EVOLUTION & DEVELOPMENT 12:2, 222–230 (2010)
DOI: 10.1111/j.1525-142X.2010.00406.x

Difference in larval type explains patterns of nonsynonymous substitutions in two ancient paralogs of the histone H3 gene in sea stars

David W. Foltz[a,*] and 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
*Author for correspondence (email: dfoltz@lsu.edu)

SUMMARY Paralogous genes frequently show differences in patterns and rates of substitution that are typically attributed to different selection regimes, mutation rates, or local recombination rates. Here, two anciently diverged paralogous copies of the histone H3 gene in sea stars, the tandem-repetitive early-stage gene and a newly isolated gene with lower copy number that was termed the "putative late-stage histone H3 gene" were analyzed in 69 species with varying mode of larval development. The two genes showed differences in relative copy number, overall substitution rates, nucleotide composition, and codon usage, but similar patterns of relative nonsynonymous substitution rates, when analyzed by the d_N/d_S ratio. Sea stars with a nonpelagic and nonfeeding larval type (i.e., brooding lineages) were observed to have d_N/d_S ratios that were larger than for nonbrooders but equal between the two paralogs. This finding suggested that demographic differences between brooding and nonbrooding lineages were responsible for the elevated d_N/d_S ratios observed for brooders and refuted a suggestion from a previous analysis of the early-stage gene that the excess nonsynonymous substitutions were due to either (1) gene expression differences at the larval stage between brooders and nonbrooders or (2) the highly repetitive structure of the early-stage histone H3 gene.

INTRODUCTION

The loss of a pelagic feeding (planktotrophic) larval stage has occurred one or more times in various groups of marine invertebrates, including gastropods, echinoderms, and ascidians (Pechenik 1999). Such a change in reproductive mode potentially affects many evolutionary processes, such as amount of gene flow, level of inbreeding and effective population size, possibility of local adaptation, rates of extinction, and rates and mode of speciation. The lack of a planktotrophic larval stage also potentially affects demographic processes including dispersal, fecundity, larval growth rates, and recruitment, and may have relevance to applied ecological questions, such as the invasion success of exotic species, the design of marine reserves, or the ability of a species to recover from habitat disturbance. There have been numerous studies on the relationship between reproductive mode and spatial population genetic structure in marine invertebrates (e.g., Arndt and Smith 1998; Johnson and Black 2006; Ayre et al. 2009; Lee and Boulding 2009; and references therein). These studies have generally found that marine invertebrate species lacking a dispersive larval stage have significantly greater population subdivision (typically measured by F_{ST}) than related species that have a dispersive larval stage, although counter-examples have also been reported (e.g., Miller and Ayre 2008). Somewhat fewer studies have compared levels of within-population genetic diversity (measured by nucleotide diversity π or genic heterozygosity H_e) in closely related invertebrate species that differ in larval type (e.g., Foltz 2003; Lee and Boulding 2009). For example, nucleotide diversity for a mitochondrial gene and a single-copy nuclear gene were both higher in two gastropod species with pelagic planktotrophic larvae when compared with related species that had benthic lecithotrophic larvae (Lee and Boulding 2009). Also, within-sample π values were significantly lower ($P < 0.0005$) in *Parvulastra exigua* (benthic lecithotrophy) as compared with the confamilial sea star species *Meridiastra calcar* (pelagic lecithotrophy), when we reanalyzed mitochondrial COI sequence data originally published by Ward et al. (2008) and Ayre et al. (2009), using only synonymous differences in a Wilcoxon signed-rank test. However, no significant difference ($P > 0.05$) in within-sample π values could be detected in a similar reanalysis of nonprotein-coding sequence variation in the sea cucumber species *Cucumaria pseudocurata* (brooded lecithotrophy) and its congener *Cucumaria miniata* (pelagic lecithotrophy), based on data of Arndt and Smith (1998). These results are generally consistent with the suggestion that species lacking a pelagic larval stage have lower long-term effective population sizes

© 2010 Wiley Periodicals, Inc.

than related species possessing a pelagic larval stage (Foltz 2003). However, the interpretation of π values in relation to reproductive mode is not as straight forward as for F_{ST} values, because other evolutionary factors (principally mutation rate heterogeneity and selective sweeps) in addition to long-term effective population size can potentially influence species-wide genetic diversity for mitochondrial genes (e.g., Galtier et al. 2009).

Here, a different possible genetic consequence of reproductive mode in marine invertebrates was examined, namely, alteration of patterns of nucleotide substitution in protein-coding genes for lineages with nonpelagic and nonfeeding (e.g., brooded lecithotrophic) larvae. The work was prompted by a recent report of a large excess of nonsynonymous substitutions in the tandem-repetitive early-stage histone H3 gene, relative to the synonymous substitution rate, in brooding lineages of forcipulate (i.e., order Forcipulatida) sea stars by Foltz and Mah (2009). The increase in nonsynonymous substitution rate in brooders, to about 1/3 of the magnitude of the synonymous substitution rate, was much larger than previously seen for mitochondrial genes in similar comparisons of brooding versus nonbrooding lineages in various marine invertebrate species (Foltz 2003; Foltz et al. 2004). Foltz and Mah (2009) hypothesized that the anomalously large excess of nonsynonymous substitutions observed for the early-stage histone H3 gene in brooding forcipulate sea stars was due to relaxed purifying selection resulting from a combination of (1) a possible change in expression pattern of the early-stage histone H3 gene associated with evolution of a nonfeeding larval form, (2) the high copy number of the early-stage gene (up to 500 copies: Cool et al. 1988), and (3) reduced effective population sizes in brooding versus nonbrooding lineages, as proposed previously for mitochondrial gene sequences (Foltz 2003; Foltz et al. 2004). They suggested that these possibilities could be tested by comparing substitution rates for the early-stage histone H3 gene with a paralogous and less-repetitive histone H3 gene.

