Common Phylogenetic Origin of Protamine-like (PL) Proteins and Histone H1: Evidence from Bivalve PL Genes

José M. Eirín-López,†[1] John D. Lewis,*[1] Le Ann Howe,*[2] and Juan Ausió**

*Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada; and †Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, A Coruña, Spain

Sperm nuclear basic proteins (SNBPs) can be grouped into three main categories: histone (H) type, protamine (P) type, and protamine-like (PL) type. Protamine-like SNBPs represent the most structurally heterogeneous group, consisting of basic proteins which are rich in both lysine and arginine amino acids. The PL proteins replace most of the histones during spermiogenesis but to a lesser extent than the proteins of the P type. In most instances, PLs coexist in the mature sperm with a full histone complement. The replacement of histones by protamines in the mature sperm is a characteristic feature presented by those taxa located at the uppermost evolutionary branches of protostome and deuterostome evolution, while the histone type of SNBPs is predominantly found in the sperm of taxa which arose early in metazoan evolution; giving rise to the hypothesis that protamines may have evolved through a PL type intermediate from a primitive histone ancestor. The structural similarities observed between PL and H1 proteins, which were first described in bivalve molluscs, provide a unique insight into the evolutionary mechanisms underlying SNBP evolution. Although the evolution of SNBPs has been exhaustively analyzed in the last 10 years, the origin of PLs in relation to the evolution of the histone H1 family still remains obscure. In this work, we present the first complete gene sequence for two of these genes (PL-III and PL-II/PL-IV) in the mussel *Mytilus* and analyze the protein evolution of histone H1 and SNBPs, and we provide evidence that indicates that H1 histones and PLs are the direct descendants of an ancient group of "orphon" H1 replication-dependent histones which were excluded to solitary genomic regions as early in metazoan evolution as before the differentiation of bilaterians. While the replication-independent H1 lineage evolved following a birth-and-death process, the SNBP lineage has been subject to a purifying process that shifted toward adaptive selection at the time of the differentiation of arginine-rich Ps.

Introduction

Sperm nuclear basic proteins (SNBPs) can be grouped into three major types: histone (H type), protamine (P type), and protamine-like (PL type) (Ausió 1995). These types of SNBP are widespread within different phylogenetic groups throughout the animal kingdom (Ausió 1995; Saperas et al. 1997). The replacement of histones by protamines in the mature sperm is a characteristic feature presented by taxa preferentially located at the uppermost evolutionary branches of protostome and deuterostome evolution (Ausió 1999), while the histone type of SNBPs is found in the sperm of taxa differentiated early in metazoan evolution such as in the sponge *Neofibularia* (Ausió et al. 1997). Despite early suggestions that protamines may have had a sporadic retroviral origin (Jankowski, States, and Dixon 1986), plenty of evidence has shown that the phylogenetic SNBP distribution is not random (Saperas et al. 1994). We have hypothesized, based on the experimental evidence from our laboratory, that protamines have evolved through a PL type intermediate from a primitive histone ancestor (Ausió et al. 1999; Eirín-López, Frehlick, and Ausió 2005).

The structural similarities between PL and H1 proteins pose very interesting questions regarding the evolutionary mechanisms to which these two groups of proteins are subject. On one hand, it has been shown that the long-term evolution of H1 histones is best described by a birth-and-death process under strong purifying selection rather than by concerted evolution (Eirín-López et al. 2004*a*). The differentiation between the replication-dependent (RD) and the replication-independent (RI) H1 lineages can be traced back to the transposition of an "orphon" group of H1 genes to a solitary genomic location early in metazoan evolution (Eirín-López et al. 2004*a*; Eirín-López et al. 2005). The subsequent evolution of both lineages led to the diversification observed inside the H1 family. On the other hand, the differentiation of SNBPs of the PL type must have also occurred early in metazoan evolution as they are present in both diploblastic and triploblastic (bilaterians) animals (Ausió 1999). The lysine to arginine transition leading to the differentiation of protamines from a PL precursor was a critical step in SNBP evolution, resulting in a strong positive selection process favoring the high arginine content of these proteins (Ausió 1999; Eirín-López, Frehlick, and Ausió 2005), making protamines one of the fastest evolving groups in nature (Oliva and Dixon 1991; Oliva 1995; Lewis et al. 2003).

Although the evolution of SNBPs has been exhaustively analyzed in the last 10 years (Ausió 1999; Lewis et al. 2004*b*; Eirín-López, Frehlick, and Ausió 2005), little attention has been paid to the origin of SNBPs in relation to the evolution of the histone H1 family, especially as it pertains to the differentiation and diversification of PL proteins, the intermediates between histones and protamines. The bivalve molluscs provide a unique opportunity to study this evolutionary progression as different species of this phylum can be taken as representative examples of all three types of SNBP (Subirana et al. 1973; Ausió 1986; Gimenez-Bonafe et al. 2002). The PL-I protein from the bivalve mollusc *Spisula* shows a high degree of homology with histone H1 (fig. 1), including a trypsin-resistant globular region containing a winged helix motif flanked by two

[1] These authors have contributed equally to this work.

[2] Present address: Department of Biochemistry and Molecular Biology, University of British Columbia, Life Sciences Centre, Vancouver, British Columbia, Canada.

Key words: birth-and-death evolution, adaptive evolution, histone H1, protamines, metazoans, winged helix domain.

E-mail: jausio@uvic.ca.

Mol. Biol. Evol. 23(6):1304–1317. 2006
doi:10.1093/molbev/msk021
Advance Access publication April 13, 2006

nonstructured terminal tails (Lewis et al. 2003). In the case of *Mytilus*, SNBPs consist of three major sperm-specific proteins: PL-II, PL-III, and PL-IV (fig. 1). These proteins replace much of the histone complement during spermiogenesis but coexist with approximately 20% of the somatic-type histones in the mature sperm (Ausió 1986; Lewis and Ausió 2002). The PL-II protein of *Mytilus* is a member of the histone H1 family, containing a conserved globular core of 84 amino acid residues that has a high similarity to both the winged helix motif of histone H1 and to the core (also a winged helix motif) of the chromatin-condensing histone H5 of chicken erythrocyte nuclei (Jutglar, Borrell, and Ausió 1991). The PL-IV protein is very small (6,500 kDa) and has a very similar composition to the lysine-rich C-terminal tail of histone H1 (Phelan et al. 1974). Indeed, PL-IV is a product of the posttranslational cleavage of a PL-II/PL-IV precursor (Carlos et al. 1993). Of the three SNBPs of *Mytilus* sperm, PL-III is present in the highest amount (Lewis and Ausió 2002). Its highly basic composition, rich in both lysine (27.5% mol/mol) and arginine (22.5% mol/mol), is intermediate to that of histones and protamines, but like protamines, it lacks any specific secondary structure in vitro (Rocchini, Rice, and Ausió 1995). If PL-III is an independent gene product under the control of an autonomous promoter, it most likely represents the initial genetic segregation of the N-terminal tail of a histone H1-like SNBP toward a protamine-like configuration.

Although PL representatives were first described in molluscs (Subirana et al. 1973; Ausió 1986), these SNBPs have also been found in many other groups including echinoderms, tunicates, and vertebrates (Ausió 1995; Chiva et al. 1995; Eirín-López, Frehlick, and Ausió 2005). In echinoderms, a PL-III–like protein called ϕ_0 has been identified (Subirana 1970) that has an amino acid sequence similar to that of the C-terminal domain of histone H1 (Prats and Cornudella 1995). In this work, we have isolated the genes for both the PL-II/PL-IV and PL-III SNBPs of *Mytilus californianus*. We have subsequently analyzed and compared the long-term evolution of SNBPs with that of histone H1, finding that both protein lineages are the direct descendants of a group of orphon H1 RD histones excluded from the main repetitive units to solitary regions in the genome early in metazoan evolution. While the RI H1 lineage evolved following a birth-and-death process, the SNBP lineage has been subject to a purifying process that shifted toward adaptive selection at the time of the differentiation of arginine-rich protamines.

Materials and Methods

Living Organisms

Specimens of *M. californianus* were collected by the authors from Point No Point (Sooke) on Vancouver Island as a part of the Science Venture Student Program.

Protein Preparation, Fractionation, and Electrophoresis

SNBPs were routinely extracted with 0.4 N HCl following the procedures described previously (Subirana and Colom 1987). Reverse phase high-performance liquid chromatography was performed on a 5-mm Vydac C18 column (25 × 3 × 0.46 cm) with 0.1% trifluoroacetic acid and eluted with varying acetonitrile gradients (Ausió 1988). Acetic acid (5%)–urea (2.5 M) polyacrylamide gels were prepared as described by Jutglar, Borrell, and Ausió (1991).

DNA Extraction and Genomic Library Construction and Screening

DNA was extracted from gonadal tissue (0.1 g) according to the protocol described by Sambrook, Fritsch, and Maniatis (1989). A *Bam*H1-digested genomic library of *M. californianus* was constructed using the Lambda ZAP II genomic library kit from Stratagene (La Jolla, Calif.). Plaques were screened using the *Mytilus trossulus* PL-II cDNA (GenBank accession number L02876; Carlos et al. 1993) as a probe (612 bp) labeled by nick translation. Hybridization was performed according to the membrane manufacturer's instructions, and these were exposed for 24 h and visualized using the PhosphorImager System (Molecular Dynamics, B & L Systems, Zoetermeer, the Netherlands). Positive clones were subcloned into pBR322, and the DNA was sequenced by the dideoxynucleotide method (Sanger, Nicklen, and Coulson 1977) using a Sequenase 2.0 kit (USB Corp, Cleveland, Ohio).

Degenerate PCR, Inverse PCR, and Genomic Walking

Degenerate primers for polymerase chain reaction (PCR) were created based on the complete amino acid sequence of PL-III from *M. trossulus* (Rocchini, Rice, and Ausió 1995). PCR was performed using the PCR Sprint thermal cycler (Interscience, Markham, Ohio) with genomic DNA as template. A touchdown profile was used for the amplification, with the annealing temperature decreasing from 65°C to 45°C over 20 cycles, followed by 10 cycles at 45°C.

Inverse PCR was carried out as described by Benkel and Fong (1996), using the primers MYTINV-F (5′-GTCCTCATCACCAAAGAAAAGGAG-3′) and MYTINV-R (5′-CTTTCCCCTTCTTGGGGTCTTGGAAC-3′). Southern analysis was used to locate positive clones. Genomic walking was performed on *Mytilus* DNA using adaptors, adaptor primers, and protocols based on Zhang and Gurr (2000). DNA was digested overnight with *Spe*I, *Nhe*I, and *Xba*I (New England Biolabs, Pickering, Ontario, Canada), adaptors were ligated at 16°C for 6 h, and PCRs were carried out using the adaptor-specific PCR primer PP1 and the gene-specific primer MYTWKF1 (5′-CAGCCTCCTCCCCCGGAAAGGCAGC-3′). A 1/40 dilution was made of the products of the first reaction, and 1 μl of this was added to a second PCR using the nested adaptor-specific PCR primer PP2 and the gene-specific primer MYTWKF2 (5′-CCAAAGAAAAGGAGGTCTGCTGGAAAG-3′). Stratagene's Herculase Enhanced DNA polymerase and buffer system were utilized for the PCRs. A hot-start and touchdown profile was used for each amplification, exactly as in Zhang and Gurr (2000).

