Recent relaxation of purifying selection on the tandem-repetitive early-stage histone H3 gene in brooding sea stars

David W. Foltz a,⁎, Christopher L. Mah b

a Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA
b Department of Invertebrate Zoology, National Museum of Natural History, MRC-163, PO Box 37012, Smithsonian Institution, Washington, DC 20007, USA

A R T I C L E   I N F O

Keywords:
Nonsynonymous substitutions
Larval habitat
Multigene family
Focipulatida
Echinoderms
Sea stars

A B S T R A C T

Patterns of nucleotide substitution differ between marine species that have a pelagic feeding (planktotrophic) larval stage and related species that lack such a stage, for both adaptive and non-adaptive reasons. Here, patterns of nucleotide and inferred amino acid substitution are analyzed for the tandem-repetitive early-stage histone H3 gene in 36 sea star species of the order Forcipulatida with documented larval habitat. The relative rate of nonsynonymous substitution (expressed as \( \omega = d_\text{S}/d_\text{N} \)) was significantly higher in lineages with a brooded non-feeding (lecithotrophic) larval form than in lineages with a planktotrophic larval form. There was also a significant excess of conservative over radical substitutions. The increase in \( \omega \) for brooders as compared to non-brooders was much greater than for previously analyzed mitochondrial sequences in echinoderms. These data are consistent with the hypothesis that purifying selection on this gene has been relaxed in brooding lineages compared to non-brooding lineages. The hypotheses of adaptive or neutral evolution are less plausible, although recent pseudogenization following a period of relaxed purifying selection could also explain the results.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The evolutionary loss of a pelagic feeding (planktotrophic) larval stage in echinoderms and other marine organisms is associated with altered patterns of nucleotide substitution for both mitochondrial and nuclear genes, compared to related species that have retained planktotrophic larvae. Some of the differences, such as elevated rates of nonsynonymous substitutions and reduced nucleotide diversities for mitochondrial protein-coding genes in brooding sea stars, where females retain fertilized eggs and non-feeding (lecithotrophic) larvae on the benthic surface, can plausibly be attributed to reduced effective population sizes in brooding lineages (Foltz, 2003; Foltz et al., 2004). These patterns are similar in magnitude to differences in nonsynonymous substitution rates for mitochondrial genes observed between island and mainland populations (Johnson and Seger, 2001; Woolfit and Bromham, 2005), or between asexual and sexual lineages (Paland and Lynch, 2006; Johnson and Howard, 2007). In contrast, for nuclear genes that are expressed only in certain tissues or at certain developmental stages, genetic differences between lineages that have a planktotrophic larva and related lineages that lack such a stage have several potential explanations (Smith et al., 2007). Genes that are involved in producing larval feeding structures, or that are expressed only in tissues specific to the planktotrophic larval form, have been hypothesized to experience a relaxation of purifying selection and genetic drift during and after the transition to a lecithotrophic larval form (Wray, 1996), leading ultimately to pseudogenization (e.g., Kissinger et al., 1997). Adaptive changes are also possible during and after the transition to a lecithotrophic larval form. For example, regulatory and signaling genes may be recruited to new roles (Raff et al., 1999; Lowe et al., 2002; Wilson et al., 2005), or the timing or location of expression of these genes may be altered (Ferkowicz and Raff, 2001; Wilson et al., 2005).