To test these possibilities, we analyzed sequences for the early-stage histone H3 gene and a newly isolated ancient paralog of this gene that we named the "putative late-stage histone H3 gene" for 69 species of sea stars. Although the two paralogs differed in copy number, codon usage, nucleotide composition, and synonymous substitution rates, the excess nonsynonymous substitutions observed in brooding lineages was equal between the paralogs, thus refuting the prediction that copy number and expression differences between paralogs would have an effect on patterns of nonsynonymous substitutions in brooding lineages versus nonbrooding lineages. The excess nonsynonymous substitutions in brooding lineages for the early-stage histone H3 gene observed in the broader phylogenetic survey reported here were also smaller than seen by Foltz and Mah (2009), which may be due to either (1) sampling differences between the studies or (2) recent pseudogenization of the early-stage histone H3 gene in some lineages.

MATERIALS AND METHODS

DNA extractions, PCR reactions and cycle sequencing reactions were performed as in Foltz (2007) and Foltz et al. (2007). Primers for amplification and sequencing of the early-stage histone H3 gene were from Foltz and Mah (2009). The design of a partly internal forward primer to amplify the paralogous H3 gene that we termed the "putative late-stage histone H3 gene" was described in supporting information Table S1. Contigs were assembled and proofread in Sequencher v. 4.8 and then exported to ClustalX v. 1.8 (Thompson et al. 1997) for alignment. The early-stage sequences were trimmed at the 5′ end to start at codon position #12, yielding an alignment length of 318 nucleotides (see supporting information Table S1 for details). We obtained new early-stage histone H3 sequences from 42 sea star specimens and putative late-stage histone H3 sequences from 71 specimens representing 69 species (see supporting information Table S2 for specimen voucher numbers, where available, and GenBank accession numbers). Most of the sequences were from museum specimens rather than freshly collected material. Taxonomic identifications were by Mah (except for GenBank sequences). Although we sampled widely in two of the most taxon-rich orders of sea stars (the Forcipulatida and Valvatida), the other orders (Brisingida, Notomyotida, Paxillosida, Spinulosida, and Velatida) mostly lacked well-preserved museums specimens, lacked data on brooding habit or gave DNA extracts from which one or both genes could not be amplified and sequenced. Likewise, a few genera (*Sclerasterias*, *Paralophaster*, and *Solaster*) are overrepresented in the data set. However, by restricting the analysis to specimens for which sequence data were available for *both* genes, any effect of unbalanced taxonomic sampling should be equalized between the paralogs. As discussed below, equalizing the sequence data sets between genes required the exclusion of early-stage sequences for two brooding lineages (*Anasterias* and several species in the genus *Leptasterias*) that had been observed previously to have high frequencies of nonsynonymous substitutions, which made a direct comparison of parameter estimates between the present study and the earlier report more difficult.

Data on mode of reproduction were obtained primarily from McEdward and Miner (2001), supplemented by additional published sources and unpublished observations on museum specimens by Mah (see supporting information Table S2 for details). Two types of lecithotrophy have been reported for sea stars: brooded lecithotrophy and pelagic lecithotrophy. Because of the rarity of pelagic lecithotrophy in the present study ($<10\%$ of the species in supporting information Table S2 are documented to have pelagic lecithotrophic larvae) and because the presence or absence of brooded offspring was in some instances inferred from museum specimens, we analyzed mode of reproduction as a presence/absence character: lineages with brooded lecithotrophic larvae (sometimes referred to later as "brooder" or "brooding") versus lineages that lack brooded lecithotrophic larvae (sometimes called "nonbrooders" or "nonbrooding").

Histone sequences are generally thought to evolve so rapidly at synonymous codon positions that saturation effects have eliminated any phylogenetic signal at deep nodes (e.g., Rooney et al. 2002). Although the 71 early-stage histone H3 sequences and the 71 novel histone H3 sequences were reciprocally monophyletic with