Southern Blot Analysis of Inverse PCR Products, Cloning, and DNA Sequencing

Half of each inverse PCR was loaded onto a 1.0% agarose gel containing ethidium bromide and visualized under UV. The gel was blotted onto Zeta-Probe GT (BioRad,

Mississauga, Ontario, Canada) using the VacuGene XL Vacuum Blotting System (Pharmacia Biotech, Québec City, Québec, Canada) following each manufacturer's instructions. The double-stranded 252-bp insert was labeled by nick translation according to Sambrook, Fritsch, and Maniatis (1989). Blots were exposed for 24 h and visualized using the PhosphorImager System (Molecular Dynamics). PCR products were purified using Wizard PCR Preps DNA Purification System (Promega, Madison, Wisc.). The purified PCR products were then cloned into pCR 2.1-TOPO vector (Invitrogen, Burlington, Ontario, Canada) following the manufacturer's instructions and transformed into TOP10 competent cells (Invitrogen). DNA was sequenced by the dideoxynucleotide method using a Sequenase 2.0 kit (USB Corp).

Evolutionary Analyses

We have included in our analyses a total of 206 amino acid sequences (see Supplementary Table, Supplementary Material online), including 91 nonredundant histone H1 somatic sequences (68 RD, 23 RI), 17 testis-specific H1 sequences, and 97 sequences for SNBPs (3 histone, 18 protamine-like, and 76 protamine sequences). Given that there are no less than 12 different nomenclatures for H1 genes, the nomenclature was adapted to that of Doenecke (Albig, Meergans, and Doenecke 1997). Multiple alignments of the amino acid sequences were conducted using the BioEdit (Hall 1999) and Clustal_X (Thompson et al. 1997) programs with the default parameters given by each program. Alignments were checked for errors by visual inspection. The alignment of the different histone H1 domains (C-terminal, globular, and N-terminal) were performed following the criteria previously established by Ramakrishnan et al. (1993) and E. Schulze and B. Schulze (1995), defining the borders of the H1 central domain (see Supplementary Alignments 1–4, Supplementary Material online).

All molecular evolutionary analyses were conducted using the computer program MEGA version 3.1 (Kumar, Tamura, and Nei 2004). The extent of the amino acid and nucleotide sequence divergence was estimated by means of the uncorrected differences (p-distance) as this approach is known to give better results when the number of positions used is relatively small due to its smaller variance (Nei and Kumar 2000). The numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site were also computed using the modified method of Nei and Gojobori (Zhang, Rosenberg, and Nei 1998), providing in both cases the transition/transversion ratio (R). Distances were calculated using the complete-deletion option in the case of amino acid phylogenies, with the exception of the complete phylogeny shown in figure 3 where the pairwise-deletion option was used, as well as for the implementation of selection tests. In both cases, the standard errors were estimated by using the bootstrap method. The presence of selection was tested in PL SNBPs by the codon-based Z-test for selection (H_1: $p_N < p_S$; Nei and Kumar 2000) and by the codon-based Fisher's exact test (H_1: $p_N > p_S$; Zhang, Kumar, and Nei 1997), being H_0: $p_S = p_N$ in both cases. The probability that the null hypothesis is rejected is indicated as $^{**}P$ ($P < 0.001$) and $^{*}P$ ($P < 0.05$).

The minimum-evolution tree-building method (Rzhetsky and Nei 1992) was used to reconstruct the phylogenetic trees, and in order to assess that our results are not dependent on this choice, phylogenetic analyses were completed by reconstructing maximum parsimony trees using the close-neighbor-interchange tree search method (see Supplementary Figures 1–4, Supplementary Material online) with search level 1 and with 10 replications for the random addition trees option. The reliability of the resulting topologies was tested by both the bootstrap and the interior-branch test methods (Felsestein 1985; Rzhetsky and Nei 1992; Sitnikova 1996), producing the bootstrap (BP) and confidence probability (CP) values, respectively, for each interior branch of the trees after 1,000 replicates. Given the known conservative nature of the bootstrap method, BP > 80% was interpreted as high statistical support for groups, whereas CP = 95% was considered statistically significant (Sitnikova, Rzhetsky, and Nei 1995). We rooted phylogenetic trees using the H1 gene of the protist *Entamoeba histolytica* as it represents one of the most primitive eukaryotes for which a H1-related protein has been characterized (Kasinsky et al. 2001).

Results

Characterization of the Genes PL-III and PL-II/PL-IV from *M. californianus*

Using genomic DNA of *M. californianus* as a template, we were successful in the amplification of a 252-bp portion of the PL-III–coding region (fig. 2*A*, lane c), which was subsequently used to design nondegenerate PCR primers for use with an inverse PCR methodology to obtain the remaining flanking sequences. Southern analyses were used to screen our PCR products for those containing our gene of interest identifying a 533-bp inverse PCR product (fig. 2*A*, lane 2). This clone contained the entire coding region of PL-III in the 5′ direction plus an additional 314 bp of upstream nucleotide sequence.

A genomic walking technique was employed to amplify the 3′ end of the *Mytilus* PL-III gene, resulting in a 261-bp fragment (fig. 2*B*) that contained the elusive remainder of the PL-III–coding region and a further 162 bp of downstream nucleotide sequence. In total, the length of the cloned region of the PL-III gene was 790 bp, which encoded a single open reading frame of 102 amino acid residues (fig. 2*C*). The protein possesses a number of distinct features, including the presence of a conserved SR repeat domain that is characteristic of many PL proteins and mammalian protamines (fig. 2*C*, solid bars). There are a number of repetitive hexapeptide amino acid sequence motifs similar to but less conserved than those found in the PL-I of *Spisula solidissima*.

A positive subclone for the PL-II/IV SNBP of *M. californianus* was obtained from a genomic library, consisting of an 1,152-bp sequence containing a single open reading frame encoding 208 amino acid residues corresponding to the PL-II/IV precursor protein. There was an additional 254 bp of 5′ leading sequence and 274 bp of downstream nucleotide sequence (fig. 2*D*).

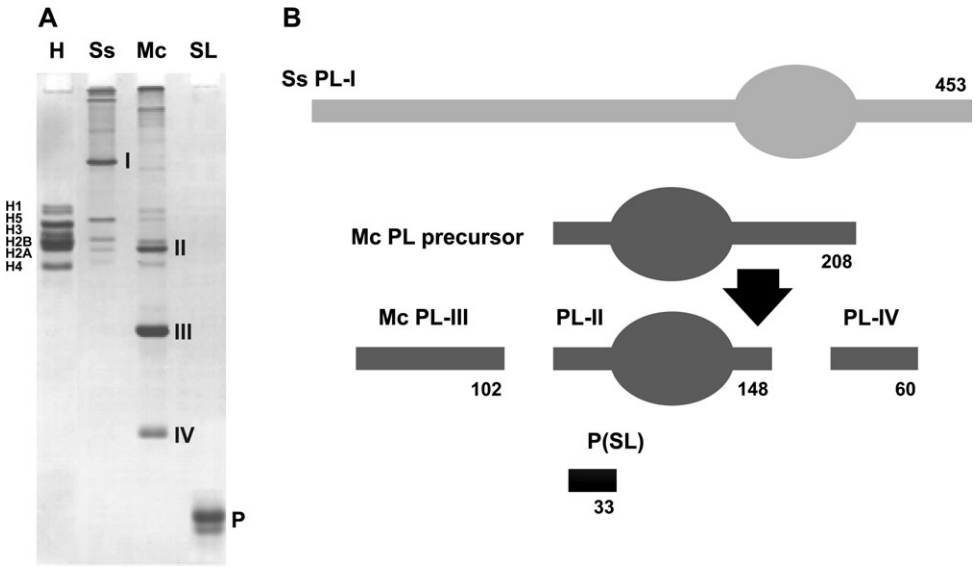


FIG. 1.—(*A*) Urea (2.5 M)-acetic acid (5%), polyacrylamide gel electrophoresis (15%) of 0.4 N HCl extracts from the sperm of: Ss, *Spisula solidissima* and Mc, *Mytilus californianus*. Salmon protamine, salmine (SL), and chicken erythrocyte histones (H) were used as markers. The roman numerals refer to the PL protein designations (PL-I, PL-II, PL-III, and PL-IV). (*B*) Schematic diagram of the PL proteins of *Spisula solidissima* (Ss), *M. californianus* (Mc), and the salmon protamine. Numbers indicate the protein length in amino acids. The posttranslational cleavage of Mc precursor to PL-II and PL-IV is shown. Evolution is proposed to have occurred from an H1-like protein (PL-I), through PL proteins (PL-II, PL-III, and PL-IV), to a protamine (P) (such as SL).

Relationships Between Histone H1 and SNBPs

In order to investigate the evolutionary relationships between SNBPs and histone H1 proteins, we reconstructed a phylogeny using H1, protamine-like (PL), and protamine (P) sequences from several representative animal taxonomic groups. The unrooted phylogeny shown in figure 3 shows the global relationships between histone H1 and SNBPs in different metazoan phyla, depicting a clear differentiation between H1 and protamines, with an intermediate position occupied by protamine-like SNBPs (see also Supplementary Alignment 1 and Supplementary Figure 1, Supplementary Material online). As previously reported (Eirín-López et al. 2004*a*), the different taxonomic groups for which H1 sequences were analyzed are well defined by the present topology, showing the monophyletic differentiation of the RI H1 lineage including orphon H1 genes from invertebrates and histones H1° and H5 from vertebrates. The functional evolution of H1 genes is also evident in the case of mammals, where H1.1-H1.5 and H1t variants cluster by type rather than by species. Although PL proteins occupy an intermediate evolutionary position, they are more closely related to H1 histones than to protamines.

The results reported in the previous section demonstrate for the very first time that the PL-III gene from *Mytilus* is an independently regulated gene which is uncoupled from PL-II/PL-IV. The PL-III gene encodes a protein which corresponds to the relatively short N-terminal region of a canonical histone H1 protein, and it displays a closer proximity to protamines in its phylogeny, most likely as a result of lacking the winged helix motif.

The functional clustering of protamines can be easily ascertained from this phylogenetic tree. In it, type 3 protamines are more closely related to the invertebrate PL-III and ϕ_0 than to any other protamine and show the highest rate of evolution among all H1 histones and SNBPs, as revealed by their longer branch lengths in the tree. Type 1 and 2 protamines of mammals seem to deviate from the previous group into a different functional clustering pattern, being more related between themselves than to any other vertebrate P1. This is in agreement with previous reports suggesting that, in mammals, type 2 protamines arose from a type 1 protamine precursor (Krawetz and Dixon 1988) but may also be due to the particular structure and evolutionary pattern shown by protamines in mammals.

