THE H3/H4 HISTONE GENE CLUSTER OF LAND SNAILS (GASTROPODA: STYLOMMATOPHORA): TS/TV RATIO, GC3 DRIVE AND SIGNALS IN STYLOMMATOPHORAN PHYLOGENY

GEORG F.J. ARMBRUSTER[1], MANJA BÖHME[2], DETLEF BERNHARD[2] AND MARTIN SCHLEGEL[2]

[1]*University of Basel, Department of Integrative Biology, Section of Conservation Biology, St Johanns Vorstadt 10, CH 4056 Basel, Switzerland;*
[2]*University of Leipzig, Department of Zoology, Institute of Molecular Evolution & Animal Systematics, Talstr. 33, D 04103 Leipzig, Germany*

(Received 6 October 2004; accepted 1 March 2005)

ABSTRACT

Histone gene primers were developed for land snails (Stylommatophora). The partial H3/H4 histone gene cluster was cloned and sequenced for 18 species. Transcription of the H3 and H4 genes was divergent (each gene is transcribed in the opposite direction) as has been found for other protostome and diploblast animals, with the exception of *Mytilus*. In the bivalve *Mytilus* transcription of both genes occurs in the same direction, i.e. land snails and bivalves seem to differ in their histone gene organization. The non-transcribed H3/H4 spacer varied in length between 279 and 691 basepairs. Nucleotide polymorphisms in this non-transcribed spacer might be of significance to study phylogenetics and systematics of closely related species and genera. As expected, the coding regions exhibited no amino acid substitution among land snail species. However, one amino acid substitution was found in comparison between land snails and *Drosophila*. The transition/transversion (TS/TV) ratio of H3 and H4 was predominately shaped by the third codon position and ranged in most cases from 1.0 to 2.0, indicating low nucleotide saturation. GC content was calculated for the third codon position (GC3 index at the 'wobble' base position). The histone GC3 values were far lower in land snails than values currently available for other genomes (i.e. mammals). This indicates that H3/H4 histone wobble bases of land snails evolve without strong GC drive. Phylogenetic trees were reconstructed from the coding regions. We used *Succinea putris* (Elasmognatha) as outgroup. *Trichia villosa* (Helicoidea) showed six apomorphic nucleotide signals. Moreover, the nucleotide signals give evidence that the Cochlicopidae, Vertiginidae and Valloniidae are paraphyletic family categories. The paraphyletic status of cochlicopid, vertiginid and valloniid gastropods is also supported by our unpublished ribosomal DNA trees.

INTRODUCTION

Histone proteins H3 and H4 occur in all animals, plants, fungi and protists. The H3 and H4 proteins, together with proteins H2A and H2B, form the core architecture in the nucleosome complex (Lewin, 2000). In non-deuterostome Metazoa and many protists, the H3 and H4 genes are typically organized in tandem without introns, separated by a non-transcribed intergenic spacer (see Miller *et al*., 1993; Rooney, Piontkivska & Nei, 2002; Piontkivska, Rooney & Nei, 2002). Moreover, both genes are transcribed in opposite direction (bold arrows in Fig. 1; see Miller *et al*., 1993; Bernhard & Schlegel, 1998; Baldo, Les & Strausbaugh, 1999). In the bivalve mollusc *Mytilus edulis*, however, the H3/H4 gene cluster is transcribed in the same direction (Albig *et al*., 2003). A further interesting aspect of H3/H4 organization is the mode of molecular evolution. Multicopies of the gene tandem are subjected to concerted evolution in the gene pool of a species. Hence, one sequence is usually sufficient to characterize the H3/H4 pattern of a species (see Coen, Strachan & Dover, 1982; Liao, 1999).

In the following contribution, general sequence patterns and phylogenetic signals of the H3/H4 gene cluster are analysed for land snails (Stylommatophora). The deduced H3 and H4 amino acid sequences were analysed for polymorphisms. We expected the amino acid sequences to be invariant because the core histones belong to highly conserved proteins in organismic evolution (see Wells & McBride, 1989; Waterborg & Robertson, 1996). Nevertheless, phylogenetic information might be found at the nucleotide level. This study elucidates the GC content of the non-coding spacer and the transition/transversion ratio (TS/TV) of the coding regions. These indices are informative in studies of nucleotide saturation in phylogenetic studies (Abouheif, Zardoya & Meyer, 1998). GC3 indices were also estimated (Birdsell, 2002; Galtier, 2003). Galtier (2003) found a high frequency of G or C nucleotides at the third codon position of histone genes of human and mouse. High GC3s are explained by an asymmetric process of copied-pasted gene conversion. The amino acid sequence of the converted sequences remains identical because of the degenerative universal code at the third codon position (i.e. the 'wobble' base). Enrichment of GC content at the 'wobble' base is driven by the fact that G and C nucleotides have more stable bonds than A and T. All these factors theoretically explain high GC3 indices in histone genes of vertebrates. Here, GC3 indices of land snails were calculated and compared with the values found in mammalian genomes.

The H3/H4 sequences are also used for investigating phylogenetic aspects within the land snail group Stylommatophora. The Stylommatophora form a group of 20,000 to 30,000 extant species belonging to 70–90 families. The stemlines of the Stylommatophora reach back to a Mesozoic radiation, over 70 Ma (see Emberton *et al*., 1990; Bieler, 1992; van Bruggen, 1995; Tillier, Masselot & Tillier, 1996). The classification system of land snails has been a controversial issue for many years, particularly

Correspondence: G.F.J. Armbruster; e-mail: g.armbruster@unibas.ch

as the phylogeny and systematic relationships at the family level are poorly understood (see Pilsbry, 1900; Nordsieck, 1985, 1990, 1992; Tillier, 1989; Emberton, 1991; Pokryszko, 1994; Tillier & Mordan, 1995, Barker, 2001). Beside morphological studies, ribosomal RNA genes have also been used to resolve the relationships within the Stylommatophora (Emberton *et al*., 1990; Tillier *et al*., 1992, 1996; Yoon & Kim, 2000). The most exhaustive study was by Wade, Mordan & Clarke (2001) using 5.8S and 28S rRNA gene sequences. The authors provided a neighbour-joining (NJ) analysis of the phylogeny of land snails, using a wide range of family taxa. However, the resolution of the NJ topology is limited. Several families receive low bootstrap support in their tree and unresolved multifurcations occur, particularly at basal nodes. Because the ribosomal sequence tree yields these multifurcations, we tested the coding parts of the H3 and H4 genes as a molecular marker in land snail systematics. Phylogenetic implications are compared with results obtained from ribosomal RNA gene trees.

MATERIAL AND METHODS

DNA extraction, histone gene primers, PCR, cloning and sequencing

DNA was extracted from freshly frozen or ethanol-preserved individuals. Two primer pairs were developed in nested positions of the core histone genes H3 and H4 (Fig. 1). Primer sequences were designed on the basis of consensus motifs of marine molluscs (H3; Colgan, Ponder & Eggler, 2000), of a coral (H4; Miller *et al*., 1993) and of a polychaete worm (H4; del Gaudio *et al*., 1998). Sequences for forward (F) and reverse (R) primers are:

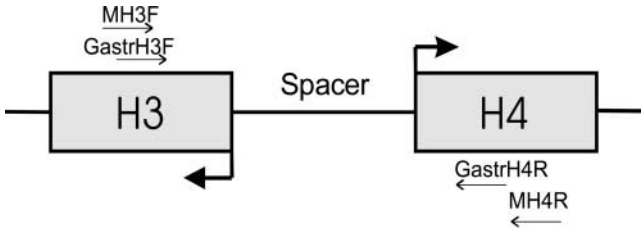

Figure 1. Tandem arrangement of the histone gene cluster H3/H4 in protostome animals and protists. Transcription of genes H3 and H4 is in opposite directions (bold arrows). Nested primer positions for land snails are indicated.

