turbonilla* as monophyletic while the Caenogastropoda together with *Omalogyra* were grouped separately from the rest of the taxa. All other nodes had no statistically significant support.

4. DISCUSSION

Molecular investigations of deep-level relationships within the Gastropoda have been made difficult due to a lack of slowly evolving genes. Hence, a number of different markers have been utilized to solve this problem. Analyses of nuclear genes like the 28S ribosomal RNA and/or the 18S ribosomal RNA have provided a number of important insights into gastropod relationships at several levels (Tillier et al. 1994, 1996; Dayrat et al. 2001; Vonnemann et al. 2005). The same applies to mitochondrial genes like the 16s ribosomal RNA (Thollesson et al. 1999) or the cytochrome oxidase subunit 1 (Remigio & Hebert 2003). Colgan et al. (2000, 2003) used the histone H3 protein in combination with other genes to clarify gastropod molecular evolution. However, many aspects of gastropod phylogeny remain unclear such as the molecular confirmation of the Heterobranchia concept based on morphological studies from Haszprunar (1985, 1988). At the moment there is no comprehensive molecular study of heterobranch phylogeny especially one including the "basal" taxa (e. g. Pyramidelloidea, Architectonicoidea, Valvatoidea, Omalogyroidea and Rissoelloidea).

In this study we wanted to present a primary molecular insight into heterobranch phylogeny while simultaneously testing the utility of the gene coding for the highly conserved protein histone H3 for resolving the deeper nodes within this taxon.

Unfortunately, the present study did not provide a robust phylogenetic hypothesis for the relationships among different lineages of Heterobranchia based on 113-Genesequences. Neither the monophyly of the Euthyneura nor a step-by-step evolution by the "basal" groups was supported. A conclusion about the monophyly of Opisthobranchia and Pulmonata could not be extracted from our data because we did not have any resolution at this point.

The first to investigate the value of histone H3 were Colgan et al. (1998). They wanted to combine small nuclear ribonucleic acid U2 data and histone H3 to investigate arthropod molecular evolution. However, partitioned data for H3 and U2 were incongruent according to Incongruence Length Difference tests. Using H3 data only, anomalous nodes appeared in their phylogenies while some possessed decay indices of 1. Therefore, their data were not sufficient to clarify relationships within major arthropod groups.

Brown et al. (1999) investigated the DNA sequence data of 34 Polychaeta species for partial histone H3, U2 snRNA and two segments of 28S rDNA (D1 and D9-10 expansion regions). When using H3 only, Brown et al. (1999) found a lack of concordance with morphological results and argued that the inclusion of all H3 data is inappropriate for the phylogenetic levels under investigations.

Colgan et al. (2000) and later Colgan et al. (2003) used partial histone H3 (327bp) for the investigation of gastropod phylogeny. In Colgan et al. (2000), where the authors used 36 sequences of histone H3 only, using the chiton *Ischnochiton australis* as an outgroup, no clades were retained in the bootstrap analyses. H3 with the third codon position excluded, retained only the higher Vetigastropoda (bootstrap support = 68 %).

In Colgan et al. (2003) in which H3 alone was used to recover phylogenetic relationships within Gastropoda, only the clade of the Patellogastropoda with a support of 52 % was recovered. When the third codon position was excluded, none of the expected groups were recovered.

Okuso et al. (2003) were the first to apply DNA sequence data to reconstruct the phylogeny of the molluscan class Polyplacophora. Their use of 59 sequences of histone H3 resolved deeper nodes than the mitochondrial genes did while the strict consensus tree nested the two Gastropo-

da *Viviparus georgianus* and *Siphonaria pectinata* within Polyplacophora.

A combined approach to the phylogeny of Cephalopoda (Mollusca) using 18S rRNA, 28S rRNA, histone H3 and CO1 underscored the aim of the study presented by Lindgren et al. (2004). The strict consensus tree for the overall optimal parameter set for 66 sequences of histone H3 alone did not show monophyly for any classes investigated.