Another shared feature between H1 histones and PLs is the presence of an evolutionary diversification process that runs parallel along the evolution of triploblastic animals. Histone H1 RD and RI subtypes are present in both protostomes and deuterostomes, suggesting that the differentiation between them took place before the protostome/deuterostome branching. With PLs, the same parallel evolutionary pattern is observed, the transition from PLs to protamines and their subsequent diversification takes place simultaneously in protostomes and deuterostomes, and thus, the origin of this protein transition also took place prior to the split between these two major groups of triploblasts (or bilaterians).

Phylogenetic Analysis of the Histone H1 Functional Domains

Given the close structural relationship between SNBPs of the PL type and different domains of the histone H1 molecule, the analysis of the evolutionary relationships between SNBPs and the N-terminal, C-terminal, and globular regions of histone H1 is of particular interest. The phylogeny

reconstructed considering the region consisting only of the N-terminal and the globular regions of H1 and PL proteins is shown in figure 4*A* (see also Supplementary Alignment 2 and Supplementary Figure 2, Supplementary Material online). The *Mytilus* PL-IV protein corresponding to a C-terminal H1 segment was excluded from this analysis.

The resulting topology is very similar to that obtained with the whole molecule, indicating that both the N-terminal and the globular domain are important determinants of the identity of these proteins within different taxonomic groups. In this analysis, the RI lineage shows a monophyletic origin, and it is precisely within this group that PL proteins are clustered. In contrast, *Mytilus* PL-III represents an exception, exhibiting a high extent of divergence from both H1 and PL proteins. This is most likely due to the absence of a globular domain in PL-III. Indeed, if the same analysis is carried out removing the globular domain (not shown), PL-III then clusters with the remaining PL proteins. The observed topology shown in figure 4*A* reveals not only a high identity between PL and H1 proteins but also a close and well-defined relationship of PLs with the H1 RI lineage.

The phylogenetic analysis using the globular and the C-terminal domains yields the topology that is shown in figure 4*B*. *Spisula*'s PL-I, PL-II/PL-IV, PL-II, and PL-IV from *Mytilus* and PL precursor from *Ostrea* have been included in this analysis (see also Supplementary Alignment 3 and Supplementary Figure 3, Supplementary Material online). The PL-III proteins with homology to the N-terminal region were excluded. Both RD and RI H1 proteins intermingle extensively in this phylogeny. Indeed, it was not possible to trace a monophyletic origin for the RI H1 lineage. As before, a phylogeny reconstructed considering exclusively the highly heterogeneous C-terminal domain is consistent with a monophyletic origin for the RI lineage.

Given that the globular domain of H1 plays an important role in determining the identity of its different subtypes, a phylogeny using only this protein region was lastly elaborated in order to avoid any interferences in the topology which could result from the high variability observed in the N- and C-terminal domains (see also Supplementary Alignment 4 and Supplementary Figure 4, Supplementary Material online). Only PL proteins containing globular domain were used in this analysis (PL-I from *Spisula*, PL-II/PL-IV from *Mytilus*, and PL precursor from *Ostrea*). The role of the globular domain, and by extension that of the winged helix structure, in providing an unequivocal "footprint" for the different H1 subtypes is evident from this phylogeny (fig. 4*C*). The topology matches almost identically the relationships established among H1 histones when using the complete protein sequences and places again PL proteins inside the RI H1 lineage. It is also interesting to note that the oocyte-specific B4/H1M maternal H1 histone from *Xenopus* is also included in the RI group close to the *Spisula* PL-I protein.

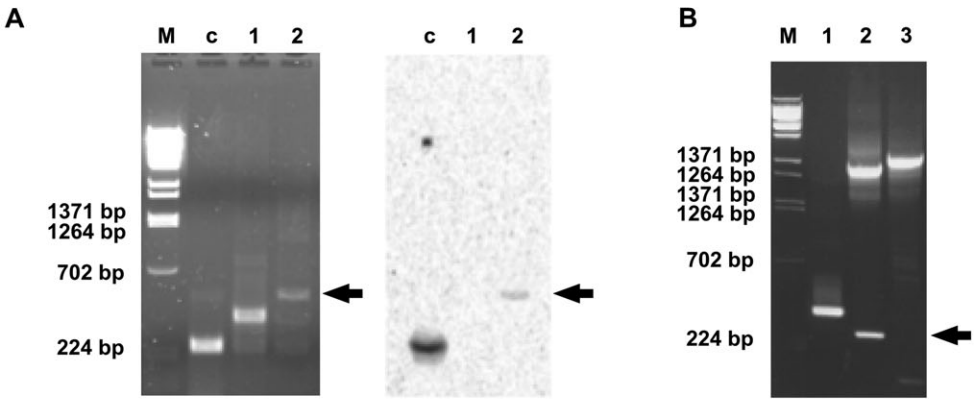
Upstream and Downstream Regulating Elements of PL Genes

Promoter regions from RI histone H1 genes exhibit characteristic and specific control elements different from those of the RD variants. The *Mytilus* SNBP genes characterized in this paper (PL-II/PL-IV and PL-III) were aligned with each other and to the PL-I gene of *Spisula* in order to assess their relatedness and to identify any conserved elements. Comparison of the gene structures of *Mytilus* PL-II/PL-IV and *Spisula* PL-I (fig. 5*A*) reveals an overall similarity of 47%, with several common features. The overall similarity of the upstream sequence of PL-II/PL-IV with that of PL-III from the same organism (fig. 5*B*) is only 30%. Approximately 40 bp upstream from the initiation codons of PL-I and PL-II/PL-IV, there is a conserved region of about 15 bp that may represent the binding site of a common or related regulatory factor, and it likely corresponds to the region of the transcription initiation site. An almost identical region is present also in the PL-III gene, indicating that this element may be important for the regulation of bivalve mollusc SNBPs genes during spermiogenesis. Another conserved region of the *Mytilus* PL-II/PL-IV and *Spisula* PL-I genes is the putative TATA box domain (fig. 5*A* and *B*), which consists of 10 of 12 identical nucleotides. Interestingly, an equivalent 16-bp conserved region is also found at the same location in PL-III, although a TATA box cannot be easily identified (fig. 5*A* and *B*). When compared with H1 promoters, this PL domain appears to replace the CAAT box of RD and the H4 box of RI which occur at approximately the same location.

The recent availability of the genome drafts of the sea urchin *Strongylocentrotus purpuratus* and the tunicate *Ciona intestinalis* made it possible to include in this analysis the proximal promoter regions of the sperm-specific histone H1 and the protamine-like P1 (P2), respectively (fig. 5*C*). The promoter regions of the protein ϕ_0 of *Holothuria* (a PL protein from echinoderms; Ruiz-Lara et al. 1993), the protamine of squid (Lewis et al. 2004*a*), and the protamine of *Drosophila* (Adams et al. 2000) were also included for comparison. The analysis shows the existence of a considerable extent of similarity between the promoters region of *Spisula* and *Mytilus* PL genes and those of the remaining SNBPs, providing further evidence for the evolutionary link between the different SNBP types.

Nucleotide Evolution of PL SNBPs

According to the vertical evolution hypothesis of SNBPs (Ausió 1995; Eirín-López, Frehlick, and Ausió 2005), PLs represent an intermediate evolutionary stage between histones and protamines. We have analyzed the numbers of synonymous and nonsynonymous nucleotide differences per site (p) among PL genes in order to assess whether positive or negative selection is acting among them (table 1). We found the highest synonymous divergence in the case of PL-I genes ($p_S = 0.484 \pm 0.022$), followed by PL-II/PL-IV ($p_S = 0.235 \pm 0.027$) and PL-III ($p_S = 0.087 \pm 0.034$). In all instances, the synonymous divergence exceeded the nonsynonymous variation, suggesting that a purifying process is involved. The presence of negative selection was tested by two different methods. First, we compared the numbers of synonymous (p_S) and nonsynonymous (p_N) substitutions using a codon-based Z-test for selection where H_1 was defined as $p_N < p_S$. Results in table 1 show that p_S is significantly greater than p_N in PL-coding regions ($P < 0.001$ for PL-I, PL-II/PL-IV,



C

```
cctagcttccaattaattctaggcgaatcttcatgtatttgaaccacatcaccaactttgatagtctgattattgtgtcctaaagccgtgtggtattccc
ttaatgatgtaagatattcatttttccatcttcgtcgaaaatgttgaataatgttgtgctgatcctgtagacgttgattcaagttgttatgtgtacttgg
ttcaaaattaatcaaattgttcatgtttgtagacggatacggtaaacatcttattctccggccataaagaagatgagaaggtgtcagtcagtcttgtatc
acccaacaagaagg

 M   P   S   P   T   R   R   S   S   K   S   R   S   K   S   R   S   R   S   R   S   A   S   S   P
ATG CCA AGC CCA ACT CGT CGT TCA TCC AAG TCC AGG TCC AAA AGC AGG AGC AGG AGC AGG TCA GCC TCC TCC CCC
 G   K   A   A   K   R   A   R   S   K   T   P   R   R   G   K   K   R   A   R   S   P   S   K   K
GGA AAG GCA GCA AAA CGT GCT CGT TCC AAG ACC CCA AGA AGG GGA AAG AAA AGG GCA AGG TCT CCA TCC AAA AAA
 A   R   R   R   S   R   S   T   K   K   T   A   A   K   R   R   K   R   S   S   S   P   K   K   R
GCA AGA AGG AGG TCT AGG TCT ACC AAG AAG ACA GCA GCT AAG AGG AGG AAG AGG TCC TCA TCA CCA AAG AAA AGG
 R   S   A   G   K   R   R   V   R   A   E   K   G   G   K   R   R   R   S   R   G   K   K   A   A
AGG TCT GCT GGA AAG AGG AGA GTA AGA GCA GAG AAA GGA GGA AAG AGA AGG AGG TCA AGG GGA AAG AAA GCC GCA
 A   K   K   *
GCA AAA AAA TGA

ctggtcgaaacacacagaacattaatgtcggagatgtagttcagattcatgacgatttatcactgagagtaaattgaactttggctgtaattgccaaatt
agtttacggcaaagatggaattacacgaccagccataattcgaaccaagaatggagcgacatct
```