MH3(F): 5-TTCTGGTAAGGACGGATCTC-3,
MH4(R): 5-AGGGCRTAGACRACATCCAT-3,
GastrH3(F): 5-GTGCTCTTCTGGTAACGACG-3,
GastrH4(R): 5-TCGGTGTAGGTGACGGCATC-3.

After running pre-analyses, we noted that PCR of gastropod H3-spacer-H4 is often difficult. PCR was performed in volumes of 50 µl, including 10–30 ng DNA, 5 µl of 10 × PCR-Buffer (Boehringer), forward and reverse primer (10 pmol each; most successful was the combination GastrH3F/GastrH4R), 0.2 mM dNTPs, 3 mM $MgCl_2$ and 1 U of Taq polymerase (Boehringer). Thermal cycling conditions were 94°C for 1 min, primer annealing at 50–55°C for 90 s, 72°C for 90 s, cycled 35 times with a prolonged last step at 72°C for 10 min. Prominent PCR fragments of appropriate length (i.e. ≈600–1100 bp) were cut out from agarose gels, cleaned and

Table 1. Origin of samples used for the H3/H4 investigation.

Sample	Sampling locality	Accession no.
Elasmognatha: Succineidae, *Succinea putris* (Linnaeus, 1758), clones #33	Leipzig, Saxony, Germany	AY559145
and #53 of two individuals		AY559146
Helicoidea: Hygromiidae, *Trichia villosa* (Draparnaud, 1805), 1 clone	Blumberg, Baden-Württemberg, Germany	AY559147
Endodontidea: Punctidae, *Punctum pygmaeum* (Draparnaud, 1801), 1 clone	Balaton, Hungary	AY559148
Clausiliidae: *Cochlodina laminata* (Montagu, 1803), 1 clone	Leipzig, Saxony, Germany	AY559149
Pupillidae: *Pupilla muscorum* (Linnaeus, 1758), 1 clone	Martinfeld, Thuringia, Germany	AY559155
Enidae: *Ena montana* (Draparnaud, 1801), 1 clone	Kirchheim/Teck, Baden-Württemberg, Germany	AY559164
Vertiginidae: *Vertigo antivertigo* (Draparnaud, 1801),	Lake Neusiedl, Austria	AY559160
3 clones of one individual (blank, #5 and #6)		AY559161
		AY559162
Vertiginidae: *Truncatellina cylindrica* (A. Férussac, 1807), 1 clone	Martinfeld, Thuringia, Germany	AY559151
Vertiginidae: *Columella edentula* (Draparnaud, 1805), 1 clone	Martinfeld, Thuringia, Germany	AY559152
Vertiginidae: *Columella edentula* (Draparnaud, 1805), clones from two	Tsuga, White Sea, Russia	AY559153
individuals (#48R, #15R)		AY559154
Cochlicopidae: *Azeca goodalli* (A. Férussac, 1821), 1 clone	Martinfeld, Thuringia, Germany	AY559150
Cochlicopidae: *Cochlicopa nitens* (von Gallenstein, 1848), 1 clone	Budyne n.O., Czech Republic	AY559156
Cochlicopidae: *Cochlicopa lubricella* (Porro, 1838), 1 clone	Bremen, Germany	AY559157
Cochlicopidae: *Cochlicopa lubrica* (O.F. Müller, 1774), 1 clone #1	Budyne n.O., Czech Republic	AY559158
Cochlicopidae: *Cochlicopa lubrica* (O.F. Müller, 1774), 1 clone #2	Geisslingen, Baden-Württemberg, Germany	AY559159
Valloniidae: *Acanthinula aculeata* (O.F. Müller, 1774), 1 clone	Mühlhausen, Thuringia, Germany	AY559163
Valloniidae: *Vallonia costata* (O.F. Müller, 1774), 1 clone	Leipzig, Saxony, Germany	AY559165
Valloniidae: *Vallonia enniensis* (Gredler, 1856), 1 clone	Kiskunshagi Parc, Hungary	AY559166
Valloniidae: *Vallonia pulchella* (O.F. Müller, 1774), 1 clone #55	Martinfeld, Thuringia, Germany	AY559167
Valloniidae: *Vallonia pulchella* (O.F. Müller, 1774), 1 clone #9	Leipzig, Saxony, Germany	AY559168
Valloniidae: *Vallonia excentrica* Sterki, 1892, 1 clone #17	Zeilitz, Saxony, Germany	AY559169
Valloniidae: *Vallonia excentrica* Sterki, 1892, 1 clone #23	Rötha, Saxony, Germany	AY559170

Family classification according to the systematic overview given in Emberton *et al.* (1990).