>90% bootstrap support when rooted on 52 chordate histone H3 sequences (details not shown), bootstrap support for basal nodes within each of these clades was often <50% (support assayed by tree reconstruction with bootstrapping in RAxML-VI-HPC v. 7.0.4, with 450 bootstrap replicates determined automatically, Stamatakis et al. 2008). Our aims in the present study were not to use these two genes for phylogenetic reconstruction per se, but to compare the patterns of synonymous and nonsynonymous substitutions between them. Therefore, for tree-based analyses such as codeml we used an "apriori tree" based on mitochondrial 12S and 16S rDNA sequences (C. L. Mah and D. W. Foltz, in preparation—see supporting information Fig. S1). The Consel program of Shimodaira and Hasegawa (2001) was used to determine the relative support for the apriori tree versus the a posteriori tree obtained in RAxML, with site-specific log-likelihoods for the two topologies determined with the baseml program in PAML v. 4.2b (Yang 2007). The early-stage and novel histone H3 genes were also compared with two classes of histone H3 gene in sea urchins (early and late) via neighbor-joining tree construction using composite maximum-likelihood distances in MEGA v. 4 (Tamura et al. 2007). This tree was rooted on the same 52 chordate histone H3 sequences mentioned above. The codeml program in PAML was used to estimate synonymous (d_S) and nonsynonymous (d_N) substitution rates per site under various models and to test different d_N/d_S ratios with log-likelihood ratio test statistics ($2\Delta\log L$), with the topology of the apriori tree input to codeml as a user-specified tree. Other options and settings in the codeml analyses were as in Foltz (2003), namely: (1) F3X4 codon frequencies, (2) no variation in d_N/d_S among sites, (3) transition/transversion rate κ was estimated from the data, (4) γ shape parameter was fixed at ∞, and (5) no molecular clock. The possibility of gene conversion between the two clades was examined with the GENECONV program of Sawyer (1999), and possible differences in nucleotide site polymorphism between the two clades were tested via a Wilcoxon signed-rank test, where within-specimen polymorphism for the early-stage and novel histone H3 genes were determined from chromatograms in a manner analogous to that used by Foltz (2007) to assay nucleotide site diversity in sea stars for a single-copy nuclear gene partial sequence, ATP synthase β subunit. Nucleotide frequencies per codon position were calculated in MEGA v. 4, and the effective number of codons (N'_c) was calculated as per Novembre (2002), using the program ENCprime (http://www.eeb.ucla.edu/Faculty/Novembre/software/software.html). A relative rate test was performed using PhylTest v. 2 (http://www.homes.bio.psu.edu/people/Faculty/Nei/Lab/phyltest2.htm), using 52 chordate histone H3 sequences as the outgroup. Relative copy number of the early- and putative late-stage histone H3 genes was estimated using quantitative PCR (qPCR), as documented in supporting information Table S3.

RESULTS

As no published late-stage histone H3 sequences exist for sea stars, we compared the sequences in supporting information Table S2 with the limited number of genomic sequences from GenBank representing early-stage histone H3 sequences in sea urchins, plus the even more limited genomic sequences for the late-stage histone H3 gene in sea urchins (see supporting information Table S2 for details on GenBank accession numbers). When rooted on 52 chordate sequences (Fig. 1), the 71 early-stage histone H3 sequences in sea stars clustered with the homologous sequences in sea urchins with 83% bootstrap support, and the 71 novel histone H3 sequences in sea stars clustered with the three late-stage histone H3 sequences in sea urchins with 90% bootstrap support. The estimate from qPCR analysis of the relative copy excess of the early-stage gene versus the putative late-stage gene was 95.8 ± 12.2. The three nonbrooding species that were analyzed by qPCR (see supporting information Table S3) did not have relative copy numbers for the early- and late-stage genes that were demonstrably different from the three brooding species included in the analysis. The putative late-stage histone H3 sequences showed differences in nucleotide frequencies per codon position and codon usage from the early-stage histone H3 sequences. In particular, the putative late-stage histone H3 sequences had much lower %A and somewhat higher %C in the third codon position, and the effective number of codons (N'_c) was smaller in the putative late-stage sequences, when compared with early-stage histone H3 sequences in the same 69 species of sea star (Table 1). We also compared the sea star sequences with the corresponding sea urchin sequences described above. The early-stage histone H3 sequences in sea stars and in sea urchins were extremely similar to each other in nucleotide composition and effective numbers of codons. The putative late-stage histone H3 sequences in sea stars differed from the three genomic sea urchin late-stage histone H3 sequences in nucleotide composition at the third codon position, but had similar effective numbers of codons.

Possible difference in overall rate of substitution between the two histone H3 gene sequences was assessed by a relative rate test. The putative late-stage sequences were evolving on average 45% as fast as the early-stage sequences, a significant difference ($P < 0.0001$). The rate differences were attributable almost entirely to synonymous substitutions, as a relative rate test applied to nonsynonymous substitutions only showed no detectable rate heterogeneity between the two histone H3 genes (details not shown). Despite the apparent slow substitution rate in the putative late-stage histone H3 gene, nucleotide site polymorphism for the putative late-stage gene

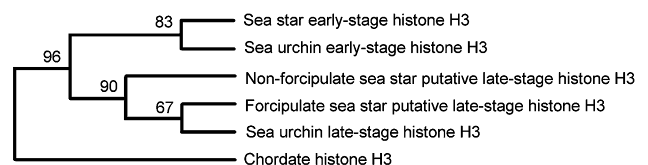

Fig. 1. Neighbor-joining tree showing bootstrap support as percentages for two classes of sea urchin histone H3 sequences and two classes of sea star histone H3 sequences, rooted on chordate sequences.

Table 1. Codon usage measured as effective number of codons, likelihood ratio test (2ΔlogL) of selection on synonymous codon substitutions (from Yang and Nielsen 2008) and nucleotide composition summaries for four groups of histone H3 sequences in echinoderms

Group	Effective no. codons	2ΔlogL (41 df)	First codon position %T	First codon position %C	First codon position %A	Second codon position %T	Second codon position %C	Second codon position %A	Third codon position %T	Third codon position %C	Third codon position %A
Average of 71 sea star early-stage H3 sequences	55.21	112.52***	9.7	30.1	28.2	23.7	25.4	29.3	15.3	34.4	21.0
Average of 71 sea star putative late-stage H3 sequences	36.09	259.30***	9.6	31.7	26.5	23.6	25.6	29.0	15.8	46.7	7.6
Average of 6 sea urchin early-stage H3 sequences	51.25	—	9.7	28.5	29.7	23.4	25.6	29.2	17.9	35.2	18.9
Average of 3 sea urchin late-stage H3 sequences	38.55	—	10.7	32.1	25.5	23.6	26.4	29.2	18.6	49.1	8.2

***$P<0.00001$

($1.08 \pm 0.11\%$) was significantly larger than for the early-stage gene ($0.06 \pm 0.03\%$), when analyzed by a Wilcoxon signed-rank test ($P<0.0001$).