D

```
ttgatgatgctacaaaattacttagagccagcatagtcctgtcttgtgtactcttctgcaagtgtattttattcggttacgagaaacataagccaagctg
ttcccaatgtgtagataatatataattcctcatcggctgattgtacatgtttctattggtgctctctatcacgtccgcgttagttacataaacaagcacc
tgtagacatcgttgtcattcttgtattttgacgtagtaaaagcaagtaaccaat

 M   P   S   P   S   R   R   S   R   S   R   S   R   S   R   S   K   S   P   K   R   S   P   A   K
ATG CCA AGC CCA AGT AGA CGT TCC AGA TCT AGG TCT AGG AGT AGG AGT AAA TCT CCA AAG AGA AGT CCA GCA AAG
 K   A   R   K   T   P   K   K   A   S   A   T   G   G   A   K   K   P   S   T   L   S   M   I   V
AAG GCA AGA AAG ACA CCA AAG AAA GCA AGC GCA ACG GGT GGA GCC AAG AAG CCA TCT ACT TTA TCC ATG ATT GTT
 A   A   I   Q   A   M   K   N   R   K   G   S   S   V   Q   A   I   R   K   Y   I   L   A   N   N
GCT GCC ATC CAA GCA ATG AAG AAC AGA AAG GGG TCT TCA GTC CAA GCT ATT AGA AAG TAC ATC CTG GCT AAC AAC
 K   G   I   N   T   S   H   L   G   S   A   M   K   L   A   F   A   K   G   L   K   S   G   V   F
AAA GGA ATC AAC ACA TCA CAC CTC GGA TCT GCA ATG AAA CTG GCT TTC GCA AAG GGA TTG AAA TCT GGT GTT TTC
 V   R   P   K   T   S   A   G   A   S   G   A   T   G   S   F   R   V   G   K   A   P   S   S   P
GTC AGA CCT AAA ACT TCC GCT GGT GCT TCT GGT GCA ACT GGT AGC TTC CGA GTT GGA AAA GCA CCT TCT TCT CCC
 K   K   K   A   K   K   A   K   S   P   K   K   K   S   S   K   K   S   K   N   K   S   N   N   A
AAG AAA AAG GCA AAG AAA GCA AAG TCA CCA AAA AAG AAG AGT TCC AAG AAA TCA AAG AAC AAA TCA AAC AAC GCT
 K   A   K   R   S   P   R   K   K   K   A   A   V   K   K   S   S   K   S   K   A   K   K   P   K
AAG GCT AAG AGG TCA CCC CGA AAG AAG AAA GCT GCA GTT AAA AAG TCA TCA AAG TCG AAG GCC AAA AAG CCA AAG
 S   P   K   K   K   K   A   A   K   K   P   A   R   K   S   P   K   K   K   A   R   K   S   P   K
TCT CCG AAG AAA AAG AAG GCT GCC AAG AAA CCC GCA AGA AAG TCT CCA AAG AAG AAA GCC AGA AAG TCT CCC AAG
 K   K   A   A   K   K   S   K   K   *
AAG AAG GCC GCC AAG AAG TCA AAG AAG TAG

acctggttagcagaaataactacacaatccaaagcctttcaggccacccaaatatttcaaaaatgtgtcttattttgttgtacttatcagtcaaggacgt
atattagttatacgtccttgtgttcttgtatcagtttatatcataatcatcaagctatgcctttattactttaattgagactacagttaatctttcccat
gattttcttgcacattttagaataataattattttcacttctaggtttacccctagtccctatatgtc
```

FIG. 2.—(*A*) Left side, 1.0% agarose gel showing the inverse PCR results. Lane c, PL-III 252-bp control fragment; lane 1, *Eco*RI-digested iPCR products; and lane 2, *Dde*I-digested iPCR products. The marker is *Bst*E-digested Lambda (λ) DNA (Gibco, Burlington, Ontario, Canada). Right side, Southern blot of the gel on the left, probed with the 252-bp PL-III fragment. Black arrows indicate the positive product. (*B*) Agarose gel (1.0%) of a 3′ genome walking experiment. Again, the positive fragment is denoted by a black arrow. Lane 1, *Nhe*I; lane 2, *Spe*I; and lane 3, *Xba*I-digested genomic DNA. (*C*) Complete sequence of the PL-III gene from *Mytilus californianus* (GenBank accession number DQ305039). The solid bar denotes the conserved RS domain. Conserved hexapeptide repeats are highlighted by dashed boxes. The total length is 790 bp. (*D*) Complete sequence of the PL-II/PL-IV gene from *M. californianus* (GenBank accession number DQ305038). Solid bars denote the conserved RS domain. The H1-like globular winged helix region is highlighted by a box. The dashed bar indicates conserved pentapeptide involved in posttranslational cleavage, this peptide is retained in the mature protein (Carlos et al. 1993). The arrow denotes the cleavage site.

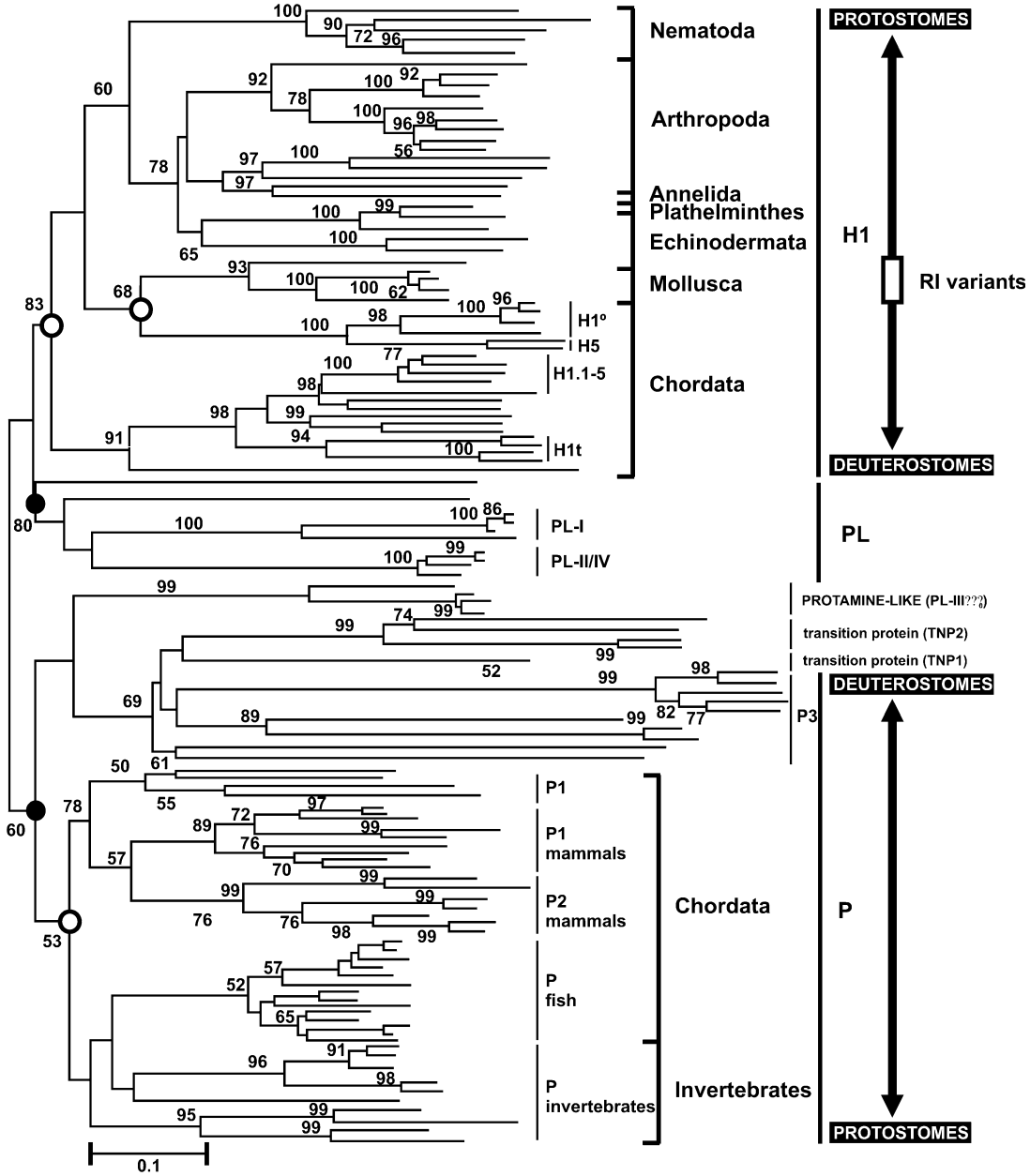


FIG. 3.—Phylogenetic relationships between histone H1 and SNBPs reconstructed from the alignment of complete amino acid sequences using uncorrected *p*-distances. The numbers for interior branches in the unrooted topology represent BP values based on 1,000 replications and only shown when a value is greater than 50%. The monophyletic origin of H1 histones of the RI H1 lineage and that of protamines is indicated by open circles, while the polyphyletic origin of PL proteins is indicated by filled circles. Taxonomic groups, as well as the different histone and SNBP subtypes, are indicated on the right side of the figure.

and for the overall mean; $P < 0.05$ for PL-III), supporting the notion that these proteins are also subject to purifying selection. It could be argued that, as with PL-II/IV and PL-III, the number of codons is relatively small, and this could lead the *Z*-test to be too liberal in rejecting the null hypothesis H_0: $p_N = p_S$. To this end, we have implemented a one-tailed codon-based Fisher's exact tests for each of the three PLs in order to confirm our results. Table 1 shows that neutral evolution can also be rejected in this case ($P < 0.001$ for PL-I and PL-II/PL-IV; $P < 0.05$ for PL-III) and given that significantly greater synonymous divergences are observed, this provides further support to the notion of purifying selection.

Discussion

PL Proteins in Molluscs and the Evolution of SNBPs

H type SNBPs are compositionally and structurally related to the histones that are found in the nuclei of somatic

cells, whereas protamine type SNBPs consists of highly specialized and highly basic arginine-rich proteins of relatively small molecular mass. The PL group is the most structurally heterogeneous group among SNBPs, consisting of basic proteins enriched in both lysine and arginine (Ausió 1995).

In the present work, we have isolated and characterized the complete sequence of the genes of the SNBPs of the mollusc *M. californianus* (fig. 1), including the PL-II/PL-IV gene and the PL-III gene. From an evolutionary standpoint, this taxonomic group is of critical interest because different species of molluscs can be taken as a representative examples of all three types of SNBP (Subirana et al. 1973; Ausió 1986; Gimenez-Bonafe et al. 2002) and also because it represents the only protostome phylum where RI histone H1 proteins have been described (Eirín-López et al. 2002; Eirín-López et al. 2004*b*).

Our results indicate that PL-II/PL-IV and the PL-III SNBPs of *Mytilus* are encoded by independent genes. While the PL-III protein bears similarity with the N-terminal tail of H1, the PL-II/PL-IV protein is posttranslationally cleaved to yield two different peptides, PL-II and PL-IV, which bear sequence similarity with the globular region and the C-terminal domain of a canonical H1, respectively. Although the presence of an intervening sequence in the PL-III–coding region from the mussel *M. edulis* had been previously described (Ruiz-Lara et al. 1993), we did not find evidence for the presence of such an intron in the genomic PL-III sequence from *M. californianus*. Rather than this being representative of a high divergence between both mussel species, we believe that the discrepancy could be the result of the well-known association of PL-III with hypervariable regions in the genome, undergoing frequent gene conversion and unequal recombination events (Heath and Hilbish 1998).