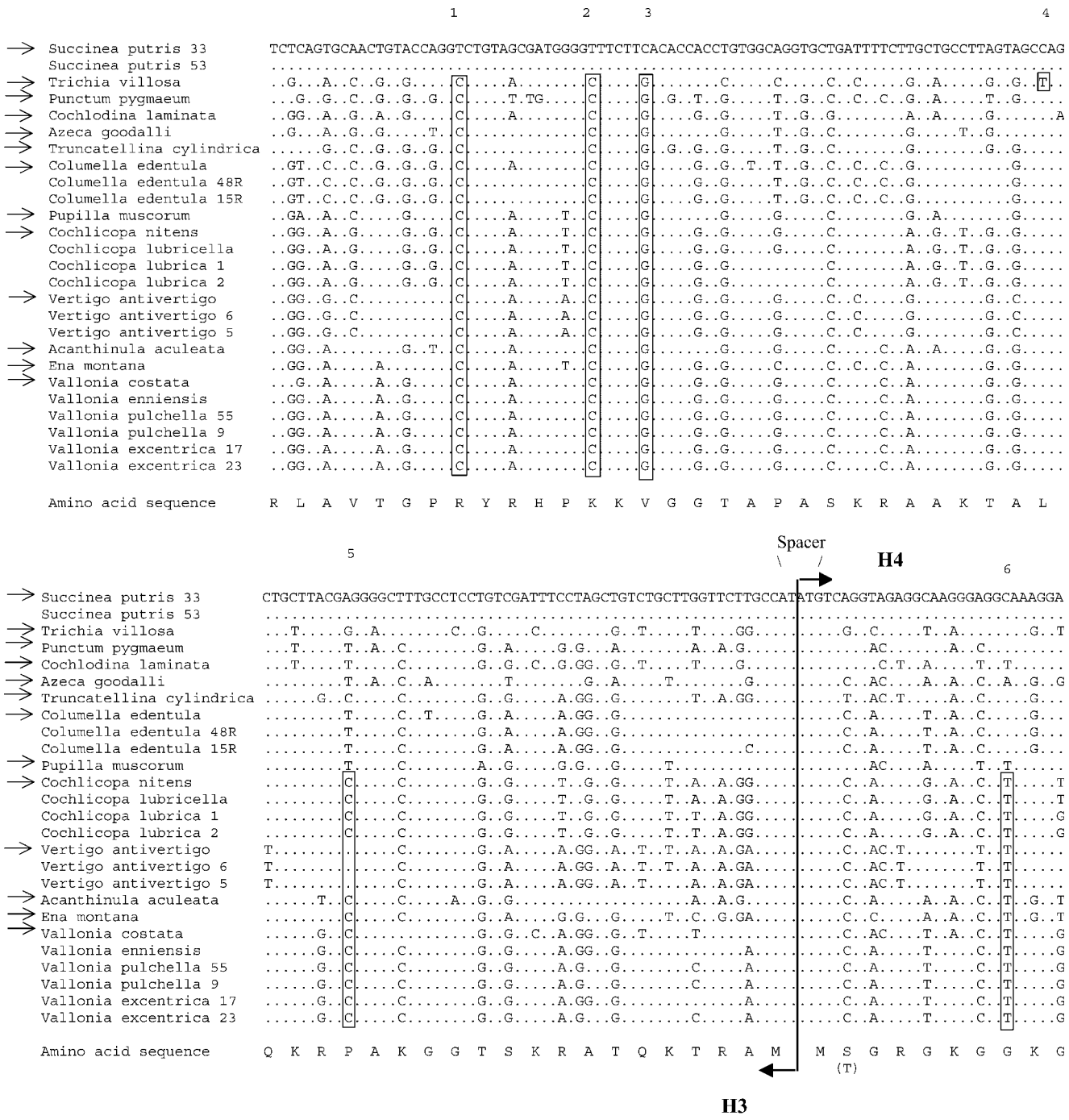

Figure 2. Alignment of the 26 coding H3 and H4 sequences of 18 land snail species. The non-coding spacer (Fig. 1) situated between H3 and H4, has been omitted (see vertical bold line). Direction of transcription is superimposed with bold arrows. The horizontal arrows (→) denote those sequences used for GC3 analysis (see Material and Methods). The amino acid sequence is shown in single letter code below the alignment, with one S/T substitution between land snails and *Drosophila* in the 5′ region of H4. Consecutive numbers above the alignment are evaluated as single nucleotide character boxes.

ligated in a pGEM-T vector (Progmega), and transformed into *Escherichia coli*. Each clone was completely sequenced in both directions with fluorescent-labelled M13 forward and reverse primers. Sequencing reactions were performed with a 7-deaza-dGTP sequencing kit (Amersham) and separated on an automated LI-COR DNA sequencer. Accession numbers are shown in Table 1. Out of 33 tested species of land snails and freshwater gastropods, sequences could be obtained from only 18 species (13 genera of nine families; Table 1). 'Negative results' were found in PCR signals of a broad spectrum of animals: *Cepaea*, *Helix* and *Perforatella* (Helicoidea), *Sphyradium* (Orculidae), *Deroceras* (Agriolimacidae), *Vitrina* (Vitrinidae), *Acavus* (Acavidae), *Discus* (Discidae), *Ruthenica* (Clausiliidae), *Physa* and *Planorbarius* (Basommatophora: Planorbidae), and *Cochlostoma* ('Prosobranchia'). In these groups amplification products were not the target gene.

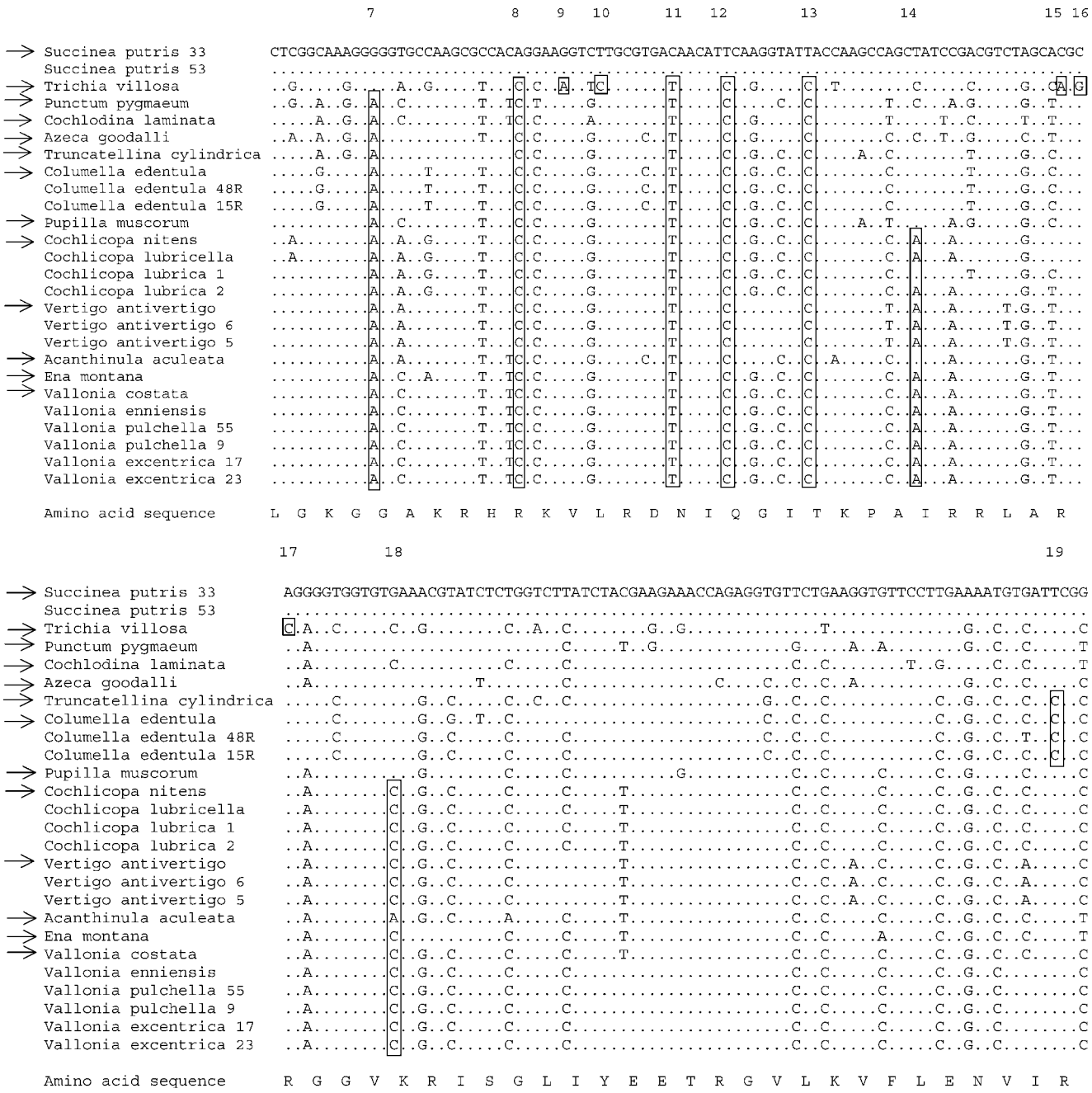

Figure 2. continued

GC content, TS/TV analyses, GC3s, and phylogenetic trees

GC content and TS/TV ratios were calculated using MEGA2 (Kumar, Tamura & Nei, 2001). Analyses of the GC3 indices at the third codon position were based on the following phylogenetic assumptions. Firstly, if continuous GC3 drive occurs during evolution, a shift of A/T to G/C could result in GC3 saturation of the histone genes. Here, reference sequences of 13 genera were taken into consideration in order to avoid GC3 intercorrelations of closely related species (marked with → in front of the sequence; Fig. 2). Secondly, the GC3 value was analysed for those third codon positions that theoretically can 'wobble' in all four nucleotides without altering the amino acid sequence. Therefore, codons abbreviated with G, A, L, S, P, V, R and T were used (letters denote international amino acid code). Using this strategy, 37 third codon positions were evaluated for the H3 gene and 43 positions for the H4 gene (Fig. 2).