Given that phylogenetic reconstruction with these relatively short sequences is problematic, this study aimed mostly to estimate and compare patterns of nucleotide substitution within and between paralogous copies of the histone H3 gene. To test that the two histone H3 genes were each evolving in an orthologous fashion, we tested for possible gene conversion events between the two genes. No significant conversion tracts were discovered (using default settings in GENECONV v. 1.81, except for gscale = 1) in either the amino acid or nucleotide sequences. The reciprocal monophyly of the early-stage and putative late-stage gene sequences in Fig. 1 provided some support for lack of gene conversion events between the two genes at the whole-sequence level. We also tested for nonorthology by comparing the topology of an "aposteriori" tree estimated from the 318 bp nucleotide alignment with the topology of an "apriori" tree derived from mitochondrial gene sequences, in which the early-stage and putative late-stage gene trees were reciprocally monophyletic and both had a topology that was identical to the mitochondrial gene tree (see supporting information Fig. S1 for details). Site-specific likelihoods were calculated separately in baseml using the GTR+γ model for the apriori and aposteriori trees and exported to the program CONSEL v. 0.1j. The mitochondrial-based apriori tree topology could not be rejected ($P>0.05$) using the approximately unbiased criterion of Shimodaira (2002), which provided additional support for orthologous evolution of both histone H3 genes.

Four models of increasing complexity for synonymous and nonsynonymous substitution rates were fit to the data. The simplest model (null model, column A in Table 2) fit one $d_N/d_S = \omega$ value to all branches of both genes. This model was rejected in favor of a more complex model (column B in Table 2) that fit different ω values to brooders (0.0790) and nonbrooders (0.0046). Model B in turn was nominally rejected in favor of model C, that fit different ω values to early-stage

Table 2. Summary statistics plus log-likelihood values (logL) for hypothesis testing, from the codeml program, showing transition/transversion rate ratios (κ) and nonsynonymous/synonymous substitution rate ratios ($d_N/d_S = \omega$) for various models that fit different ω values to brooding lineages (termed "brooders" below) and nonbrooding lineages ("nonbrooders") for the early-stage and late-stage histone H3 genes in sea stars

Quantity	Model: A, 1 ω value*	Model: B, 2 ω values	Model: C, 3 ω values	Model: D, 4 ω values
κ	2.57	2.58	2.58	2.58
ω	0.00825	0.0046 nonbrooders 0.0790 brooders	0.0025 early-stage, nonbrooders 0.0090 late-stage, nonbrooders 0.0789 early-stage+late-stage, brooders	0.0025 early-stage nonbrooders 0.0090 late-stage, nonbrooders 0.0837 early-stage, brooders 0.0743 late-stage, brooders
logL	−4624.65	−4604.34	−4601.24	−4601.22
2ΔlogL	—	40.62, $P<0.0001$	6.20, $P<0.02$	0.04, $P>0.05$

Test statistics (2ΔlogL) each have 1 degree of freedom and test models B, C, and D against the simpler model to the immediate left.
*Null hypothesis.

sequences in nonbrooding lineages, putative late-stage sequences in nonbrooding lineages, and early-stage+putative late-stage genes in brooding lineages (see also Figs. 2 and 3). Model C could not be rejected in favor of a more complex model that fit four different ω values specific to all combinations of gene × larval type. We concluded that ω values were similar between the two histone H3 genes for brooding lineages, but possibly heterogeneous between genes for nonbrooding lineages. The codeml analysis comparing models B and C in Table 2 was not confirmed by the nonparametric relative rate test on nonsynonymous substitutions mentioned above, and the significance of the parametric test disappeared if a correction for multiple testing was applied. The amino acid replacements inferred in codeml were shown in supporting information Fig. S1. There were no amino acid replacements fixed between the two histone H3 genes, and most replacements occurred near the tips of the tree rather than toward the root. Of the 30 inferred amino acid replacements, 19 involved codon #90 and valine ↔ isoleucine replacements. If this variable codon was excluded from the codeml analysis, the estimates of ω were reduced by approximately 50%, and the two ω-value model (B in Table 2) could not be rejected in favor of more complicated models with additional ω values. Alternatively, if the 36 species in supporting information Table S2 for which the presence/absence of brooding behavior was inferred indirectly from museum specimens were removed from the codeml analysis (together with the removal of one of a pair of replicate sequences for *Diplasterias brandti* and for *Diplasterias meridionalis*), the estimates of ω were also reduced, and the two simpler models (A and B in Table 2) were each rejected in favor of model C with $P<0.01$. The most robust conclusion from these analyses was that brooding versus nonbrooding was a significant source of variation in ω values, but the presence of a significant gene effect (early-stage vs. putative late stage) depended on which sites and sequences were included in the analysis.