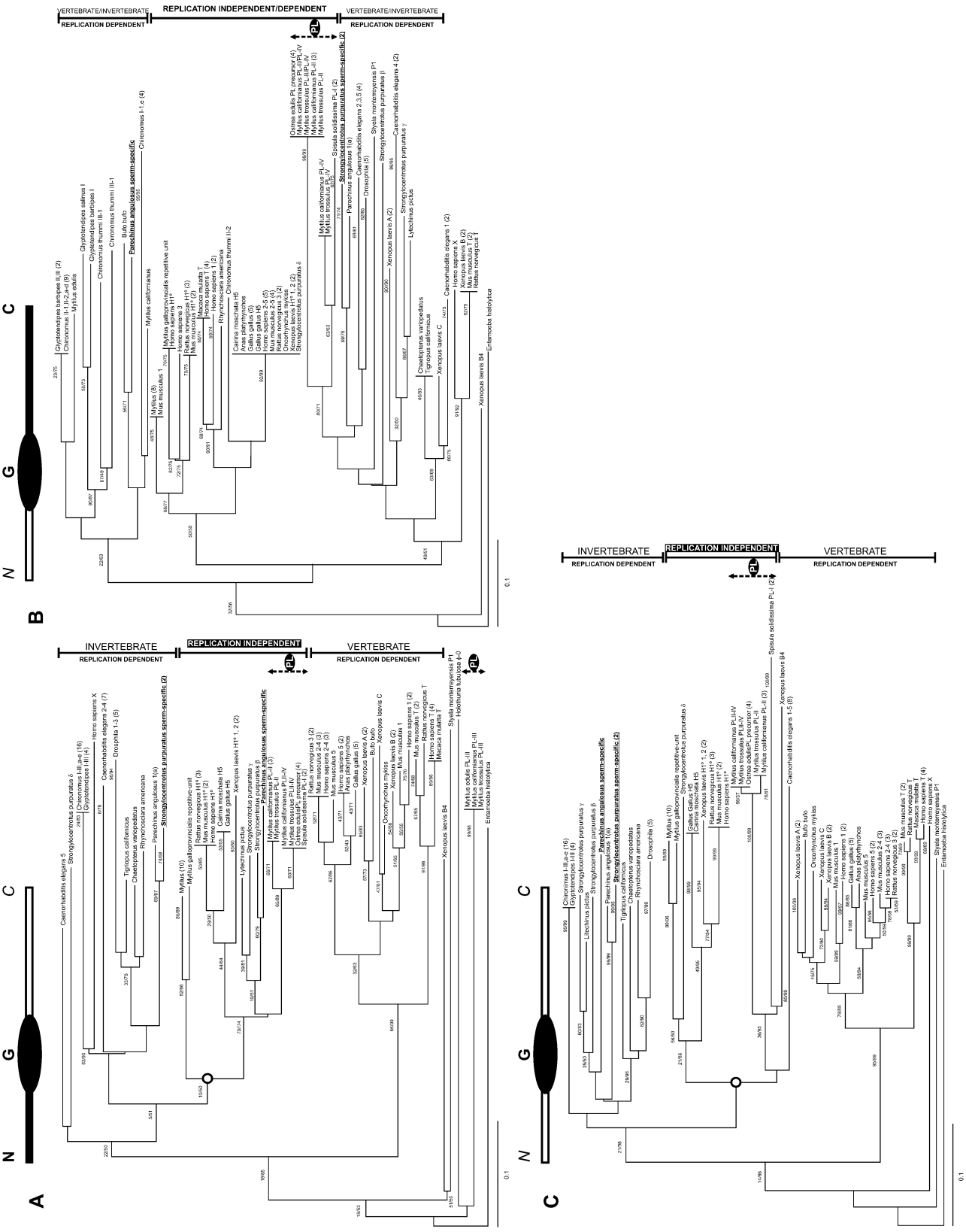
The analysis of PL-II/PL-IV and PL-III promoter regions shows the presence of distinct control elements involved in their gene expression regulation (fig. 5*A* and *B*). Although PL-III apparently lacks a canonical TATA box which is present in PL-II/PL-IV and also in the promoter region of *Spisula* PL-I (fig. 5*C*), it shares with them a 16-bp conserved element within the same region. This sequence is probably involved in the stage-specific developmental expression of these genes during spermiogenesis.

Interestingly, the three PL proteins of *Mytilus* show homologies with different structural domains of the PL-I SNBP from *Spisula* (Ausió and van Holde 1987) in the same domain-based fashion as they do with histone H1 (fig. 1). Whether the relationship between the PL-II/PL-IV and PL-III genes of *Mytilus* and PL-I gene from *Spisula* is the result of a segregation process from an ancestor PL-I precursor in *Mytilus* or whether *Spisula*'s PL-I has been the result of the fusion of PL-II/PL-IV and PL-III ancestor genes still remains unclear. However, a gene fusion event is less likely than a gene segregation event in evolutionary terms. In addition, the segregation hypothesis would be consistent with the vertical evolution of SNBPs (Ausió 1999; Eirín-López, Frehlick, and Ausió 2005), in which the gradual segregation of either a N- or C-terminal domain of H1 histones would have ultimately led to the differentiation of protamines through PL intermediates, which subsequently underwent a lysine to arginine transition (Lewis et al. 2004*b*). Comparison of the nucleotide-coding sequences of the different mollusc PL groups shows, as would be expected, that the lowest divergences are found in PL-II/PL-IV and PL-III. This could be taken as an indication of the recent divergence of these genes (not enough time has elapsed to accumulate a high nucleotide variation) and is in agreement with their segregation from an ancestor PL-I precursor, which otherwise exhibited high substitution numbers, as in the case of *Spisula*.

Further support for the segregation hypothesis can be ascertained from the phylogeny shown in figure 3, whose topology places PL-III closer to protamines than to any other PL protein, in agreement with the vertical SNBP evolution hypothesis as shown in figure 4*A–C*. While PL-II/PL-IV/PL-I–related SNBPs underwent a lysine to arginine transition and an additional segregation of the C-terminal domain that resulted in protamines in some groups of deuterostomes (Ausió et al. 1999; Lewis et al. 2004*b*), the role and/or relation of PL-III in this process, beyond molluscs, remains yet to be established.

RI H1 Histones and the Evolution of PL Proteins

The homology shared by H1 and PLs extends beyond the structural level, especially when evolutionary and functional considerations are further taken into account. The histone hypothesis for the vertical evolution of SNBPs predicts that only H or PL-precursor SNBP types would be present in those taxa that arose early in metazoan evolution, whereas the more specialized PL and P types would represent a characteristic feature of those taxa located at the uppermost evolutionary branches of bilaterian (protostomes and deuterostomes) evolution. This assumption is clearly supported by the analysis shown in figure 3, which implicitly shows that the differentiation of the PL lineage occurred before the differentiation between diploblastic and triploblastic animals, followed by a parallel vertical "mode" of evolution defined as H $\leftrightarrow$ PL $\leftrightarrow$ P during the evolution of bilaterians (Ausió et al. 1999; Eirín-López, Frehlick, and Ausió 2005). The functional clustering of protamines is also supported by this phylogenetic tree, and in fact, all these proteins also cluster well with vertebrate transition proteins (TNP), a set of protamines that precede the incorporation of protamines to sperm chromatin during the histone to protamine transition (Meistrich 1989). Such grouping provides support to the notion that PL and TNP proteins may be ontogenetically related (Ausió 1995). On the other hand, the close relationship observed between type 3 protamines and transition proteins further supports their common evolutionary origin arising from the ancient duplication of a domain containing type 1 and type 2 protamine genes before the radiation of rodents, mammals, and Artiodactyla (Kramer and Krawetz 1998). It is important to note that protamine 3 is a relatively recently described new type of mammalian protamine whose gene is situated between protamine 2 and transition protein 2 in the mammalian protamine gene cluster and whose substantial amount of aspartic acid suggest that protamine 3 is not likely to be a DNA-binding protein (Kramer and Krawetz 1998).


A
B
C
N
G
C
INVERTEBRATE
REPLICATION DEPENDENT
REPLICATION INDEPENDENT
VERTEBRATE
REPLICATION DEPENDENT
REPLICATION INDEPENDENT/DEPENDENT
VERTEBRATE/INVERTEBRATE
REPLICATION DEPENDENT
VERTEBRATE/INVERTEBRATE
REPLICATION DEPENDENT
INVERTEBRATE
REPLICATION DEPENDENT
REPLICATION INDEPENDENT
VERTEBRATE
REPLICATION DEPENDENT
Caenorhabditis elegans 5
Strongylocentrotus purpuratus δ
Chironomus I-III-a-c (16)
Glyptotendipes I-III (4)
Homo sapiens X
Caenorhabditis elegans 2-4 (7)
Drosophila 1-3 (5)
Tigriopus californicus
Chaetopterus variopedatus
Rhynchosciara americana
Parechinus angulosus 1(a)
Strongylocentrotus purpuratus sperm-specific (2)
Mytilus (10)
Mytilus galloprovincialis repetitive-unit
Rattus norvegicus H1° (3)
Mus musculus H1° (2)
Homo sapiens H1°
Carinia mnochata H5
Gallus gallus H5
Lytechinus pictus
Strongylocentrotus purpuratus β
Strongylocentrotus purpuratus sperm-specific
Mytilus californianus PL-I (3)
Mytilus trossulus PL-II
Mytilus californianus PL-I/II-V
Mytilus trossulus PL-I/II-V
Ostrea edulis PL precursor (4)
Spisula solidissima PL-I (2)
Rattus norvegicus 3 (2)
Mus musculus 2-4 (3)
Mus musculus 5
Homo sapiens 5 (2)
Homo sapiens 2-4 (3)
Gallus gallus (5)
Xenopus laevis A (2)
Bufo bufo
Oncomythus mykiss
Xenopus laevis C
Xenopus laevis B (2)
Mus musculus 1
Homo sapiens 1 (2)
Mus musculus T (2)
Rattus norvegicus T
Macaca mulatta T
Xenopus laevis B4
Styela montereyensis P1
Holothuria tubulosa γ-δ
Mytilus edulis PL-III
Mytilus californianus PL-III
Mytilus trossulus PL-III
Entamoeba histolytica
0.1
Glyptotendipes barbipes II,III (2)
Chironomus I-II-2,a-d (9)
Mytilus edulis
Glyptotendipes salinus I
Glyptotendipes barbipes I
Chironomus thummi III-1
Bufo bufo
Parechinus angulosus sperm-specific
Chironomus I-1 α (4)
Mytilus californianus
Mytilus (8)
Mus musculus 1
Mytilus galloprovincialis repetitive unit
Homo sapiens H1°
Rattus norvegicus H1° (3)
Mus musculus H1° (2)
Homo sapiens T (4)
Macaca mulatta T
Rhynchosciara americana
Chironomus thummi III-2
Carinia mnochata H5
Mus musculus H5
Gallus gallus (5)
Gallus gallus H5
Homo sapiens 2-5 (5)
Rattus norvegicus 3 (2)
Oncomythus mykiss
Xenopus laevis H1° 1, 2 (2)
Strongylocentrotus purpuratus δ
Ostrea edulis PL precursor (4)
Mytilus californianus PL-I/II-V
Mytilus trossulus PL-I/II-V
Mytilus californianus PL-I (3)
Mytilus californianus PL-I/II-V
Mytilus trossulus PL-I/II-V
Spisula solidissima PL-I (2)
Strongylocentrotus purpuratus sperm-specific (2)
Parechinus angulosus 1(a)
Drosophila (5)
Caenorhabditis elegans 2,3,5 (4)
Styela montereyensis P1
Strongylocentrotus purpuratus β
Caenorhabditis elegans 4 (2)
Strongylocentrotus purpuratus γ
Lytechinus pictus
Chaetopterus variopedatus
Tigriopus californicus
Xenopus laevis C
Caenorhabditis elegans 1 (2)
Homo sapiens δ
Xenopus laevis B (2)
Mus musculus T (2)
Rattus norvegicus T
Xenopus laevis B4
Entamoeba histolytica
0.1
Chironomus I-III-a-c (16)
Glyptotendipes I-III (4)
Strongylocentrotus purpuratus γ
Strongylocentrotus purpuratus β
Parechinus angulosus 1(a)
Strongylocentrotus purpuratus sperm-specific (2)
Tigriopus californicus
Chaetopterus variopedatus
Rhynchosciara americana
Drosophila (5)
Mytilus (10)
Mytilus galloprovincialis repetitive-unit
Gallus gallus H5
Carinia mnochata H5
Xenopus laevis H1° 1, 2 (2)
Rattus norvegicus H1° (3)
Mus musculus H1° (2)
Homo sapiens H1°
Mytilus californianus PL-I/II-V
Mytilus trossulus PL-I/II-V
Ostrea edulis PL precursor (4)
Mytilus trossulus PL-I/II
Mytilus californianus PL-I (3)
Spisula solidissima PL-I (2)
Xenopus laevis B4
Caenorhabditis elegans 1-5 (6)
Xenopus laevis A (2)
Bufo bufo
Oncomythus mykiss
Xenopus laevis C
Xenopus laevis B (2)
Mus musculus 1
Homo sapiens 1 (2)
Gallus gallus (5)
Anas platyrhynchos
Mus musculus 5
Mus musculus 2-4 (3)
Homo sapiens 2-4 (3)
Rattus norvegicus 3 (2)
Mus musculus T (2)
Rattus norvegicus T
Homo sapiens T (4)
Homo sapiens X
Styela montereyensis P1
Entamoeba histolytica
0.1