Sequences were aligned using the multiple alignment program Clustal X (Thompson *et al*., 1997). As an outgroup we used two sequences of *Succinea putris* (Elasmognatha). Species of the elasmognath group belong to an old and basal branch of terrestrial gastropods (Tillier *et al*., 1996; Wade & Mordan, 2000).

The non-transcribed spacer between H3 and H4 (Fig. 1) was variable in length and nucleotide composition and could not be aligned across all taxa (see Results). Hence, phylogenetic trees (Figs 4, 5) were based on 354 nucleotides of the concatenated

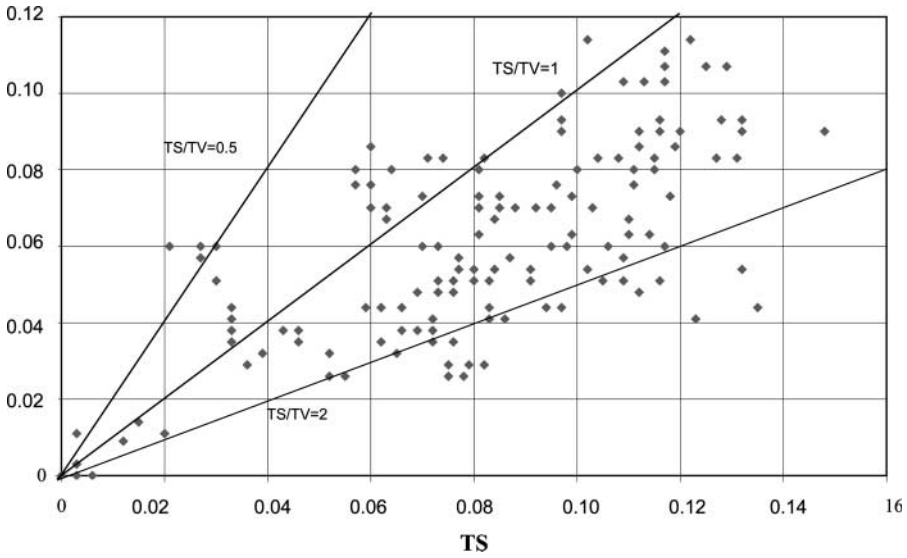

Figure 3. Scatter plot of transition/transversion ratio (TS/TV) in pairwise sequence comparison of the concatenated histone H3 and H4 gene sequences.

histone H3 and H4 gene sequences without the spacer. A maximum likelihood tree (ML; Fig. 4) was calculated using PAUP* 4.0 b10 (Swofford, 2002). To find the most appropriate model of DNA substitution we carried out a hierarchical likelihood ratio test with Modeltest v. 3.06 (Posada & Crandall, 1998). Using the Akaike information criteria (Akaike, 1973), the model selected was TVM + I+ Γ (transversion model with invariable sites and gamma distribution) with base frequencies of A = 0.2160, C = 0.2621, G = 0.2923, T = 0.2296; a rate matrix of [A–C] = 0.7480, [A–G] = 3.0493, [A–T] = 1.3608, [C–G] = 0.2224, [C–T] = 3.0493, [G–T] = 1.0000; a proportion of invariable sites of 0.6328; and a gamma distribution shape parameter of G = 1.8685. The same model was applied in a Bayesian analysis (Fig. 5), carried out with MrBayes (Huelsenbeck & Ronquist, 2001) and 1,000,000 generations, with a sampling frequency of 100 generations. From the 10,000 trees found, the first 1,000 were discarded as burn-in, the remaining trees were used to calculate the consensus tree. A neighbour-joining tree (NJ; Saitou & Nei, 1987) was also constructed using TREECON (Van de Peer & De Wachter, 1994) using Kimura's (1980) two-parameter model (see bootstrap values implemented in Fig. 4). Maximum parsimony analysis (MP) was carried out with PAUP* version 4.0 b10, with heuristic search using 10 stepwise additions of sequences and TBR branch swapping option (tree not shown; see text). To test the robustness of NJ and MP, bootstrap analyses with 1,000 replicates were performed.

RESULTS

The consensus alignment of the clones is shown in Fig. 2. The H3 and H4 genes are arranged in opposed direction of transcription (bold arrows in Fig. 2). The non-coding spacer between the H3 and H4 genes is omitted in Figure 2 because the spacer sequences cannot be aligned across all taxa. Spacer length varied from 279 bp in the vertiginid *Vertigo antivertigo* to 691 bp in the helicoid *Trichia villosa* (Table 2). Moreover, the length of the spacer can differ within a nominal species (e.g. in *Columella edentula*; Table 2). The alignment covers 50 codon positions in the H3 gene, and 68 codons in the H4 gene. Deduced amino acid sequence of the histones H3 and H4 is displayed below the nucleotide sequence alignment (Fig. 2). Within land snails, no amino acid variation was found. However, one amino acid substitution is indicated in the histone gene H4 of land snails and *Drosophila*. The substitution S/T is marked in the alignment.

The overall GC content was higher in the coding sequences than in the non-coding spacer (Table 2), i.e. 52–63% of GC in the histone H3 and H4 genes, and 39–53% of GC in the non-coding spacer. The TS/TV ratio of the H3/H4 gene sequences is shown in Fig. 3. Most pairwise sequence comparisons fall in the TS/TV ratio between 1.0 and 2.0. However, outliers can also be seen in Figure 3 (i.e. TS/TV $>$ 2.0 and TS/TV $<$1.0). The results of the GC3 analysis are listed in Table 3. The average value of the 13 genera was 54% GC3 in H3, and 49% GC3 in H4. The lowest and the highest GC3 indices were both found in the H3 gene, i.e. 30% GC3 (in the outgroup sequence of *Succinea putris*) and 78% GC3 (in the sequence of *Truncatellina cylindrica*).

Figures 4 and 5 show the phylogenetic trees inferred from the coding histone H3 and H4 regions. Sequences of congeneric species (*Cochlicopa lubrica*, *C. lubricella* and *C. nitens*, as well as *Vallonia excentrica*, *V. pulchella*, *V. enniensis* and *V. costata*) form distinct groups. The helicoid *Trichia villosa* branches off at the first node. In the next higher clade, two species groups are well supported: (i) *Truncatellina cylindrica* and *Columella edentula* (56% likelihood and 92% bootstrap support, Fig. 4, and a posterior probability of 86%, Fig. 5), and (ii) the species cluster of *Vertigo*, *Vallonia*, *Acanthinula*, *Ena* and *Cochlicopa* (21% likelihood, 70% bootstrap support in Neighbour Joining, and posterior probability of 84%). Maximum Parsimony (MP) analysis did not improve the resolution of the branching patterns (data not shown). Again, the helicoid sequence splits off first, and the above mentioned groups (i) and (ii) appeared with 85% and 51% bootstrap probability. Interestingly, the histone tree topologies provide evidence that the families Vertiginidae, Cochlicopidae and Valloniidae might be artificial units. The potential paraphyletic status is superimposed with asterisks (*** for cochlicopids, ** for vertiginids and * for valloniids). The systematic position of *Cochlodina*, *Punctum*, *Azeca* and *Pupilla* remained unresolved in the analyses, with branching values of less than 50%.