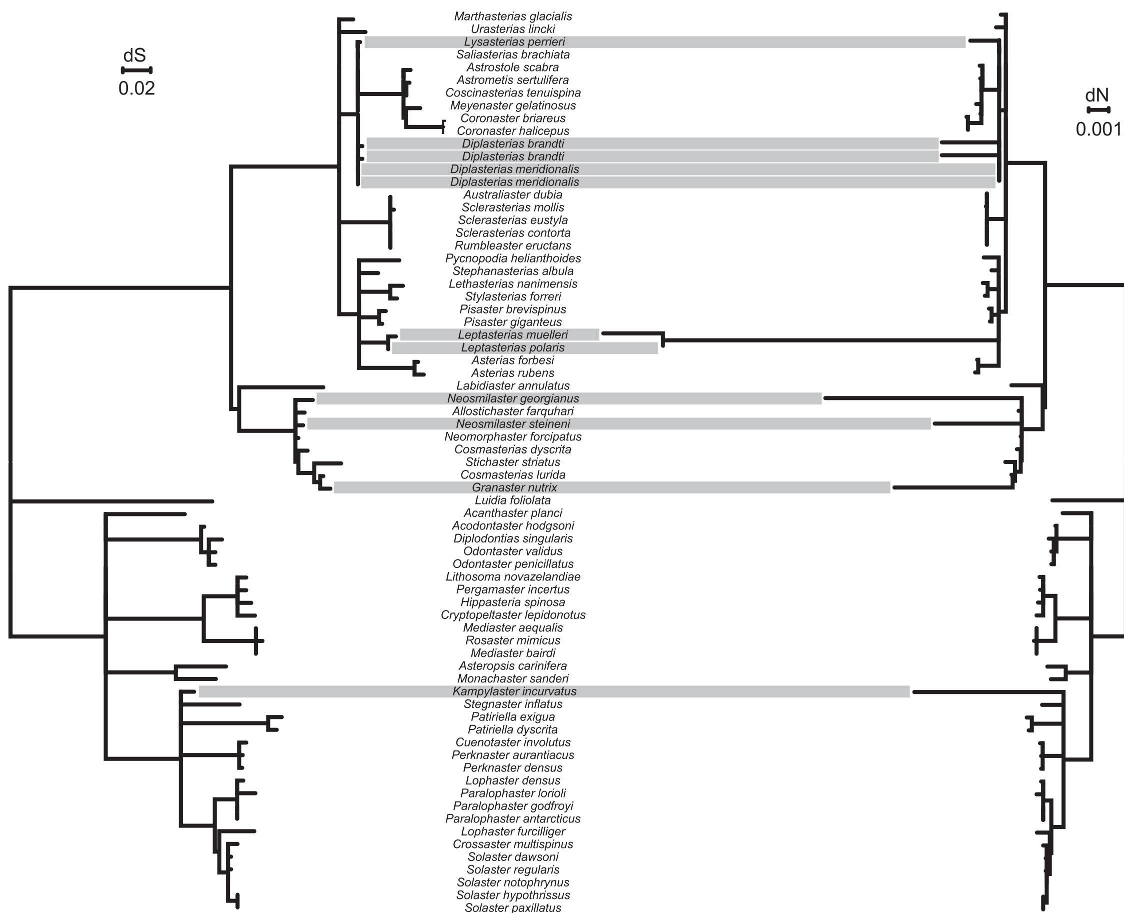

Fig. 2. Phylogenetic trees constructed using maximum-likelihood estimates of the expected number of synonymous (left) and nonsynonymous (right) substitutions per site for the early-stage histone H3 gene in sea stars, under a model (model C in Table 2) that allows brooding species (whose names are shaded) to have different rates of substitution than nonbrooding species. The scale bars show the estimated number of nucleotide substitutions per synonymous site (d_S) or per nonsynonymous site (d_N).