Less attention has been paid to the possibility that developmentally-regulated genes that are tandem-repetitive, such as members of the histone gene family in echinoderms, may also experience relaxed purifying selection in species that lack planktotrophic larvae. Simulation analysis of two selection models (truncation selection and exponential fitness) showed that the equilibrium frequency of deleterious alleles per copy in a multigene family was generally an increasing function of copy number, recombination rate and mutation rate, and a decreasing function of the selection coefficient and the probability of a gene conversion event per copy per generation (Ohta, 1989). These results suggest that members of a multigene family may harbor higher levels of deleterious alleles than single-copy genes, even before considering the possible consequence of a relaxation of purifying selection associated with larval form. The tandem-repetitive early-stage histone H3 gene family in sea stars shows a pattern of organization and evolution similar to that observed in sea urchins, with approximately 500 copies per haploid genome and a high level of synonymous substitutions (and no nonsynonymous substitutions) between distantly-related species.
(Maxson et al., 1983; Cool et al., 1988), but as yet there are very few histone H3 sequence data available for brooding lineages. As part of a molecular phylogenetic study of forcipulate sea stars, sequence data were collected for 109 codons of the early-stage histone H3 gene in 36 forcipulate sea star species with documented larval type. This data set allowed us to test the null hypothesis of no difference in relative nonsynonymous substitution rate ($\omega = d_N/d_S$) between brooding and non-brooding lineages and to determine that brooding lineages had a dramatically elevated $\omega$ value when compared to non-brooding lineages belonging to the same order, when tested by either parametric or non-parametric methods.

2. Materials and methods

DNA was extracted from fresh, frozen or alcohol-preserved sea star tube feet as in Foltz (2007). PCR reactions were set up and cycle sequencing was performed as in Foltz et al. (2007a). Primers for PCR and sequencing were modified from those designed by Daniel A. Janies (Ohio State University, Columbus, Ohio, USA, personal communication):

$$H3F\ 5’-ACA\ ATG\ GGY\ GGY\ ACY\ AAG\ CAG\ ACA\ GC-3’$$
$$H3R\ 5’-GTT\ GGA\ TGY\ GYT\ TGG\ GCA\ TGA\ TGG\ T-3’$$

The analyzed region, plus the adjacent priming sites, comprised around 92% of the coding region of this small, intron-less gene.

Sequence data obtained with the above primers for Anasterias antarctica showed extensive secondary peaks suggestive of amplification of two or more divergent paralogous gene copies. This extract was re-amplified and sequenced with the following degenerate internal primers, which gave cleaner results:

**Modified F**  $5’-\text{CGT\ ACT\ AAG\ CAC\ GGA\ GCW\ MGY\ AAR\ AG-3’}$
**Modified R**  $5’-\text{GTT\ GGA\ TGT\ CTT\ TGG\ GCA\ TGA\ TGG\ TNA\ CC-3’}$

Multi-copy genes like the early-stage histone H3 gene may evolve by processes different from those of single-copy nuclear genes, such as concerted evolution, gene conversion or birth-death dynamics (Nei and Rooney, 2005). Also, the existence of multiple copies of the same gene creates the possibility of a high frequency of sequence variants that could confound the usual interpretation of a sequence chromatogram. However, the polymorphism (i.e., secondary peaks) of third-codon positions within individual sea star early-stage histone H3 chromatograms ($N = 34$ comparisons by paired $t$-test) was not significantly greater than previously observed for introns of two single-copy nuclear protein-coding genes for the same individuals (Foltz, 2007), suggesting that sequence variants of the early-stage histone H3 gene that are created by mutation are rapidly homogenized within and between gene copies, either by concerted evolution or by other processes.

![Fig. 1. Maximum likelihood tree for 40 unique early-stage histone H3 sequences in forcipulate sea stars with known larval type, rooted on Z. fulgens. Nodes are labeled with bootstrap proportions as percentages when > 50%, and the scale bar shows the estimated number of nucleotide substitutions per site. In Figs. 1 and 2, numbers in square brackets differentiate conspecific sequences, as per Supplementary Table 4. The following clades with moderate-to-high bootstrap support were completely congruent between the histone H3 gene tree and a gene tree derived from mitochondrial and nuclear sequences (Foltz et al., 2007b): (1) Astrometis + Coscinasterias + Coronaster + Diplasterias (= Cryptasterias), (2) Stylasterias + Lethasterias, (3) Asterias, (4) Pisaster, (5) Leptasterias + Evasterias, and (6) Marthasterias + Urasterias.](image-url)
We obtained new early-stage histone H3 sequences from 49 forcipulate specimens, supplemented with two sequences from GenBank (see Supplementary Table 4 for GenBank accession numbers). Taxonomic identification of the Leptasterias aequa basil species complex was from Foltz et al. (2008); other identifications were by Mah (except for GenBank sequences). Although accurate fitting of different substitution rates to brooding and non-brooding lineages requires nearly complete taxon sampling (see Foltz, 2003 for discussion), errors in assigning the correct reproductive mode to particular branches as a result of incomplete sampling would tend to equalize rates between brooding and non-brooding lineages. This conclusion assumes that shifts in reproductive mode are infrequent within a clade, and that long terminal branches for brooding lineages composed of a single sampled species (e.g., Granaster nutrix and Neosmilaster steineni in Figs. 1 and 2) include an unknown but potentially long period of time before the brooding mode evolved.