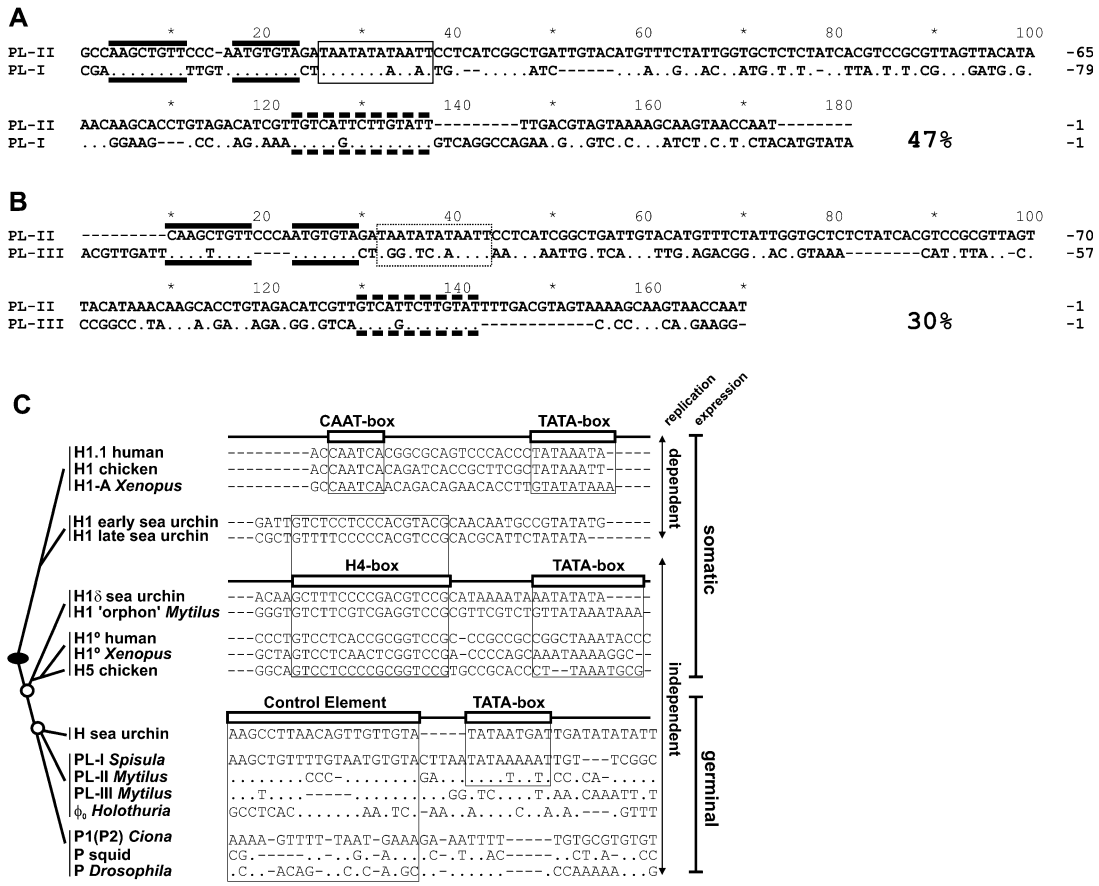


FIG. 5.—(*A*) Alignment of the upstream regions of the PL-II/IV gene from *Mytilus californianus* with those of the PL-I from *Spisula solidissima*. The conserved putative initiation binding site is indicated by dashed bars, the TATA box is indicated by a box, and the conserved putative H4 box–like region is indicated by solid bars. The overall nucleotide similarity in percentile is indicated on the right. (*B*) Alignment of the upstream regions of the PL-II/IV gene of *Mytilus* with the PL-III gene from *Mytilus*. The dashed box indicates the lack of a conserved TATA box in the PL-III gene. (*C*) Comparison of proximal promoter regions in different types of H1 and SNBPs. While RD H1 histones exhibit a characteristic CAAT box just upstream of the TATA box, both RI H1 histones and SNBPs lack this element. The H4 box is typical of RI H1 genes and is also found in RD sequences from sea urchins. All the SNBPs show a putative control element which is conserved to different extents within the different protein types (dots and dashes in the alignments represent matching bases and gaps with respect to the reference sequences in the PL and P types). The branching pattern on the left depicts the hypothetical long-term evolution of H1 and SNBPs derived from the phylogenies shown in figure 3.

Interestingly, a similar evolutionary process is observed in the evolution of the histone H1 family of chromosomal proteins where the differentiation between the RI and RD subtypes took place as early in bilaterian evolution as before the split between protostomes and deuterostomes (Eirín-López et al. 2004*a*), leading to a parallel birth-and-death evolution of both lineages (Eirín-López et al. 2005).

On the other hand, phylogenetic reconstructions that discriminate among the three major structural domains of the metazoan H1 molecule (fig. 4) strongly support the notion that the globular region containing the winged helix structure is a critical determinant of H1 identity in the different groups of metazoans, as indicated by the matching topologies of the trees obtained by this analysis (fig. 4*A* and *C*). Interestingly, PL proteins are more closely related to the RI H1 lineage, clustering in the monophyletic group containing these variants (fig. 4*A* and *C*), a finding that has not only evolutionary but also functional relevance as both

←

FIG. 4.—Phylogenetic Neighbor-Joining trees showing the relationships between H1 and PL proteins based on comparisons of the protein segments corresponding to the N-terminal and globular domains (*A*), the globular and the C-terminal domains (*B*), and the globular domain (*C*) of histone H1. The reconstructions are based on uncorrected *p*-distances using the complete-deletion option. The numbers for interior branches in the unrooted topology represent BP and also CP values, respectively, both based on 1,000 replications and only shown when a value is greater than 50%. The H1 or SNBP subtypes corresponding to different species are indicated, with the number of sequences analyzed in parentheses. The sperm-specific histone sequences (H type) are shown in boldface and underlined. The position of the PL SNBPs and those of the RD and the RI H1 lineages are highlighted on the right side of the phylogenies, and the common monophyletic origin for the PL and the RI H1 proteins is indicated by open circles. In all instances, the trees were rooted with the sequence of the H1-like protein from the protist *Entamoeba* as it represents one of the most primitive H1 histones known.

Table 1
Average Numbers of Synonymous (p_S) and Nonsynonymous (p_N) Nucleotide Differences Per Site and Tests of Selection in Protamine-like Genes

					Codon-Based *Z*-Test[b]		Codon-Based Fisher's Exact Test[c]
	R[a]	*p* (SE)	p_S (SE)	p_N (SE)	Statistic	*P* Value	*P* Value
PL-I	0.8	0.240 (0.008)	0.484 (0.022)	0.160 (0.011)	13.899	0.000	0.000
PL-II/PL-IV	1.1	0.081 (0.009)	0.235 (0.027)	0.032 (0.008)	7.442	0.000	0.000
PL-III	0.5	0.037 (0.011)	0.087 (0.034)	0.020 (0.009)	2.207	0.015	0.019
Overall	0.6	0.482 (0.011)	0.608 (0.017)	0.441 (0.015)	6.815	0.000	0.000

NOTE.—SE, standard error.
[a] Average transition/transversion ratio.
[b] H_0: $p_N = p_S$; H_1: $p_N < p_S$.
[c] H_0: $p_N = p_S$; H_1: $p_N > p_S$.

RI H1 and PL occur in terminally differentiated systems. The phylogenetic analysis obtained from the globular and the C-terminal regions represents an exception. In this instance, both RD and RI H1 proteins appear to be extensively interspersed (fig. 4*B*). The intrinsic protein disorder presented by C-terminal tails as well as the reiteration of different small amino acid motifs make very difficult to conduct sequence alignments of these regions. While the interspersed pattern of RD and RI proteins could be initially attributed to the high intrinsic protein disorder presented by the H1 C-terminal tail (Hansen et al. 2005), a recent report describing that this region becomes fully structured upon interaction with DNA (Roque et al. 2005) seems to debilitate this argument. Thus, a possible explanation could involve internal genomic processes such as the amplification of short motifs in H1 terminal regions, whose evolution through point mutation and further slippage would have been responsible for the differentiation of the N- and C-terminal regions with their intrinsic low sequence complexity in histone H1 (Ponte, Vila, and Suau 2003).