The coding regions were also evaluated in a manual analysis to find apomorphic signals (see the consecutive numbers above the alignment; Fig. 2). The sequences of *Succinea putris* exhibited

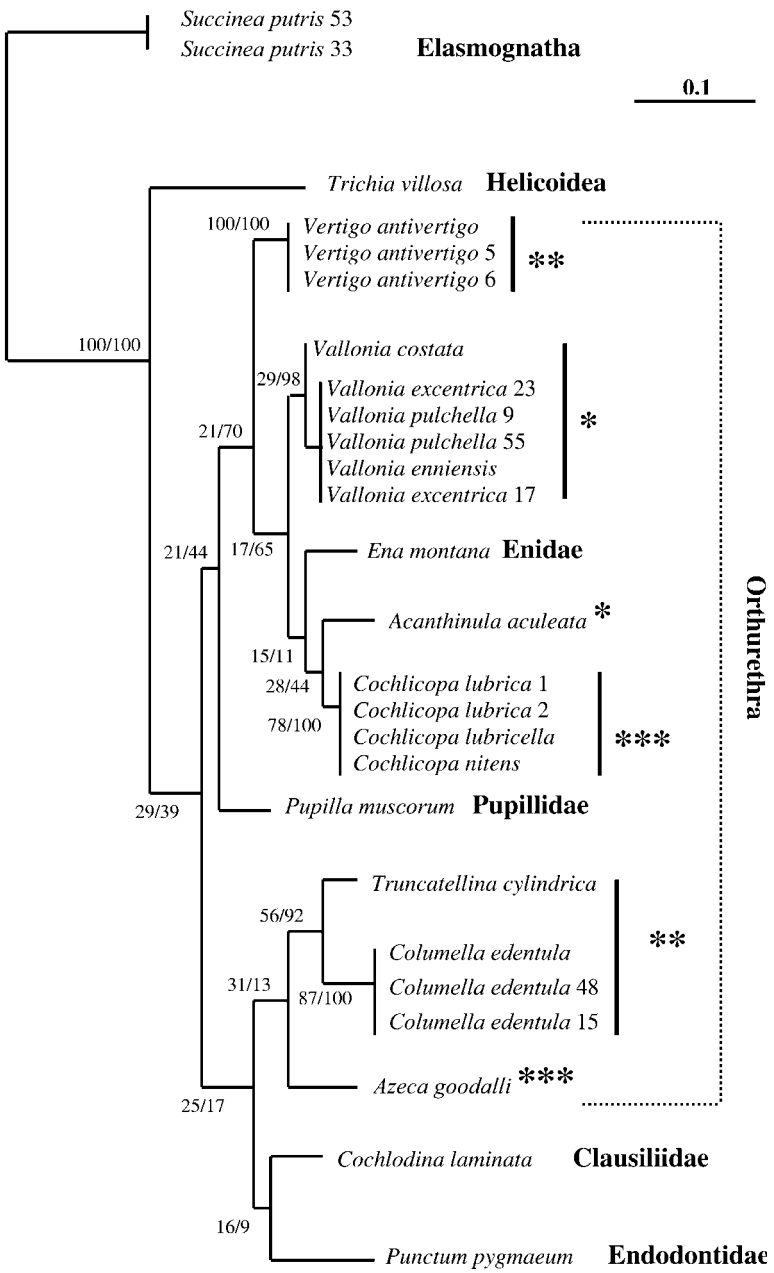

Figure 4. Maximum-likelihood tree obtained from the concatenated H3 + H4 coding sequences based on the alignment given in Figure 2. Values at the nodes denote maximum likelihood probabilities (first number) and bootstrap values of the NJ analysis (second number). The asterisks indicate potential paraphyletic family groups (*** for Cochlicopidae; ** for Vertiginidae, * for Valloniidae).

seven markers, which were defined as plesiomorphic character states of the outgroup. Following this definition, the entire ingroup exhibited seven autapomorphies (box positions 1, 2, 3, 8, 11, 12, 13; Fig. 2). Within this ingroup, the helicoid snail *Trichia villosa* had the highest amount of exclusive markers (six identified apomorphies; positions 4, 9, 10, 15, 16, 17). The next higher clade (without *Trichia villosa*) showed one apomorphy (No. 7 = A). The closeness of the investigated orthurethran snails (genera *Azeca*, *Truncatellina*, *Columella*, *Pupilla*, *Cochlicopa*, *Vertigo*, *Acanthinula*, *Ena* and *Vallonia*) was not confirmed, i.e. there was no distinct apomorphic signal in the H3/H4 sequence alignment supporting a common origin. Nevertheless, the above mentioned species group (i) of *Truncatellina* and *Columella* showed one identified autapomorphy (19 = A), and the species cluster (ii) of *Vertigo*, *Vallonia*, *Acanthinula*, *Ena* and *Cochlicopa* displayed sequence similarities at four positions (box positions 5, 6, 14, 18). The latter nucleotide patterns give a hint for a common origin of group (ii) although the respective characters can also be found in other taxa.

DISCUSSION

General patterns and TS/TV ratio

Colgan *et al.* (2000) provided the first data on histone H3 genes of marine and limnic molluscs. Their study points to high fixation of the deduced amino acid sequences. However, Colgan and coworkers did not extend their study to include the H4 gene and the histone spacer. Our investigation confirms that histones H3 and H4 belong to highly