Data on mode of reproduction were obtained primarily from Emlet, McEdward and Strathmann (1987) and McEdward and Miner (2001), supplemented by Janosik et al. (2008). Two types of lecithotrophy have been reported from forcipulate sea stars: brooded lecithotrophy and pelagic lecithotrophy. Although pelagic lecithotrophy is relatively common in some sea star families, such as the Benthopectinidae, the Echinasteridae, the Asterinidae and the Solasteridae (McEdward and Miner, 2001), it is relatively rare among forcipulate sea stars, and only two of the 36 species with data on reproductive mode (Urasterias lincki and Zoroaster fulgens) have pelagic lecithotrophic larvae. In the present analysis, therefore, difference in larval nutrition (lecithotrophic versus planktotrophic) is largely confounded with difference in larval habitat (brooded versus pelagic). Because of the rarity of pelagic lecithotrophy in the present study and for simplicity, we have used the terms ‘brooder’ or ‘brooding’ to refer to species with brooded lecithotrophic larvae and the terms ‘lecithotrophy’ or ‘lecithotrophic’ without further qualification to refer collectively to both types of lecithotrophy.

Tree reconstruction with bootstrapping for the 40 unique sequences in the data set was done by using approximate maximum-likelihood (ML) computation in RAxML-VI-HPC v. 7.0.4 (Stamatakis et al., 2008). Bootstrap proportions (300 pseudoreplicates determined automatically by RAxML) were calculated using the Consense program of Phylip v. 3.6 (evolution.genetics.washington.edu/phylip/doc/consense.html). The codeml program in PAML v. 4.2b (Yang, 2007) was used to estimate synonymous \( (d_S) \) and nonsynonymous \( (d_N) \) substitution rates per site and test different \( d_N/d_S \) ratios with log-likelihood ratio test statistics \( (2 \Delta l) \). The topology of the best ML tree obtained from RAxML was input to codeml as a user-specified tree. Because likelihood ratio tests are potentially subject to

![Fig. 2. Phylogenetic trees constructed using maximum-likelihood estimates of the expected number of synonymous (left) and nonsynonymous (right) substitutions per site under a two-ratio model that allows brooding lineages (gray branches) to have different rates than non-brooding lineages (black branches). The scale bars show the estimated number of nucleotide substitutions per synonymous site \( (d_S) \) or per nonsynonymous site \( (d_N) \).](image-url)
nonconservative behavior due to model mis-specification and other causes (e.g., Friedman and Hughes, 2007), we also performed exact tests of heterogeneity in patterns of synonymous and nonsynonymous substitutions between brooding and non-brooding lineages as in Zhang, Kumar and Nei (1997), using ancestral codon reconstruction in codeml v. 4.2b. Various molecular clock models were tested with log-likelihood ratio test statistics from the basal program in PAML v. 4.2b, using the best-fit model of nucleotide substitution obtained with the Akaike Information Criterion (AIC) in Modeltest v. 3.7 (Posada and Buckley, 2004) and the topology from Fig. 1 as a user tree. For this analysis, the tree was rooted on Z. fulgens. Details of all models analyzed with the PAML package are in Supplementary Table 1. Nonsynonymous substitutions were classified into six groups by polarity and volume. Within-group substitutions were characterized as conservative and between-group substitutions were characterized as radical, and nonsynonymous sites were classified as radical or conservative using the program HON-NEW (Zhang, 2000).