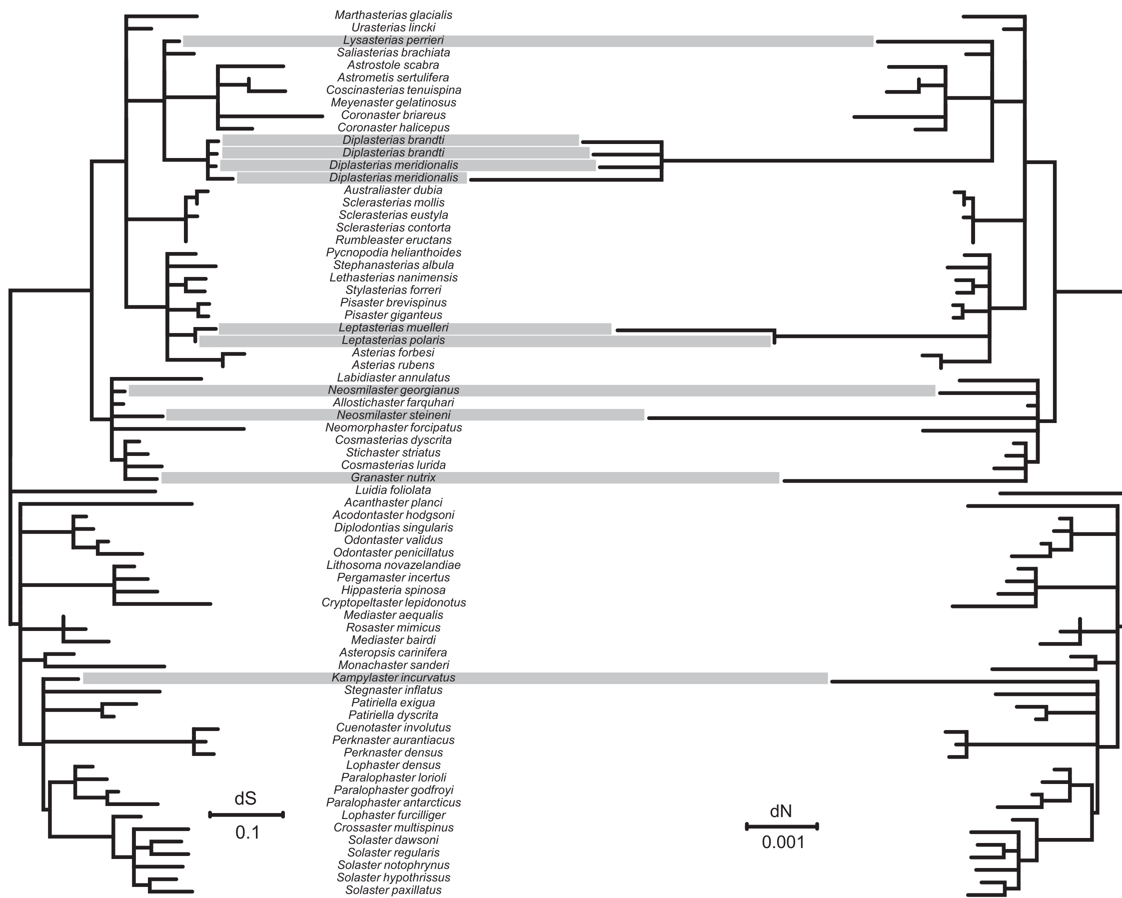

Fig. 3. Phylogenetic trees corresponding to those in Fig. 2, but for the putative late-stage histone H3 gene.

DISCUSSION

Maxson et al. (1987) suggested that early- and late-stage histone gene sets would be found to be characteristic of all echinoderm classes, and that these two genes diverged at least 200 Ma. The present study provided the first direct support for the existence of a late-stage histone gene in sea stars. Furthermore, because sea urchins and sea stars are part of an ancient radiation of echinoderms that is older than 475 Myr (Sprinkle and Guensburg 1997), the orthologous evolution of the two histone H3 genes observed here suggests that they are the product of an ancient duplication that is at least as old as the split between these two classes. In the sea urchin *Strongylocentrotus purpuratus* genome, two larval histone H3 gene families (early/α vs. late), a cleavage-stage (CS) histone H3 gene and an additional histone H3 gene common to all metazoans that is termed H3.3 have been reported (Marzluff et al. 2006). A fifth type of histone, testis-specific, does not have an H3 representative in the *S. purpuratus* genome. The early-stage histone H3 gene in sea urchins is part of a tandem array with the H1, H2a, H2b, and H4 genes, with several hundred copies of this gene array, mostly found in a single large tandem repetitive unit. In contrast, the late-stage histone H3 gene is present in eight dispersed copies in the *S. purpuratus* genome. The remaining histone H3 genes (CS and H3.3) are each single-copy in the *S. purpuratus* genome, and they also differ from the early- and late-stage genes in that they contain introns. Although an absolute estimate of the copy number of the putative late-stage histone H3 gene in sea stars was not obtainable from the relative quantitation via qPCR in supporting information Table S3, the relative copy excess for the early-stage gene (95.8 ± 12.2), plus the estimate of 500 copies of the early-stage histone H3 gene per haploid genome in forcipulate and valvatidan sea stars by Cool et al. (1988), suggest that there may be approximately 4–7 copies of the late-stage histone H3 gene in the sea star genome. This figure is close to the above-quoted value for *S. purpuratus*. The similarity of the novel histone H3 sequences in sea stars reported here to the sea urchin late-stage gene in (1) primary sequence, (2) estimated copy number, and (3) codon usage (Table 1), plus the absence of introns in the sequenced region, all support the identification of the novel H3 gene in sea stars as the ortholog of the late-stage histone H3 gene in sea urchins. Although the forcipulate sea star late-stage sequences

analyzed in Fig. 1 clustered with sea urchin late-stage sequences (rather than with nonforcipulate sea star late-stage sequences, as would be expected), the support for the forcipulate+sea urchin clade was relatively weak. This result may be another example of the difficulty of resolving basal relationships with a short and fast-evolving gene (as was noted in "Materials and Methods"), or it may reflect the fact that only three sea urchin late-stage histone H3 sequences are currently available in GenBank.

It is possible that the putative late-stage histone H3 gene in sea stars resembles the late-stage gene in sea urchins in having a dispersed repetitive structure and in being expressed at a low level during early embryogenesis and at a higher level after gastrulation (Marzluff et al. 2006). Given the long divergence time between these two echinoderm classes that was noted above and the known differences between sea urchins and sea stars in expression pattern for the early-stage histone H3 gene (Cool et al. 1988), these suggestions about the putative late-stage histone H3 gene should be verified by expression data in larval and adult sea star tissue. The only evidence in the present study that suggests a dispersed repetitive organization for the putative late-stage histone H3 gene in sea stars is the higher nucleotide site polymorphism in this gene when compared with the early-stage gene and also when compared with the polymorphism level for a single-copy nuclear gene partial sequence obtained from the same specimens (see supporting information Table S2 for details). This higher polymorphism level is consistent with the prediction that sequence homogenization via gene conversion is less effective in a low copy number dispersed repetitive gene family than in a high copy number tandem repetitive gene family, resulting in greater copy-to-copy polymorphism in the former (Graham 1995). While the lower polymorphism of the early-stage gene in sea stars might be due to the tandem array being located in a chromosomal region with a lower than average mutation rate, that suggestion conflicts with the greater rate of synonymous substitution observed in the early-stage compared with the putative late-stage gene (see also next paragraph).