Finally, the clustering pattern of the PL proteins within the monophyletic group of the RI H1 histones has very important consequences in terms of the functional aspects of their evolutionary process. As was already mentioned, both lineages are specific for terminally differentiated systems, somatic in the case of RI H1 proteins (Doenecke et al. 1997) and germinal in the case of PL proteins (Ausió 1999; Eirín-López, Frehlick, and Ausió 2005). In contrast to RD H1 histones, RI H1 and PL protein genes are found in solitary locations in the genome and, in addition, are expressed as polyadenylated transcripts (Ausió 1999; Eirín-López et al. 2004*a*; Eirín-López et al. 2005). The comparisons of the promoter regions in PL genes show the presence of a conserved control element of 16 bp just upstream the TATA box (fig. 5*C*) which coincides with the position occupied by the H4 box in RI H1 proteins (Van Wijnen et al. 1992; Peretti and Khochbin 1997). Although there is no apparent similarity between both elements at the nucleotide level, they are distinct from the CAAT box, which is present at this location in the RD H1 subtypes.

The Long-Term Evolution of RI H1 histones and PL SNBPs

It is now well established that, while the long-term evolution of H1 histones is best described by a birth-and-death process under strong purifying selection (Nei and Hughes 1992; Eirín-López et al. 2004*a*; Eirín-López et al. 2005; Nei and Rooney 2005), the evolution of protamines is subject to a positive sex-driven selection, common to many genes expressed in male reproductive tissues (Eberhardt 1985; Wyckoff, Wang, and Wu 2000). To analyze the type of evolutionary constraints operating on PL SNBPs, it is critical to determine the position occupied by these chromosomal proteins in the midst of this apparent "shift" in the selection type experienced by the H1 and P types. The results obtained from the tests of selection performed on PL proteins unequivocally unveil the presence of purifying selection acting at the protein level, most likely determined by functional constraints. Thus, PL genes diverge extensively at the nucleotide level through synonymous substitutions.

Therefore, it is likely that the shift from negative to positive selection occurred at the time of the transition of a highly differentiated arginine-rich protamine (Lewis et al. 2004*b*). At this point, arginine-rich protamines would have been initially rapidly favored by positive selection mainly due to their higher affinity for DNA (Puigdomenech et al. 1976; Ausió, Greulich, and Watchel 1984). In addition, other features such as the higher flexibility imparted on DNA packaging (Cheng et al. 2003) and the role of polyarginine tracts at the time of sperm-egg fertilization (Ohtsuki et al. 1996) may have been important further determinants (Eirín-López, Frehlick, and Ausió 2005; Frehlick et al. 2006).

It is probable that the orphon origin proposed for RI H1 histones (Eirín-López et al. 2005) can be also extended to SNBPs, a notion that was indirectly suggested by a previous work (E. Schulze and B. Schulze 1995). The evolutionary relationships and processes undergone by histone H1 and SNBPs are summarized in figure 6, where they are ascribed to different taxonomic groups along a simplified phylogeny of the metazoan phyla based on combined analyses of morphology and molecular data (Giribet 2002). This useful representation shows that the exclusion of specialized H1 genes from the main repetitive histone genomic clusters to a solitary location in the genome early in the evolution of metazoans could have resulted in an orphon group. The independent evolution of this group may have ultimately led to the parallel differentiation and diversification of both the RI H1 somatic and the SNBP germinal lineages in protostomes and deuterostomes along the evolution of triploblastic animals.

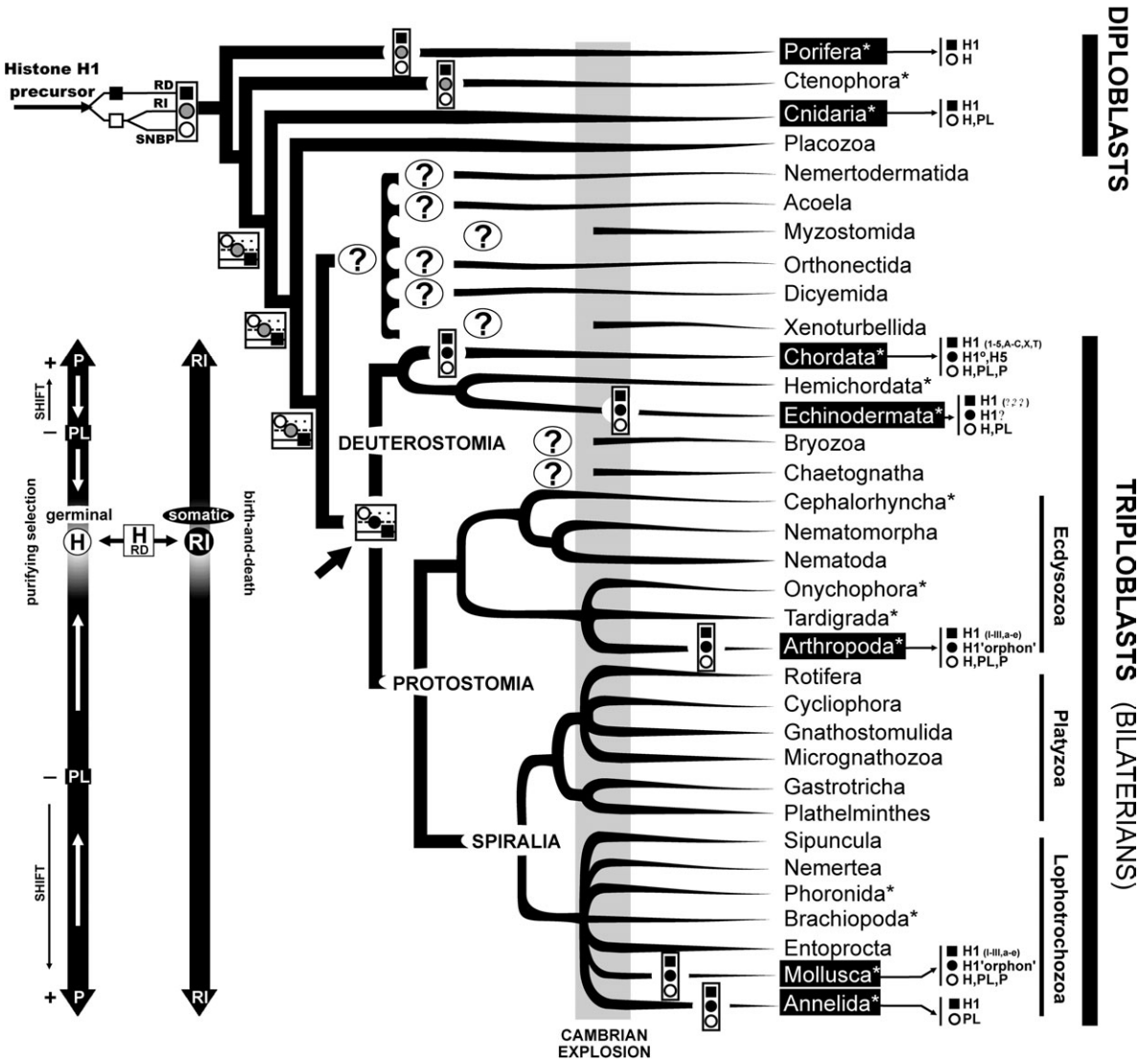


FIG. 6.—Simplified phylogeny of the metazoan phyla based on combined analyses of morphology and molecular data adapted from (Giribet 2002) showing the origin and evolution of the RD H1 (solid boxes), RI H1 (gray/solid circles), and protamine-like (PL-I) (open circles) lineages. The exclusion of a histone H1 precursor from the main units to a solitary location in the genome resulted in an orphon lineage. Further duplication events led to the differentiation of the RD H1 lineage in first instance and later on RI H1 and PL-I lineages diverged from a common ancestor. Most likely, the differentiation between germinal and somatic H1 lineages took place early in metazoan evolution. Otherwise, the complete functional differentiation of the three lineages would had arisen as early in the evolution of triploblastic animals as before the split between protostomes and deuterostomes (indicated by a black arrow), resulting in a parallel evolution subject to birth-and-death in the case of somatic H1 histones and to purifying selection in the case of germinal histones. The evolutionary processes to which the three histone lineages are subject are summarized in the left side of the phylogeny. While birth-and-death evolution remained as the main mechanism responsible for the functional diversification of RI and RD H1 lineages, PL-I proteins were subject to a purifying selection process leading to the transition of highly specialized PL proteins (i.e., PL-III in the bivalve mollusk *Mytilus* and ϕ_0 in the echinoderm *Holoturia*). The subsequent transition to protamines (P) undergone by PLs in the sperm of organisms at the furthermost tips of bilaterian evolution resulted in a shift toward a positive sex-driven selection, which was responsible for the high evolutionary rate presented by these proteins. Taxon names with asterisks indicate phyla with Cambrian fossil record. The gray window represents a temporary lapse symbolizing the Cambrian explosion, with some of the major irresolutions in the tree postulated to have occurred during that time framework. Uncertain relationships are highlighted by a question mark.

Supplementary Material

Supplementary Alignments 1–4, Figures 1–4, and Table are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). The sequences described in the present work have been deposited in the GenBank database with accession numbers DQ305038 and DQ305039.

Acknowledgments

We are very grateful to Lindsay J. Frehlick for carefully reading the manuscript and for helpful editorial suggestions as well as to four anonymous reviewers for useful discussions and technical suggestions. This work was supported by grants from Natural Sciences and Engineering Research Council of Canada, grant number OGP 0046399 (to J.A.)

and by a Postdoctoral Marie Curie International Fellowship within the 6th European Community Framework Programme (to J.M.E.-L.).

Literature Cited

Adams, M. D., S. E. Celniker, R. A. Holt et al. (193 co-authors). 2000. The genome sequence of Drosophila melanogaster. Science **287**:2185–2195.

Albig, W., T. Meergans, and D. Doenecke. 1997. Characterization of the H1.5 genes completes the set of human H1 subtype genes. Gene **184**:141–148.

Ausió, J. 1986. Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve mollusks. Comp. Biochem. Physiol. B Biochem. Mol. Biol. **85**:439–449.

———. 1988. An unusual cysteine-containing histone H1-like protein and two protamine-like proteins are the major nuclear proteins of the sperm of the bivalve mollusc Macoma nasuta. J. Biol. Chem. **263**:10141–10150.

———. 1995. Histone H1 and the evolution of the nuclear sperm-specific proteins. Pp. 447–462 *in* B. G. M. Jamieson, J. Ausió, and J. L. Justine, eds. Advances in spermatozoal phylogeny and taxonomy. Memoires du Museum National d'Histoire Naturelle, Paris, France.

———. 1999. Histone H1 and evolution of sperm nuclear basic proteins. J. Biol. Chem. **274**:31115–31118.

Ausió, J., K. O. Greulich, and E. Watchel. 1984. Characterization of the fluorescence of the protamine thynnine and studies of binding to double-stranded DNA. Biopolymers **23**:2559–2571.

Ausió, J., J. T. Soley, W. Burguer, J. D. Lewis, D. Barreda, and K. M. Cheng. 1999. The histidine-rich protamine from ostrich and tinamou sperm. A link between reptile and bird protamines. Biochemistry **38**:180–084.

Ausió, J., and K. E. van Holde. 1987. A dual chromatin organization in the sperm of the bivalve mollusc Spisula solidissima. Eur. J. Biochem. **165**:363–371.