3. Results

Analysis using RAxML recovered a tree (Fig. 1) that was mostly congruent with published forcipulate trees based on mitochondrial and nuclear rDNA sequences (Foltz et al., 2007b, see Fig. 1 for details) and on mitochondrial and nuclear single-copy sequence data (Foltz et al., 2008). For both previously-published and early-stage histone H3 gene trees, there was a lack of phylogenetic resolution for basal relationships as well as poor resolution among some recently-diverged species in the genus Leptasterias. Also, cloned early-stage histone H3 sequences from GenBank for two forcipulate genera (Pisaster [X07503, X07504, X54112] and Pycnopodia [X54114]) were identical to PCR-derived sequences from the same taxa collected in the present study, which allowed us to identify the amplified region as the early-stage histone H3 gene. These results suggest that the histone H3 sequences analyzed here are of the early embryonic stage type, and are evolving in an orthologous fashion.

Data on larval type were available for 36 forcipulate species, 10 brooders and 26 non-brooders (Fig. 2). The inferred ancestral histone H3 amino acid sequence (see Supplementary Table 3) was shared by 25 of 26 non-brooding forcipulate species and by four of 10 brooding species. The remaining seven species in Fig. 2 showed one or more inferred amino acid substitutions (details in Supplementary Table 3). The two species in Fig. 2 with pelagic lecithotrophic larvae (U. lincki and Z. fulgens) did not show any nonsynonymous substitutions. The $\omega$ ($d_s/d_o$) ratio obtained for a 1-rate model fitted to all 36 species was 0.020; the corresponding ratios for the 2-rate model were 0.001 for non-brooders and 0.330 for brooders (details in Supplementary Table 1). The 2$\Delta$ value was 69.8 with 1 degree of freedom (df), so brooders had a significantly ($P<0.0001$) higher $\omega$ value than non-brooders. The 2-rate model also performed better when tested against an alternate model in which the $\omega$ ratio for brooders was fixed at the neutral value ($\omega = 1$; Yang, 2007) and the $\omega$ ratio for non-brooders was estimated from the data (2$\Delta$ = 92.2, 1 df, $P<0.005$) The best-fit model of nucleotide substitution (TIM+G) was used to obtain log-likelihood values for two clock models: [1] a global clock and [2] a local clock in which brooders had a different substitution rate than non-brooders. The local clock model did not fit the data significantly better than the global clock model (2$\Delta$ = 1.36, 1 df, $P>0.05$), so there was no evidence of nucleotide substitution rate heterogeneity associated with larval type. Three different groups of brooding sea stars in Fig. 2 were tested non-parametrically and separately for heterogeneity in numbers of synonymous and nonsynonymous substitutions, as compared to closely-related non-brooding lineages. For each comparison, the difference in $\omega$ ratio between brooding and non-brooding lineages was significant, after allowance for multiple testing, when analyzed by the more conservative exact test of Zhang, Kumar and Nei (1997, see Supplementary Table 2 and Supplementary Fig. 1 for details). This result suggests that the conclusion that brooding lineages of forcipulate sea stars have excess numbers of nonsynonymous substitutions is robust to possible violations of the assumptions of the parametric tests presented above.

One-third of the sites in the 327 bp alignment were variable, and about 83% of the codons (90 of 109) had one or more nucleotide substitutions. Only one of the nonsynonymous substitutions was demonstrably polymorphic within a species (see Supplementary Table 3). Of the 16 nonsynonymous substitutions, 12 were conservative and four were radical, using the six-group classification of Zhang (2000). Considering only nonsynonymous substitutions that occurred in two or more individuals (which were less likely than singleton substitutions to be PCR artifacts, sequencing errors or within-species polymorphisms), there were six conservative substitutions and no radical substitutions. Because approximately 33% of the nonsynonymous sites in the sequenced gene fragment were conservative and 67% were radical, there was a significant excess of conservative substitutions by a binomial test ($P<0.005$).