Paralogs sometimes show asymmetric rates of divergence, a situation that has been attributed either to a change in selective regime in one or both paralogs (Conant and Wagner 2003; Kim and Yi 2006) or to differences in recombination and/or mutation rates between paralogs (Zhang and Kishino 2004; Seoighe and Scheffler 2005; Clément et al. 2006; Studer and Robinson-Rechavi 2009). Here, rate heterogeneity between paralogs was restricted to synonymous substitutions. Such a result could be due to either a lower mutation rate for the putative late-stage gene versus the early-stage gene, or to more stringent selection on synonymous substitutions in the putative late-stage gene, owing to the greater codon usage bias observed in the putative late-stage gene. The extent to which codon usage bias reflects mutational versus selective processes is controversial (e.g., Duret 2002; Hershberg and Petrov 2008). We tested the relative influence of mutation and selection on codon usage in the two paralogs with the approach of Yang and Nielsen (2008). An approximate likelihood ratio test ($2\Delta\log L$) comparing a mutation-bias only model (FMut/Sel0) against a mutation-bias+selection model (FMut/Sel) that was constructed in codeml v. 4.2b with 41 degrees of freedom found evidence of selection on synonymous codons in both paralogs (Table 1). However, estimating ω under the FMut/Sel model did not change the results in Table 2 appreciably, and mostly worked to equalize ω values between the early-stage and putative late-stage genes (details not shown). There was no evidence for adaptive evolution in either gene; the consensus amino acid sequence inferred from each gene was identical to the consensus H3 sequence in *S. purpuratus* (Marzluff et al. 2006). Except for the variation at amino acid position #90 between isoleucine and valine (which is also variable among late-stage histone H3 gene copies in *S. purpuratus*), all of the variation occurred at the tips of the tree, at the species or genus level (see supporting information Fig. S1). This phylogenetic distribution suggests that the nonsynonymous substitutions observed in the two histone H3 genes in sea stars are mostly recent and mildly deleterious mutations.

Foltz and Mah (2009) hypothesized that the excess nonsynonymous substitutions observed for the early-stage histone H3 gene in brooding lineages versus nonbrooding lineages was due to some combination of (1) relaxed selection in brooding lineages owing to change in expression pattern associated with evolution of a nonfeeding larval form, (2) a higher effective mutation rate or relaxed selection associated with the high copy number and tandem-repetitive structure of the early-stage gene, and (3) relaxed selection in brooding lineages owing to reduced effective population sizes, as seen previously for mitochondrial gene sequences (Foltz 2003; Foltz et al. 2004). The observation here that the $\omega = d_N/d_S$ ratio in brooding sea star lineages is nearly identical between the paralogous histone H3 genes is inconsistent with suggestions (1) and (2), leaving a demographic explanation (reduced dispersal and reduced long-term effective population size in lineages that lack a pelagic larval form) as the most likely explanation. The ω for brooding lineages observed here (approximately 0.08) is much lower than the corresponding value (0.33) reported by Foltz and Mah (2009). Some of the difference between the studies may reflect the less intensive sampling here of brooding clades like the genus *Leptasterias*. Less sampling means that most of the brooding lineages in supporting information Fig. S1 are represented by single specimens, whose subtending branches include an unknown but potentially long period of time before the evolution of a brooding mode of reproduction. This uncertainty would tend to equalize ω values between brooding and nonbrooding lineages (Foltz and Mah 2009), and thus make it more difficult to reject the null hypothesis in Table 1. This sampling design

also makes it difficult to compare ω values obtained in the present study to earlier analyses of mitochondrial protein-coding genes. In addition, many (9 of 16) of the nonsynonymous substitutions inferred in the earlier study occurred in either *Anasterias antarctica* or the *Leptasterias aequalis/hexactis* clade, and those nine substitutions included all of the observed trans-species substitutions. Unfortunately, putative late-stage histone H3 sequences could not be obtained for either *A. antarctica* or the *L. aequalis/hexactis* clade. The exclusion of these taxa from the present study may also have lowered the overall ω value for the early-stage gene in brooding lineages, particularly if the concentration of nonsynonymous substitutions in particular clades is due to recent pseudogenization of the early-stage gene (see discussion in Foltz and Mah 2009). Three instances of putative pseudogenes in early-stage histone H3 sequences have been observed: (1) the nonbrooding *Paralophaster lorioli* in supporting information Table S2 has a deletion of codons 24–33, (2) *Goniopecten demonstrans* (GQ288605) has a frame-shifting deletion of codons 63–71, and the first two bases of codon 72, (3) *L. aleutica* (EU707618) and *Leptasterias camtschatica* (EU707620), two recently diverged sister species, share a tandem duplication of codons 25–28. All of these species (except *P. lorioli*) lack data on larval type and thus were not included in the analyses here. Whether the apparent loss of functional gene copies for the early-stage histone H3 gene in these lineages is compensated by altered expression of other copies of the histone H3 gene (e.g., orphon genes) is currently unknown. Ultimately, as with other aspects of the present study, data on expression levels for the early-stage and putative late-stage gene in different larval and adult stages are required to fully understand the evolution of these sequences.

Acknowledgments

We thank Dr. Robert Van Syoc for arranging loan of material from the California Academy of Sciences and Dr. Mark A. Batzer for suggestions. Jerilyn A. Walker and Dr. Scott W. Herke provided assistance with the qPCR analyses. James W. Halk, Tap Bui and Caitlyn Blanchard assisted in the laboratory. Financial support was provided by NSF grant OPP-0631245 to C. L. Mah and an award from the Office of Research and Economic Development at Louisiana State University to D. W. Foltz.

REFERENCES

Arndt, A., and Smith, M. J. 1998. Genetic diversity and population structure in two species of sea cucumber: differing patterns according to mode of development. *Mol. Ecol.* 7: 1053–1064.

Ayre, D. J., Minchinton, T. E., and Perrin, C. 2009. Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Mol. Ecol.* 18: 1887–1903.