Ausió, J., M. L. Van Veghel, R. Gómez, and D. Barreda. 1997. The sperm nuclear basic proteins (SNBPs) of the sponge Neofibularia nolitangere: implications for the molecular evolution of SNBPs. J. Mol. Evol. **45**:91–96.

Benkel, B. F., and Y. Fong. 1996. Long range-inverse PCR (LR-IPCR): extending the useful range of inverse PCR. Genet. Anal. Biomol. Eng. **13**:123–127.

Carlos, S., D. F. Hunt, C. Rocchini, D. P. Arnott, and J. Ausió. 1993. Post-translational cleavage of a histone H1-like protein in the sperm of Mytilus. J. Biol. Chem. **268**:195–199.

Cheng, A. C., W. W. Chen, C. N. Fuhrmann, and A. D. Frankel. 2003. Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. J. Mol. Biol. **327**:781–796.

Chiva, M., N. Saperas, C. Cáceres, and J. Ausió. 1995. Nuclear basic proteins from the sperm of tunicates, cephalochordates, agnathans and fish. Pp. 501–514 *in* B. G. M. Jamieson, J. Ausió, and J. L. Justine, eds. Advances in spermatozoal phylogeny and taxonomy. Memoires du Museum National d'Histoire Naturelle, Paris, France.

Doenecke, D., W. Albig, C. Bode, B. Drabent, K. Franke, K. Gavenis, and O. Witt. 1997. Histones: genetic diversity and tissue-specific gene expression. Histochem. Cell Biol. **107**:1–10.

Eberhardt, W. G. 1985. Sexual selection and animal genitalia. Harvard University Press, Cambridge.

Eirín-López, J. M., L. J. Frehlick, and J. Ausió. 2005. Protamines, in the footsteps of linker histone evolution. J. Biol. Chem. **281**:1–4.

Eirín-López, J. M., A. M. González-Tizón, A. Martínez, and J. Méndez. 2002. Molecular and evolutionary analysis of mussel histone genes (Mytilus spp.): possible evidence of an "orphon origin" for H1 histone genes. J. Mol. Evol. **55**:272–283.

———. 2004*a*. Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of *orphon* H1 genes. Mol. Biol. Evol. **21**:1992–2003.

Eirín-López, J. M., M. F. Ruiz, A. M. González-Tizón, A. Martínez, L. Sánchez, and J. Méndez. 2004*b*. Molecular evolutionary characterization of the mussel Mytilus histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with "orphon" features. J. Mol. Evol. **58**:131–144.

———. 2005. Common evolutionary origin and birth-and-death process in the replication-independent histone H1 isoforms from vertebrate and invertebrate genomes. J. Mol. Evol. **61**:398–407.

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

Frehlick, L. J., J. M. Eirín-López, A. Prado, H. W. Su, H. E. Kasinsky, and J. Ausió. 2006. Sperm nuclear basic proteins of two closely related species of Scorpaeniform fish (Sebastes maliger, Sebastolobus sp.) with different sexual reproduction and the evolution of fish protamines. J. Exp. Zool. A Comp. Exp. Biol. **305**:277–287.

Gimenez-Bonafe, P., E. Ribes, P. Sautiere, A. González, H. E. Kasinsky, M. Kouach, P. E. Sautiere, J. Ausió, and M. Chiva. 2002. Chromatin condensation, cysteine-rich protamine, and establishment of disulphide interprotamine bonds during spermiogenesis of *Eledone cirrhosa* (Cephalopoda). Eur. J. Cell Biol. **81**:341–348.

Giribet, G. 2002. Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? Mol. Phylogenet. Evol. **24**:345–357.

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

Hansen, J., X. Lu, E. Ross, and R. Woody. 2005. Intrinsic protein disorder, amino acid composition, and the histone terminal domains. J. Biol. Chem. **281**:1853–1856.

Heath, D. D., and T. J. Hilbish. 1998. Mytilus protamine-like sperm-specific protein genes are multicopy, dispersed, and closely associated with hypervariable RFLP regions. Genome **41**:587–596.

Jankowski, J. M., J. C. States, and G. H. Dixon. 1986. Evidence of sequences resembling avian retrovirus long terminal repeats flanking the trout protamine gene. J. Mol. Evol. **23**:1–10.

Jutglar, L., J. I. Borrell, and J. Ausió. 1991. Primary, secondary, and tertiary structure of the core of a histone H1-like protein from the sperm of Mytilus. J. Biol. Chem. **266**:8184–8191.

Kasinsky, H. E., J. D. Lewis, J. B. Dacks, and J. Ausió. 2001. Origin of H1 histones. FASEB J. **15**:34–42.

Kramer, J. A., and S. A. Krawetz. 1998. Genesis of a novel human sequence from the protamine PRM1 gene. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. **120**:467–473.

Krawetz, S. A., and G. H. Dixon. 1988. Sequence similarities of the protamine genes: implications for regulation and evolution. J. Mol. Evol. **27**:291–297.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5**:150–163.

Lewis, J. D., and J. Ausió. 2002. Protamine-like proteins: evidence for a novel chromatin structure. Biochem. Cell Biol. **80**:353–361.

Lewis, J. D., Y. Song, M. de Jong, S. Bagha, and J. Ausió. 2003. A walk though vertebrate and invertebrate protamines. Chromosoma **111**:473–482.

Lewis, J. D., M. De Jong, S. Bagha, A. Tang, W. F. Gilly, and J. Ausió. 2004*a*. All roads lead to arginine: the squid protamine gene. J. Mol. Evol. **58**:673–680.

Lewis, J. D., N. Saperas, Y. Song, M. J. Zamora, M. Chiva, and J. Ausió. 2004*b*. Histone H1 and the origin of protamines. Proc. Natl. Acad. Sci. USA **101**:4148–4152.

Meistrich, M. L. 1989. Histone and basic nuclear protein transitions in mammalian spermatogenesis. Pp. 165–182 *in* L. S. Hnilica, G. S. Stein, and J. L. Stein, eds. Histones and other basic nuclear proteins. CRC Press, Orlando, Fla.

Nei, M., and A. L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27–38 *in* K. Tsuji, M. Aizawa, and T. Sasazuki, eds. 11th histocompatibility workshop and conference. Oxford University Press, Oxford.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution in multigene families. Annu. Rev. Genet. **39**: 121–152.

Ohtsuki, K., Y. Nishikawa, H. Saito, H. Munakata, and T. Kato. 1996. DNA-binding sperm proteins with oligo-arginine clusters function as potent activators for egg CK-II. FEBS Lett. **378**:115–120.

Oliva, R. 1995. Sequence, evolution and transcriptional regulation of avian-mammalian P1 type protamines. Pp. 537–548 *in* B. G. M. Jamieson, J. Ausió, and J. L. Justine, eds. Advances in spermatozoal phylogeny and taxonomy. Memoires du Museum National d'Histoire Naturelle, Paris, France.

Oliva, R., and G. H. Dixon. 1991. Vertebrate protamine genes and the histone-to-protamine replacement reaction. Prog. Nucleic Acid Res. **40**:25–94.

Peretti, M., and S. Khochbin. 1997. The evolution of the differentiation-specific histone H1 gene basal promoter. J. Mol. Evol. **44**:128–134.

Phelan, J. J., J. Colom, C. Cozcolluela, J. A. Subirana, and R. D. Cole. 1974. A lysine-rich protein from spermatozoa of the mollusc Mytilus edulis. J. Biol. Chem. **249**:1099–1102.

Ponte, I., R. Vila, and P. Suau. 2003. Sequence complexity of histone H1 subtypes. Mol. Biol. Evol. **20**:371–380.

Prats, E., and L. Cornudella. 1995. The gene encoding the sperm-specific basic nuclear protein fo from sea cucumber. Pp. 491-500 *in* B. G. M. Jamieson, J. Ausió, and J. L. Justine, eds. Advances in Spermatozoal Phylogeny and Taxonomy. Memoires du Museum National d'Histoire Naturelle, Paris, France.

Puigdomenech, P., P. Martínez, J. Palau, E. M. Bradbury, and C. Crane-Robinson. 1976. Studies on the role and mode of operation of the very-lysine-rich histones in eukaryote chromatin. Nuclear-magnetic-resonance studies on nucleoprotein and histone phi 1-DNA complexes from marine invertebrate sperm. Eur. J. Biochem. **65**:357–363.

Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature **362**:219–223.

Rocchini, C., P. Rice, and J. Ausió. 1995. Complete sequence and characterization of the major sperm nuclear basic protein from Mytilus trossulus. FEBS Lett. **363**:37–40.

Roque, A., I. Iloro, I. Ponte, J. L. Arrondo, and P. Suau. 2005. DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. J. Biol. Chem. **280**:32141–32147.

Ruiz-Lara, S., E. Prats, M. T. Casas, and L. Cornudella. 1993. Molecular cloning and sequence of a cDNA for the sperm-specific protein phi 1 from the mussel Mytilus edulis. Nucleic Acids Res. **21**:2774.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**: 945–967.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor, New York.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

Saperas, N., J. Ausió, D. Lloris, and M. Chiva. 1994. On the evolution of protamines in bony fish: alternatives to the "retroviral horizontal transmission" hypothesis. J. Mol. Evol. **39**: 282–295.

Saperas, N., M. Chiva, D. C. Pfeiffer, H. E. Kasinsky, and J. Ausió. 1997. Sperm nuclear basic proteins (SNBPs) of agnathans and chondrichthyans: variability and evolution of sperm proteins in fish. J. Mol. Evol. **44**:422–431.

Schulze, E., and B. Schulze. 1995. The vertebrate linker histones H1°, H5, and H1M are descendants of invertebrate "orphon" histone H1 genes. J. Mol. Evol. **41**:833–840.

Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. Mol. Biol. Evol. **13**:605–611.

Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. Mol. Biol. Evol. **12**: 319–333.

Subirana, J. A. 1970. Nuclear proteins from a somatic and a germinal tissue of the echinoderm *Holothuria tubulosa*. Exp. Cell Res. **63**:253–260.

Subirana, J. A., and J. Colom. 1987. Comparison of protamines from freshwater and marine bivalve molluscs: evolutionary implications. FEBS Lett. **220**:193–196.

Subirana, J. A., C. Cozcolluela, J. Palau, and M. Unzueta. 1973. Protamines and other basic proteins from spermatozoa of molluscs. Biochim. Biophys. Acta **317**:364–379.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

Van Wijnen, A. J., F. M. I. van Den Ent, J. B. Lian, J. L. Stein, and G. S. Stein. 1992. Overlapping and CpG methylation-sensitive protein-DNA interaction at the histone H4 transcriptional cell cycle domain: distinctions between two human H4 gene promoters. Mol. Cell. Biol. **12**:3273–3287.

Wyckoff, G. J., W. Wang, and C. I. Wu. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304–309.

Zhang, J., S. Kumar, and M. Nei. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. **14**:1335–1338.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

Zhang, Z., and S. J. Gurr. 2000. Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. Gene **253**:145–150.

Billie Swalla, Associate Editor

Accepted April 1, 2006