4. Discussion

Two classes of histone H3 sequence exist in sea urchins: early-stage (α-class) genes that are expressed between fertilization and the late blastula stage, and late-stage (γ-class) genes that are expressed after the late blastula stage. Early-stage histone H3 genes are organized in a unit with histone H1, H2A, H2B and H4 genes, with several hundred tandem-repetitive copies of the unit in one or more chromosome locations per haploid genome. Late-stage histone H3 genes exist as single copies or in H3/H4 gene pairs, with approximately 10 copies per haploid genome (Maxson et al., 1983). Early-stage histone H3 genes in the sea star Pisaster are also organized in a unit with histone H2A, H2B and H4 genes, with several hundred tandem-repetitive copies per haploid genome, although the gene order and expression level in pre- and post-fertilization eggs is different than in sea urchins (Howell et al., 1987; Cool et al., 1988). In particular, there are no detectable amounts of early-stage histone H3 mRNA transcripts in Pisaster eggs until 12 h post-fertilization (7th cleavage), in contrast to the eggs of advanced sea urchins, which accumulate up to 10$^6$ copies of maternal early-stage histone H3 mRNA transcripts in the female pronucleus (Maxson et al., 1983; Raff et al., 1984; Howell et al., 1987).

Histone polypeptides represent a classic example of phylogenetically-deep amino acid sequence conservation (e.g., Maxson et al., 1983; Malik and Henikoff, 2003), which presumably reflects the selective constraints imposed by the need to interact both with DNA and with other histone monomers during nucleosome formation. Given [1] the high degree of inferred amino acid sequence conservation previously reported for the early-stage histone H3 gene in sea stars (Cool et al., 1988) and for other histone sequences, and [2] that the only previous reports of relaxed purifying selection on histone gene sequences involve euchromatic microorganisms (Piontkivska et al., 2002; Malik and Henikoff, 2003), our initial expectation was that variation in this gene in forcipulates would be concentrated at synonymous sites. The discovery of abundant nonsynonymous substitutions in brooding lineages of forcipulate sea stars, at approximately 1/3 the synonymous rate and much greater than the relative excess of nonsynonymous substitutions observed for the mitochondrial COI gene in brooding versus non-brooding echinoderms (Foltz, 2003; Foltz et al., 2004), was unexpected. At present, too little is known of gene expression patterns in brooding forcipulate sea stars to propose a functional reason for the relaxed purifying selection inferred here. A hypothesis for future research is that the timing of expression of the early-stage histone H3 gene may be altered in lecithotrophic lineages compared to planktotrophic lineages, analogous to the heterochronic shift in expression for a developmentally-important gene that was observed in a sea urchin species with lecithotrophic larvae, when compared to a related species.
with planktotrophic larvae (Ferkowicz and Raff, 2001). If the expression of the early-stage histone H3 gene is greatly abbreviated in duration in species with lecithotrophic larvae, compared to related species with planktotrophic larvae, the result could be a relaxation of purifying selection in the former. As mentioned above, lecithotrophy and brooding are confounded in the present study, in that only two of the studied species (U. lincki and Z. fulgens) have pelagic lecithotrophic larvae. Neither of those species showed any nonsynonymous substitutions in the present study, suggesting that larval habitat rather than larval nutrition is the important factor influencing early-stage histone H3 gene evolution in sea stars. However, pelagic lecithotrophy is common in some non-broopulae sea star families, which would allow lecithotrophic versus planktotrophic lineages to be analyzed more rigorously, without the complicating effect of larval habitat (pelagic versus brooded). When an early-stage histone H3 sequence from the family Asterinidae (DQ676891) was used to BLAST two EST databases for the same species (Patiria miniata, 1249 total ESTs) and a related asterinid (P. pectinifera, 56381 total ESTs), 160 hits were obtained for libraries constructed from gonadal and 45-hour embryonic tissue. However, all 160 cDNA sequences clustered with sea urchin late-stage histone H3 sequences rather than with sea star early-stage histone H3 sequences (details not shown). The failure to recover any early-stage histone H3 sequences among these cDNAs is consistent with the observation (discussed above) that early-stage histone H3 transcripts are present at undetectable levels in sea star eggs and embryos until 12 h post-fertilization. Unfortunately, both asteroid species used for EST library construction have pelagic planktotrophic larvae, so these results do not shed any light on possible expression in the studied species (Ohta, 1989; Graham, 1995). Thus, the effect of reduced nonsynonymous substitution rates (ω) in brooders compared to non-brooders. However, the increase in ω value for brooders versus non-brooders observed here (0.330 vs. 0.001) is much larger than observed previously for comparable studies of mitochondrial COI sequences (e.g., 0.014 vs. 0.002, Foltz et al., 2004), which suggests that some factor in addition to reduced effective population size is also involved. Another hypothesis for future research is that tandem-repetitive genes are already under a form of relaxed purifying selection due to their high-copy number, and are expected to accumulate deleterious alleles at a higher per-gene rate than for single-copy or low-copy number genes (Ohta, 1989; Graham, 1995). Thus, the effect of reduced effective population size in brooding lineages would be exaggerated for high-copy number nuclear genes, because after the evolution of a brooded lecithotrophic larval type, pre-existing deleterious alleles could spread through the tandem array by gene conversion or unequal recombination. This hypothesis could be tested by obtaining late-stage histone H3 gene sequences, as well as estimates of gene copy number for both early-stage and late-stage histone H3 genes, in brooding and non-brooding sea star lineages. The prediction is that late-stage histone H3 gene sequences (if present in low-copy number in sea stars as in sea urchins), should show an elevated nonsynonymous substitution rate in brooding lineages that is more comparable to the mitochondrial rate than to the early-stage histone H3 rate.