Kjer et al. (2006) investigated the molecular phylogeny of Hexapoda (supermatrix approach with 137 taxa; 375bp). They only recovered one ordinal level node when using only H3.

According to our phylogenetic results and the results in the papers listed above histone H3 alone provides no new marker for studying deep molecular evolution of the Heterobranchia due to the high grade of conservation and the low phylogenetic signal for deeper nodes.

There were several indices defined by the results of our statistical tests supporting this assumption. We observed a high codon usage bias (see tab. 2) in our alignment which was also indicated by an increasing frequency of C- and G-ending codons and fewer A- and U-ending codons (Shields et al. 1988). High C+G content at silent sites reflects the effect of selection (Shields et al. 1988) while selective constraints against certain codons might reduce the amount of phylogenetic noise caused by synonymous substitution at either first or third codon positions (Brown et al. 1999). However, our data suggest that a bias in codon usage will not necessarily be indicative of the phylogenetic utility of a sequence. Despite a high codon usage bias our computed phylogenetic trees showed a poor resolution of the deeper nodes. An explanation for this could be that the pressure to obtain the favoured codon had partially obscured the phylogenetic signal. Colgan et al. (1998, 2000) made similar observations and concluded that apparent, high codon-usage bias as found for the H3 data does not necessarily result in high phylogenetic consistency for DNA sequences. In the studies presented by Brown et al. (1999) a lack of agreement of the H3 analyses with morphology occurs despite very high codon usage bias. They concluded that whilst selective constraints may have reduced the absolute rate of synonymous substitutions, the pressure in favor of (homoplastic) restitution of the favoured codon has at least a partially obscured phylogenetic signal. Hence, codon usage bias does not necessarily mean that a gene sequence will be phylogenetically useful.

The degree of bias ($\sum\chi^2/n$) showed a high value of 0,617. It was similar to the values observed in gastropods (0.60) (Colgan et al. 2000) and polychaetes (0.665) (Brown et al. 1999) and higher than the values of *Drosophila melanogaster* (Fitch & Strausbaugh 1993) and arthropods (0.37) (Colgan et al. 1998).

Another indication suggesting the problems of H3 as a marker for studying deep molecular evolution was the high grade of conservation indicated by the lack of parsimony-informative sites in the amino acid alignment.

Additional important evidence was the observed substitution saturation at the third codon position. A saturation is caused by multiple-hits which render homoplasious changes. Homoplasy on the basis of saturation in substitution is one of the major problems in molecular phylogenetics (Tillier et al. 1996). This problem generally becomes more relevant at progressively higher taxonomic levels (Boore & Brown 1998). If numerous substitutions occur at the same position, a hiding or completely erasing of the ancient phylogenetic signal could be the result (Lopez et al. 1999). In order to avoid a decrease of the phylogenetic information contained in the sequences, we excluded the third codon position in further analyses. However, this was the position with the most variable sites (78 of 82 positions) in our data set. With the exclusion there was no phylogenetic information left for a resolution of the deeper nodes. Trees which resemble combs at the base resulted (see fig. 3). The entropy was calculated for each position in the alignment to further assess the influence of the 3 codon positions. A high entropy value implies that the nucleotides occur almost equally at this position within the alignment. An almost equal distribution of the nucleotides at one position indicates a less selective constraint allowing a higher frequency of substitutions. The average values for each of the three codon positions indicate a less selective constraint for the third codon position hence supporting the previous result of this codon position being saturated.

It is questionable if given a larger data set, the noise will eventually succumb to the signal. There are few examples where the expansion of an ambiguous data set has resulted in a convincing phylogeny (Boore & Brown 1998).