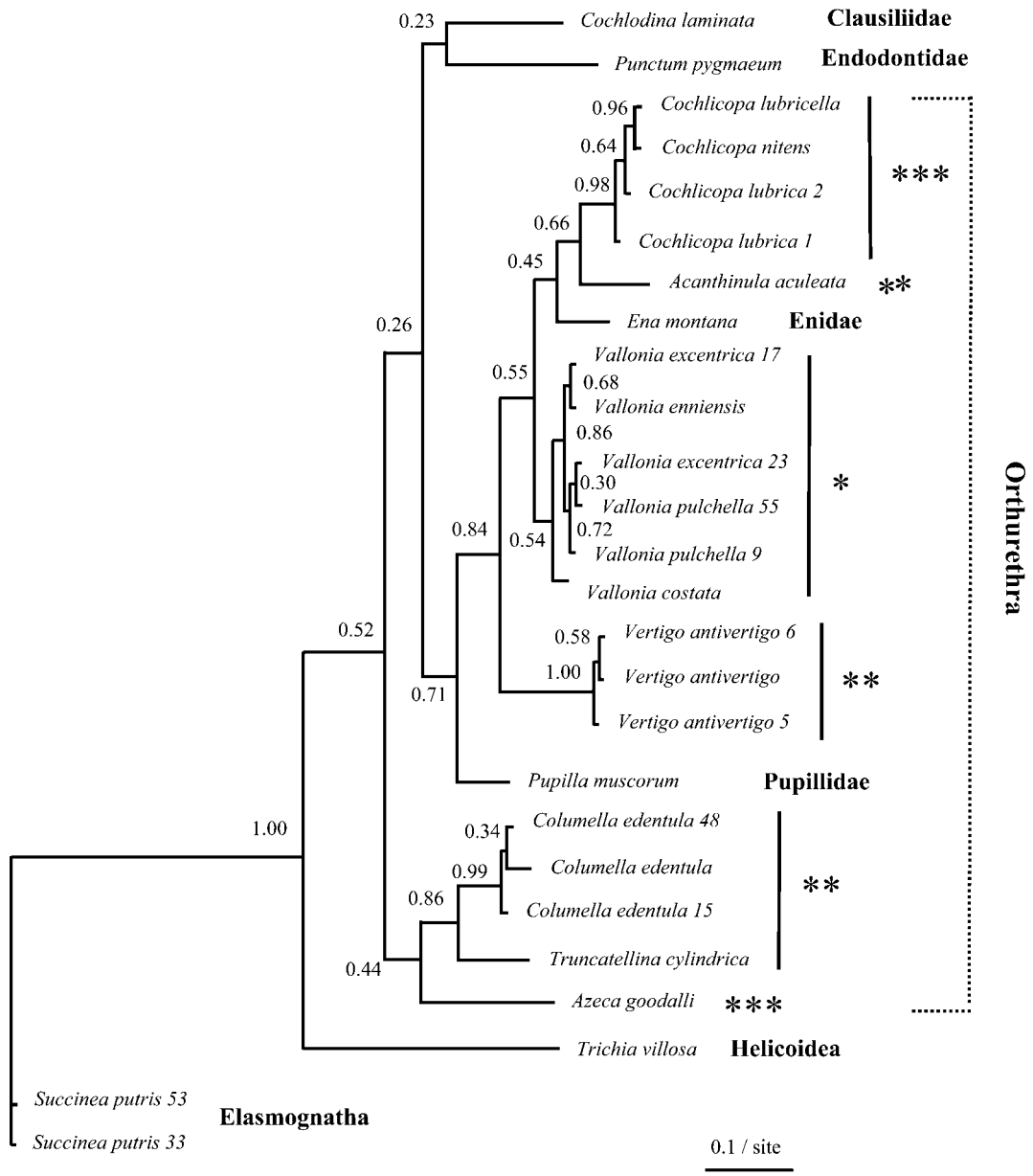

Figure 5. Bayesian tree obtained from the concatenated H3 + H4 coding sequences based on the alignment given in Figure 2. Values at the nodes refer to posterior probabilities for the branching pattern. The asterisks denote potential paraphyletic family groups (*** for Cochlicopidae; ** for Vertiginidae, * for Valloniidae).

conserved proteins in animal evolution. No differences in the deduced amino acid sequences were found in land snails, and one amino acid substitution was detected in comparison to *Drosophila* (Fig. 2).

Divergently orientated transcription of the genes H3 and H4 has been described in numerous non-deuterostome Metazoa and protists. Our data show that stylommatophoran gastropods also have divergent transcription of H3 and H4 genes (bold arrows in Fig. 2). However, we can confirm this finding only for the 18 species investigated since the H3/H4 genes could not be amplified for other snail groups (see Material and Methods). The divergent transcription is in contrast to the bivalve *Mytilus edulis* (Albig *et al*., 2003) in which transcription of the H3 and H4 genes is orientated in identical direction, i.e. on the same DNA strand. Thus, there is strong evidence that the genomes of land snails and *Mytilus* may differ in the organization of their H3/H4 gene cluster. The H3-spacer-H4 gene cluster of other molluscs still needs to be investigated to clarify the orientation of transcription for more taxa.

The length of the non-transcribed H3/H4 spacer (279–691 bp; Table 2) fits well with data available for invertebrates, e.g. 263 bp in the coral *Acropora formosa* (Miller *et al*., 1993) and about 800 bp in the polychaete worm *Chaetopterus variopedatus* (del Gaudio *et al*., 1998). Closely related land snail species (e.g. *Cochlicopa lubrica*, *C. lubricella* and *C. nitens*) exhibited point mutations in the H3/H4 spacer (data not shown). Even within a nominal species (*Columella edentula*) spacer length varied. Such a finding is in contrast to concerted evolution and homogenization of the H3-spacer-H4 gene cluster within a species. We therefore suppose that length variation and nucleotide polymorphisms in this histone spacer should be studied more intensively at the taxonomic level of genera, species, populations and individuals.

Table 2. Nucleotide content in the coding part of H3 and H4, and of the non-coding spacer, with reference to length in base pairs (bp).

Species	Coding part of the H3 + H4 sequence							Non-coding spacer between H3 and H4						
	G	C	T	A	GC	AT	Length (bp)	G	C	T	A	GC	AT	Length (bp)
Punctum pygmaeum	32.5	24.3	23.4	19.8	56.8	43.2	354	22.6	16.3	32.1	29.0	38.9	61.1	686
Trichia villosa	32.8	26.6	22.0	18.6	59.4	40.6	354	17.2	29.4	23.2	30.2	46.6	53.4	691
Pupilla muscorum	31.4	25.7	22.9	20.1	57.1	43.0	354	22.6	21.2	26.9	29.3	43.8	56.2	624
Vallonia costata	31.6	26.8	22.3	19.2	58.4	41.5	354	25.6	21.1	25.9	27.4	46.7	53.3	540
Vallonia pulchella (55)	32.2	27.1	20.9	19.8	59.3	40.7	354	26.7	22.4	25.2	25.7	49.1	50.9	544
Vallonia pulchella (9)	32.2	27.1	20.9	19.8	59.3	40.7	354	26.7	22.5	25.2	25.6	49.2	50.8	543
Vallonia enniensis	32.8	26.8	20.9	19.5	59.6	40.4	354	27.4	22.6	24.8	25.2	50.0	50.0	548
Vallonia excentrica (23)	32.2	27.1	20.9	19.8	59.3	41.7	354	26.7	22.5	25.4	25.4	49.2	50.8	543
Vallonia excentrica (17)	32.8	26.8	20.9	19.5	59.6	41.4	354	26.8	22.1	25.2	25.9	48.9	51.1	548
Vertigo antivertigo	30.5	25.1	22.9	21.5	55.6	44.4	354	24.4	21.1	30.5	24.0	45.5	54.5	279
Vertigo antivertigo (5)	30.5	25.4	22.6	21.5	55.9	44.1	354	24.4	21.1	30.5	24.0	45.5	54.5	279
Vertigo antivertigo (6)	30.5	25.1	22.9	21.5	55.6	44.4	354	24.4	21.1	30.5	24.0	45.5	54.5	279
Ena montana	30.8	26.6	21.5	21.2	57.4	42.7	354	21.7	21.3	29.1	27.9	43.0	57.0	506
Cochlicopa nitens	32.8	25.4	21.8	20.1	58.2	41.9	354	23.8	19.9	28.7	27.6	43.7	56.3	543
Cochlicopa lubricella	32.8	25.4	21.8	20.1	58.2	41.9	354	23.6	19.7	28.9	27.8	43.3	56.7	543
Cochlicopa lubrica (1)	32.8	25.7	22.0	19.5	58.5	41.5	354	21.9	20.0	30.7	27.3	41.9	58.1	524
Cochlicopa lubrica (2)	32.8	25.7	22.0	19.5	58.5	41.5	354	21.9	20.0	30.7	27.3	41.9	58.1	524
Acanthinula aculeata	30.5	26.0	22.0	21.5	56.5	43.5	354	20.1	21.8	28.2	29.9	41.9	58.1	298
Succinea putris (33)	29.1	22.9	26.6	21.5	52.0	48.1	354	20.4	17.2	36.9	25.6	37.6	62.5	309
Succinea putris (56)	29.1	22.9	26.6	21.5	52.0	48.1	354	20.4	17.2	36.6	25.9	37.6	62.5	309
Azeca goodalli	31.1	25.7	23.7	19.5	56.8	43.2	354	27.0	21.7	23.8	27.5	48.7	51.3	488
Columella edentula	32.9	27.5	21.8	17.8	60.3	39.6	354	21.5	25.2	27.9	25.5	46.7	53.4	326
Columella edentula (15R)	33.3	28.8	20.3	17.5	62.1	37.8	354	21.6	26.2	26.0	26.2	47.8	52.2	389
Columella edentula (48R)	33.3	28.2	20.9	17.5	61.5	38.4	354	22.4	26.0	26.3	25.3	48.4	51.6	388
Truncatellina cylindrica	34.7	28.8	20.1	16.4	63.5	36.5	354	23.6	27.1	26.4	22.9	50.7	49.3	402
Cochlodina laminata	31.9	23.7	24.6	19.8	55.6	44.4	354	25.0	28.5	24.3	22.2	53.5	46.5	432

The overall GC content is higher in the coding H3/H4 regions than in the non-transcribed spacer (Table 2). Such differences between coding genes and non-coding regions are well known (see Miller *et al*., 1993). The TS/TV scatter plot of the coding H3 and H4 (Fig. 3) shows that the threshold of <0.5 is seldom exceeded. Values of <0.5 point to high nucleotide saturation, with weak phylogenetic information because of too many transversions (see Abouheif *et al*., 1998). Thus, the H3/H4 gene cluster with its TS/TV range of 1.0–2.0 may be a promising genomic region for reconstructing phylogenetic relationships (see below).