Compared to the hypothesis of relaxed purifying selection on the early-stage histone H3 gene in brooding sea stars, adaptive or purely neutral (i.e., pseudogenization) explanations are less likely, given [1] the fact that ω = 1 in brooding lineages and [2] the scarcity of previous reports of adaptive nucleotide substitutions in this gene for closely-related species in other phyla. There were no evident ecological correlates with the occurrence of nonsynonymous substitutions that would suggest adaptive evolution; nonsynonymous substitutions were observed in both temperate and high-latitude lineages as well as in both shallow-water and deep-water lineages. Additional evidence against pseudogenization is the absence of nonsense or frameshift mutations over approximately 200 million years of early-stage histone H3 sequence evolution in the order Forcipulatida (dating from Foltz, 2007). However, variations of the basic pseudogene hypothesis are somewhat more plausible. Possibly several paralogous early-stage histone arrays exist in the sea star genome, and in brooding lineages one or more functional arrays co-exist with a pseudogenized array that predominates in the PCR. Another possibility is that the early-stage histone H3 sequences in brooders went through a period of relaxed purifying selection during the evolution of a brooded lecithotrophic larval type, but subsequently became pseudogenes. Given the recent age of all inferred amino acid substitutions in Supplementary Table 3 (none of the dated amino acid substitutions is older than 4.3 Mya, and half are less than 1.4 Mya), it could be that there has been insufficient time for the early-stage histone H3 sequences in brooders to acquire the genetic signature of a pseudogene.

Acknowledgments

We thank Dr. Robert Van Syoc for arranging loan of material from the California Academy of Sciences and Drs. Mark A. Batzer and John C. Larkin for suggestions and comments on the manuscript. James W. Halk and Cali M. Meyer assisted in the laboratory. Financial support was provided by NSF grant OPP-0631245 to CL Mah and an award from the Office of Research and Economic Development at Louisiana State University to DW Foltz.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmargen.2009.06.002.

References