Surprisingly, there was a good resolution on genera level. Analyses conducted with maximum parsimony and Bayesian inference (with all data) recovered all (or nearly all) genera mostly with statistically significant supported nodes. However, our findings should be consider preliminary and further studies including more genera are necessary to test the possible utility of histone H3 for the resolution of recent splits. In other studies (Colgan et al. 1998; Okuso et al. 2003; Lindgren et al. 2004) some (but not all) genera were found to be monophyletic but due to the question the authors intended to answer, their studies

in regards to taxon sampling lacked genera represented by more than one species.

In conclusion, still slowly evolving genes for the resolution of the deeper nodes in gastropod phylogeny are missing. To test the Heterobranchia concept as outlined by HASZPRUNAR (1988) other markers have to be found with sufficient variability but no substitution saturation.

Acknowledgements. We are grateful to Hermann Dreyer for providing the H3 primers and to the following colleagues and friends for providing samples: Gilianne and Jon Brodie (Townsville), Alan Hodgson (Grahamstown), Michael Schrödl (Munich), Enrico Schwabe (Munich), Verena Vonnemann (Bochum), Heike Wägele (Bonn), Tim Wollesen (Frankfurt), as well as Bruce Marshall (Te Papa Tongarewa, Wellington) for his support during a collecting trip to New Zealand. This project was funded by the German Science Foundation (KL 1303/4-1).

REFERENCES

BOORE, J. L. & BROWN, W. M. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Current Opinion in Genetics and Development **8(6)**: 668–674.

BROWN, S., ROUSE, G., HUTCHINGS, P. & COLGAN, D. 1999. Assessing the usefulness of histone H3, U2 snRNA and 28S rDNA in analyses of polychaete relationships. Australian Journal of Zoology **47**: 499–516.

COLGAN, D. J., MCLAUCHLAN, A., WILSON, G. D. F., LIVINGSTON, S. P., EDGECOMBE, G. D., MACARANAS, J., CASSIS, G. & GRAY, M. R. 1998. Histone H3 and U2 snRNA sequences and arthropod molecular evolution. Australian Journal of Zoology **46**: 419–437.

COLGAN, D. J., PONDER, W. F. & EGGLER, P. E. 2000. Gastropod evolutionary rates and phylogenetic relationships. Zoologica Scripta **29(1)**: 29–63.

COLGAN, D. J., PONDER, W. F., BEACHAM, E. & MACARANAS, J. M. 2003. Gastropod phylogeny based on six segments from four genes representing coding or non-coding and mitochondrial or nuclear DNA. Molluscan Research **23**: 123–148.

DAYRAT, B., TILLIER, A., LECOINTRE, G. & TILLIER, S. 2001. New Clades of Euthyneuran Gastropods (Mollusca) from 28S rRNA Sequences. Molecular Phylogenetics and Evolution **19(2)**: 225–235.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**: 783–791.

FITCH, D. H. A. & STRAUSBAUGH, L. D. 1993. Low Codon Bias and High Rates of Synonymous Substitution in *Drosophila hydei* and *D. melanogaster* Histone Genes. Molecular Biology and Evolution **10(2)**: 397–413.

GRANDE, C., TEMPLADO, J., CERVERA, J. L. & ZARDOYA, R. 2004a. Phylogenetic relationships among Opisthobranchia (Mollusca: Gastropoda) based on mitochondrial *cox 1*, *trn V*, and *rrnL* genes. Molecular Phylogenetics and Evolution **33**: 378–388.

GRANDE, C., TEMPLADO, J. CERVERA, J. L. & ZARDOYA, R. 2004b. Molecular Phylogeny of Euthyneura (Mollusca: Gastropoda). Molecular Biology and Evolution **21(2)**: 303–313.

HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**: 95–98.

HASZPRUNAR, G. 1985. The Heterobranchia – a new concept of the phylogeny of the higher Gastropoda. Zeitschrift für Zoologie, Systematik und Evolutionsforschung **23**: 15–37.

HASZPRUNAR, G. 1988. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. Journal of Molluscan Studies **54**: 367–441.

LOPEZ, P., FORTERRE, P. & PHILIPPE, H. 1999. The root of the tree of life in the light of the covarion model. Journal of Molecular Evolution **49**: 496–508.