Clément, Y., Tavares, R., and Marais, G. A. B. 2006. Does lack of recombination enhance asymmetric evolution among duplicate genes? Insights from the *Drosophila melanogaster* genome. *Gene* 385: 89–95.

Conant, G. C., and Wagner, A. 2003. Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13: 2052–2058.

Cool, D., Banfield, D., Honda, B. M., and Smith, M. J. 1988. Histone genes in three sea star species: cluster arrangement, transcriptional polarity, and analyses of the flanking regions of H3 and H4 genes. *J. Mol. Evol.* 27: 36–44.

Duret, L. 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12: 640–649.

Foltz, D. W. 2003. Invertebrate species with nonpelagic larvae have elevated levels of nonsynonymous substitutions and reduced nucleotide diversities. *J. Mol. Evol.* 57: 607–612.

Foltz, D. W. 2007. An ancient repeat sequence in the ATP synthase β subunit gene of forcipulate sea stars. *J. Mol. Evol.* 65: 564–573.

Foltz, D. W., Hrincevich, A. W., and Rocha-Olivares, A. 2004. Apparent selection intensity for the cytochrome oxidase subunit I gene varies with mode of reproduction in echinoderms. *Genetica* 122: 115–125.

Foltz, D. W., and Mah, C. L. 2009. Recent relaxation of purifying selection on the tandem-repetitive early-stage histone H3 gene in brooding sea stars. *Mar. Genomics* 2: 113–118.

Foltz, D. W., Nguyen, A. T., Nguyen, I., and Kiger, J. R. 2007. Primers for the amplification of nuclear introns in sea stars of the family Asteriidae. *Mol. Ecol. Notes* 7: 874–876.

Galtier, N., Nabholz, B., Glémin, S., and Hurst, G. D. D. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18: 4541–4550.

Graham, G. J. 1995. Tandem genes and clustered genes. *J. Theor. Biol.* 175: 71–87.

Hershberg, R., and Petrov, D. A. 2008. Selection on codon bias. *Annu. Rev. Genet.* 42: 287–299.

Johnson, M. S., and Black, R. 2006. Effects of mode of reproduction on genetic divergence over large spatial and temporal scales in intertidal snails of the genus *Bembicium* Philippi (Gastropoda: Littorinidae). *Biol. J. Linn. Soc.* 89: 689–704.

Kim, S. H., and Yi, S. V. 2006. Correlated asymmetry of sequence and functional divergence between duplicate proteins of *Saccharomyces cerevisiae*. *Mol. Biol. Evol.* 23: 1068–1075.

Lee, H. J., and Boulding, E. G. 2009. Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods: the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. *Mol. Ecol.* 18: 2165–2184.

Marzluff, W. F., Sakallah, S., and Kelkar, H. 2006. The sea urchin histone gene complement. *Dev. Biol.* 300: 308–320.

Maxson, R., Mohun, T., Gormezano, G., and Kedes, L. 1987. Evolution of late H2A, H2B, and H4 histone genes of the sea urchin, *Strongylocentrotus purpuratus*. *Nucleic Acids Res.* 15: 10569–10582.

McEdward, L. R., and Miner, B. G. 2001. Larval and life—cycle patterns in echinoderms. *Can. J. Zool.* 79: 1125–1170.

Miller, K. J., and Ayre, D. J. 2008. Population structure is not a simple function of reproductive mode and larval type: insights from tropical corals. *J. Anim. Ecol.* 77: 713–724.

Novembre, J. A. 2002. Accounting for background nucleotide composition when measuring codon usage bias. *Mol. Biol. Evol.* 19: 1390–1394.

Pechenik, J. A. 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Mar. Ecol. Prog. Ser.* 177: 269–297.

Rooney, A. P., Piontkivska, H., and Nei, M. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol. Biol. Evol.* 19: 68–75.

Sawyer, S. A. 1999. *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*. Department of Mathematics, Washington University, St. Louis, http://www.math.wustl.edu/ ~ sawyer.

Seoighe, C., and Scheffler, K. 2005. Very low power to detect asymmetric divergence of duplicated genes. In A. McLysaght and D. H. Huson (eds.). *Comparative Genomics: RECOMB 2005 International Workshop, RCG 2005, Dublin, Ireland, September 18–20, 2005, Proceedings (Lecture Notes in Computer Science/Lecture Notes in Bioinformatics)*. Springer-Verlag, Berlin, pp. 142–152.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51: 492–508.

Shimodaira, H., and Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.

Sprinkle, J., and Guensburg, T. E. 1997. Early radiation of echinoderms. *Paleontol. Soc. Pap.* 3: 205–224.

Stamatakis, A., Hoover, P., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Syst. Biol.* 57: 758–771.
Studer, R. A., and Robinson-Rechavi, M. 2009. How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet.* 25: 210–216.
Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596–1599.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876–4882.
Ward, R. D., Holmes, B. H., and O'Hara, T. D. 2008. DNA barcoding discriminates echinoderm species. *Mol. Ecol. Res.* 8: 1202–1211.
Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586–1591.
Yang, Z., and Nielsen, R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.* 25: 568–579.
Zhang, Z., and Kishino, H. 2004. Genomic background predicts the fate of duplicated genes: evidence from the yeast genome. *Genetics* 166: 1995–1999.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Apriori tree from mitochondrial sequences used for analysis of histone H3 sequence evolution in sea stars.

Table S1. Primer design strategy for hypothesized late-stage histone H3 gene in sea stars.

Table S2. Specimen voucher and GenBank accession numbers for histone H3 sequences.

Table S3. Quantitative PCR protocol.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.