GC3 drive in the H3/H4 gene cluster

The organization of the H3 and H4 genes of deuterostome animals is different to the tandem-situation (Fig. 1) found in most protostomes. The H3 and H4 genes of deuterostomes are clustered in a loose arrangement on the chromosomes, including single and neighbouring loci (Rooney *et al*., 2002; Galtier, 2003).

Table 3. Results of the GC3 analysis of the partial H3 and H4 genes for land snails (see Material and Methods).

	Histone gene H3	Histone gene H4
Average GC3 value	54%	49%
Minimum value	30% (in *Succinea putris*)	37% (in *Cochlodina laminata*)
Maximum value	78% (in *Truncatellina cylindrica*)	63% (in *Truncatellina cylindrica* and in *Columella edentula*)

In the deuterostome genome of human and mouse, histone copies have an enriched GC content at the third codon position with GC3 values between 70% and 95% (i.e. a high GC3 drive), with an *absolute minimum* of 55% GC3 (Galtier, 2003: 66). Table 3 shows that the H3 and H4 copies of land snails vary between 30% and 78% GC3. The *average* value was 54% in H3 and 49% in H4. Hence, a generally high GC content at the third codon position was not detected for the H3 and H4 genes of the investigated land snail genera. Although speculative, it could be that asymmetric gene conversion and GC3 drive might somehow be suppressed by the tandem organization of the H3/H4 cluster (see Fig. 1). Possibly, physiological factors may also be involved in GC3 drive mechanisms (Galtier, 2003: 67). Since GC3 histone data for other invertebrates are not yet available, we think that additional studies of protostomes/diploblasts *vs* deuterostomes are of significance to understand the mechanisms of the 'wobble' base evolution in this protein-coding gene cluster. For example, more work should be done in comparing entire H2A, H2B, H3 and H4 gene stretches.

Phylogenetic signals in the present data set

The ingroup has seven apomorphic histone signals in common (Box No. 1, 2, 3, 8, 11, 12, 13; see Fig. 2). This finding demonstrates a clear separation of the ingroup from the elasmognath outgroup.

The trees show that the helicoid *Trichia villosa* branches off at the first node (Figs 4, 5). This result supports the opinion of Wade *et al*. (2001) who interpret the Helicoidea as a basal branch in the system of the Stylommatophora. Interestingly, a high number of histone nucleotide apomorpies

(N = 6) are detected in *T. villosa* (see boxes in Fig. 2), and the next higher clade is characterized by one autapomorphy (No. 7 = A; Fig. 2).

Some nominal land snail families seem to be artificial categories. Schileyko's (1998a, b) comment on the Cochlicopidae, Vertiginoidea and Valloniidae is worth mentioning. Schileyko interprets the validity of these family categories as doubtful because of insufficient morphological characteristics. In the molecular study of Wade & Mordan (2000) and Wade *et al.* (2001) based on 5.8S and 28S ribosomal RNA genes, the taxonomic status of the three families was not discussed in depth. The histone trees suggest that the Cochlicopidae and Vertiginidae are paraphyletic families (Figs 4, 5). We suggest that the species *Azeca goodalli* and *Truncatellina cylindrica* (+ *Columella edentula*) should be removed from the Cochlicopidae and Vertiginidae, respectively. These species seem to belong to other evolutionary lineages than currently presumed in morphological systematics. Moreover, there is evidence that *Acanthinula aculeata* may represent a separate lineage outside the Valloniidae, because of its distant position from *Vallonia* species (see Figs 4, 5). Taking the present findings together, histone genes give support for the opinion of Schileyko that these are 'doubtful' family categories.

The resolution of deep branches is limited in molecular phylogeny of gastropods (Remigio & Hebert, 2003). The basal nodes receive low support in the original ribosomal gene tree of Wade *et al.* (2001: 416) and in our histone trees. These multifurcations seem due to an explosive radiation of terrestrial gastropods during the Upper Cretaceous/Palaeocene period (Tillier *et al.*, 1996). Deep-branch resolution can be re-addressed to the systematic validity of the Orthurethra. This group was again characterized as monophyletic (Wade *et al.* 2001: 416, 420). On the other hand, Pokryszko (1994) suggested orthurethran snails as a potentially paraphyletic group of gastropods. This would mean, for example, that the linearly stretched ureter of the Orthurethra is a homoplasy, or has evolved convergently. We also presume that the Orthurethra could indeed be paraphyletic since bootstrapping and posterior probabilities give no clear signal for a monophyletic origin (cf. Figs 4, 5).

In conclusion, the H3-spacer-H4 histone gene cluster provides a new marker for studying molecular evolution and systematics of the Stylommatophora. Further work is required to improve PCR success rate. The traditionally recognized families Cochlicopidae, Vertiginidae and Valloniidae seem to be artificial units in the stylommatophoran system. We previously investigated the phylogenetic status of these three families with ribosomal RNA gene sequences (accession nos AY546469–AY546471) and the methods given in Wade *et al.* (2001). The family status of the Cochlicopidae, Vertiginidae and Valloniidae again appeared paraphyletic (data and sequence trees are available from the corresponding author). However, neither H3/H4 nor ribosomal RNA gene trees can accurately resolve all branches in land snail phylogeny. Several nodes receive low bootstrap or posterior probability values, e.g. the monophyletic origin of the superfamily 'Orthurethra' again appears doubtful. This could be an effect of explosive radiation during evolution. Comparison of more genes and more gastropod taxa, and additional morphological studies, are necessary to elucidate particularly the basal splits. This may help to unravel the ramification in the extant family lineages. Resolving these problems is an ongoing challenge for both morphologists and molecular biologists.

ACKNOWLEDGEMENTS

We gratefully acknowledge discussions with P. Mordan (Natural History Museum, London), W. Rähle (Zoological Institute of University of Tübingen), A. Schileyko (Severtzov Institute of Russian Academy of Science, Moscow), and D. Colgan (Australian Museum, Sydney). Moreover, two anonymous reviewers gave valuable comments to improve the manuscript. We also thank W. Schier, Y. Haffner, A. Schneider and D. Pfefferle for their laboratory work, and M. Glaubrecht (Museum of Natural History, Berlin) for providing gastropod tissue samples. Thanks also to G. Fritzsch (IZBI, Leipzig University) for his help in computer analyses.

REFERENCES

ABOUHEIF, E., ZARDOYA, R. & MEYER, A. 1998. Limitations of metazoa 18S rDNA sequence data: implication for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *Journal of Molecular Evolution*, **47**: 394–405.

AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. In: *Proceedings of the second international symposium on information theory* (B.N. Petrov & F. Csaki, eds), 267–281. Akedemia Kiado, Budapest.

ALBIG, W., WARTHORST, U., DRABENT, B., PRATS, E., CORNUDELLA, L. & DOENECKE, D. 2003. *Mytilus edulis* core histone genes are organized in two clusters devoid of linker histone genes. *Journal of Molecular Evolution*, **56**: 597–606.