KJER, K. M., CARLE, F. L., LITMAN, J. & WARE, J. 2006. A Molecular Phylogeny of Hexapoda. Arthropod Systematics & Phylogeny **64(1)**: 35–44.

KNUDSEN, B., KOHN, A. B., NAHIR, B., MCFADDEN, C. S. & MOROZ, L. L. 2006. Complete DNA sequence of the mitochondrial genome of the sea-slug, *Aplysia californica*: Conservation of the gene order in Euthyneura. Molecular Phylogenetics and Evolution **38**: 459–469.

LINDGREN, A. R., GIRIBET, G. & NISHIGUCHI, M. K. 2004. A combined approach to the phylogeny of Cephalopoda (Mollusca). Cladistics **20**: 454–486.

MAXSON, R., COHN, R. & KEDES, L. 1983. Expression and Organisation of Histone Genes. Annual Review of Genetics **17**: 239–277.

MCINERNEY, J. 1997. GCUA: General Codon Usage Analysis. Version 1.1. The National History Museum, London.

NICHOLAS, K. B. & NICHOLAS, H. B. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

OKUSU, A., SCHWABE, E., EERNISSE, D. & GIRIBET, G. 2003. Towards a phylogeny of chitons (Mollusca, Polyplacophora) based on combined analysis of five molecular loci. Organisms, Diversity & Evolution **3**: 281–302.

POSADA, D. & CRANDALL, K. A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**: 817–818.

REMIGIO, E. A. & HEBERT, P. D. N. 2003. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. Molecular Phylogenetics and Evolution **29**: 641–647.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**: 1572–1574.

SHIELDS, D. C., SHARP, P. M., HIGGINS, D. G. & WRIGHT, F. 1988. "Silent" Sites in Drosophila Genes are not neutral: Evidence of Selection among Synonymous Codons. Molecular Biology & Evolution **5(6)**: 704–716.

SWOFFORD, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods). 4.0Beta for Linux/UNIX. Sinauer Associates, Sunderland, MA.

THOLLESSON, M. 1999. Phylogenetic analysis of Euthyneura (Gastropda) by means of the 16S rRNA gene: use of a fast gene for higher-level phylogenies. Proceedings of the Royal Society of London Series B **266**: 75–83.

TILLIER, S., MASSELOT, M, GUERDOUX, J. & TILLIER, J. 1994. Monophyly of major Gastropod taxa tested form partial 28S rRNA sequences, with emphasis on Euthyneura and hot-vent limpets Peltospiroidea. The Nautilus **2**: 122–140.

TILLIER, S., MASSELOT, M. & TILLIER, A. 1996. Phylogenetic relationships of the pulmonata gastropods from rRNA sequences, and age of the stylommatophoran radiation. Pp. 267–284 in: Taylor, J. (ed.) Origin and Evolutionary Radiation of the Mollusca. Oxford Science, London.

VONNEMANN, V., SCHRÖDL, M. A., KLUSSMANN-KOLB & WÄGELE, H. 2004. Reconstruction of the Phylogeny of the Opisthobranchia (Mollusca: Gastropoda) by means of 18S and

28S rRNA Gene Sequences. Journal of Molluscan Studies **71**: 113–125.

XIA, X. & XIE, Z. 2001. DAMBE: Data analysis in molecular biology and evolution. Journal of Heredity **92**: 371–373.

XIA, X.; XIE, Z.; SALEMI, M.; CHEN, L. & WANG, Y. 2003. An index of substitution saturation and its application. Molecular Phylogenetics and Evolution **26**: 1–7.

Authors' addresses: Angela DINAPOLI (corresponding author) and Annette KLUSSMANN-KOLB, Institut for Ecology, Evolution and Diversity – Phylogeny and Systematics, J. W. Goethe-University, Frankfurt am Main, Germany, a.dinapoli@bio.uni-frankfurt.de.