BALDO, A. M., LES, D.H. & STRAUSBAUGH, L.D. 1999. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of *Sophophora*. *Molecular Biology and Evolution*, **16**: 1511–1520.

BARKER, G.M. 2001. Gastropods on land: phylogeny, diversity and adaptive morphology. In: *The biology of terrestrial molluscs* (G.M. Barker, ed.), 1–146. CABI Publishing, Oxford.

BERNHARD, D. & SCHLEGEL, M. 1998. Evolution of histone H4 and H3 genes in different ciliate lineages. *Journal of Molecular Evolution*, **46**: 344–354.

BIELER, R. 1992. Gastropod phylogeney and systematics. *Annual Review of Ecology and Systematics*, **23**: 311–338.

BIRDSELL, J.A. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular Biology and Evolution*, **19**: 1181–1197.

COEN, E., STRACHAN, T. & DOVER, G. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *Journal of Molecular Biology*, **158**: 17–35.

COLGAN, D.J., PONDER, W.F. & EGGLER, P.E. 2000. Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. *Zoologica Scripta*, **29**: 29–63.

DEL GAUDIO, R., POTENZA, N., STEFANONI, P., CHIUSANO, M.L. & GERACI, G. 1998. Organization and nucleotide sequence of the cluster of five histone genes in the polychaete worm *Chaetopterus variopedatus*: first record of a H1 histone gene in the phylum Annelida. *Journal of Molecular Evolution*, **46**: 64–73.

EMBERTON, K.C. 1991. Polygyrid relations: a phylogenetic analysis of 17 subfamilies of land snails. *Zoological Journal of the Linnean Society*, **103**: 207–224.

EMBERTON, K.C., KUNCIO, G.S., DAVIS, G.M., PHILLIPS, S.M., MONDEREWICZ, K.M. & GUO, Y.H. 1990. Comparison of recent classification of stylommatophoran land-snail families, and evaluation of large-ribosomal-RNA sequencing for their phylogenetics. *Malacologia*, **31**: 327–352.

GALTIER, N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends in Genetics*, **19**: 65–68.

HUELSENBECK, J.P. & RONQUIST, F. 2001. MrBayes: Bayesian inference of phylogeny. *Biometrics*, **17**: 754–755.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**: 111–120.

KUMAR, S., TAMURA, K. & NEI, M. 2001. *MEGA2: Molecular Evolutionary Genetics Analysis*. Pennsylvania State University, University Park, PA.

LEWIN, B. 2000. *Genes VII*. Oxford University Press, Oxford.

LIAO, D. 1999. Concerted evolution: molecular mechanism and biological implications. *American Journal of Human Genetics*, **64**: 24–30.

MILLER, D.J., HARRISON, P.L., MAHONY, T.J., MCMILLAN, J.P., MILES, A., ODORICO, D.M. & TEN LOHUIS, M.R. 1993. Nucleotide sequence of the histone gene cluster in the coral *Acropora formosa* (Cnidaria; Scleractinia): features of histone gene structure and organization are common to diploblastic and triploblastic metazoans. *Journal of Molecular Evolution*, **37**: 245–253.

NORDSIECK, H. 1985. The system of the Stylommatophora (Gastropoda), with special regard to the systematic position of the Clausiliidae, I. Importance of the excretory and genital systems. *Archiv für Molluskenkunde*, **116**: 1–24.

NORDSIECK, H. 1990. Phylogeny and system of the Pulmonata (Gastropoda). *Archiv für Molluskenkunde*, **121**: 31–52.

NORDSIECK, H. 1992. Systematic revision of the Helicoidea (Gastropoda: Stylommatophora). *Harvard University Museum of Comparative Zoology Special Occasional Publications*, **8**: 1–79.

PILSBRY, H.A. 1900. On the zoological position of *Achatinella* and *Partula*. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **52**: 561–567.

PIONTKIVSKA, H., ROONEY, A. & NEI, M. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Molecular Biology and Evolution*, **19**: 689–697.

POKRYSZKO, B. 1994. On the mono(?)phyly and classification of the Orthurethra/Pupilloidea (Gastropoda: Pulmonata: Stylommatophora). *Genus*, **5**: 371–390.

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.

ROONEY, A.P., PIONTKIVSKA, H. & NEI, M. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Molecular Biology and Evolution*, **19**: 68–75.

SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**: 406–425.

SCHILEYKO, A.A. (1998a). Treatise on recent terrestrial pulmonate molluscs. Part 1. *Ruthenica*, Suppl. **2**: 94–104.

SCHILEYKO, A.A. (1998b). Treatise on recent terrestrial pulmonate molluscs. Part 2. *Ruthenica*, Suppl. **2**: 129–143.

SWOFFORD, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). 4.0b10 edn. Sinauer Associates, Sunderland, Massachusetts.

THOMPSON, J.D., GIBSON, T.J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D.G. 1997. The Clustal X window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**: 4876–4882.

TILLIER, S. 1989. Comparative morphology, phylogeny and classification of land slugs and snails (Gastropoda: Pulmonata: Stylommatophora). *Malacologia*, **30**: 1–303.

TILLIER, S. & MORDAN, P. 1995. The anatomy and systematics of the New Caledonian land snail genus *Draparnaudia* Montrouzier, 1859 (Pulmonata: Orthurethra). *Zoological Journal of the Linnean Society*, **113**: 47–91.

TILLIER, S., MASSELOT, M., PHILIPPE, H. & TILLIER, A. 1992. Phylogènie moléculaire des Gastropoda (Mollusca) fondée sur le séquençage partiel de l'ARN ribosomique 28S. *Comptes Rendus de l'Academie des Sciences, Paris, Série 3*, **314**: 79–85.

TILLIER, S., MASSELOT, M. & TILLIER, A. 1996. Phylogenetic relationships of the pulmonate gastropods from rRNA sequences, and tempo and age of the stylommatophoran radiation. In: *Origin and evolutionary radiation of the Mollusca* (J.D. Taylor, ed.), 267–284. Oxford University Press, Oxford.

VAN BRUGGEN, A.C. 1995. Biodiversity of the Mollusca: time for a new approach. In: *Biodiversity and conservation of the Mollusca* (A.C. van Bruggen, S.M. Wells & T.C.M. Kemperman, eds), 1–19. Backhuys Publishers, Oegstgeest-Leiden.

VAN DE PEER, Y. & DE WACHTER, R. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for Microsoft Windows environment. *Computer Applications in the Biosciences*, **10**: 569–570.

WADE, C.M. & MORDAN, P.B. 2000. Evolution within the gastropod molluscs; using the ribosomal RNA gene-cluster as an indicator of phylogenetic relationships. *Journal of Molluscan Studies*, **66**: 565–570.

WADE, C.M., MORDAN, P.B. & CLARKE, B. 2001. A phylogeny of the land snails (Gastropoda: Pulmonata). *Proceedings of the Royal Society London, Series B*, **268**: 413–422.

WATERBORG, J.H. & ROBERTSON, A.J. 1996. Common features of analogous replacement histone H3 genes in animals and plants. *Journal of Molecular Evolution*, **43**: 194–206.

WELLS, D. & MCBRIDE, C. 1989. A comprehensive compilation and alignment of histones and histone genes. *Nucleic Acids Research*, **17** (Suppl.): r311–r346.

YOON, S.H. & KIM, W. 2000. Phylogeny of some gastropod mollusks derived from 18S rDNA sequences with emphasis on the Euthyneura. *Nautilus*, **114**: 84–92.