Duplication and Selection on Abalone Sperm Lysin in an Allopatric Population

Nathaniel L. Clark,[1,2] *Geoffrey D. Findlay,*[1,2] *Xianhua Yi, Michael J. MacCoss, and Willie J. Swanson*

Department of Genome Sciences, University of Washington

While gene duplication is a major source of evolutionary novelty, the importance of this process in reproductive protein evolution has not been widely investigated. Here, we report the first known case of gene duplication of abalone sperm lysin in an allopatric subspecies found in the Eastern Atlantic, *Haliotis tuberculata coccinea*. Mass spectrometry identified both copies of the lysin protein in testis tissue, and 3-dimensional structural modeling suggests that both proteins remain functional. We also detected positive selection acting on both paralogs after duplication and found evidence of a recent selective sweep. Because *H. t. coccinea* occurs in geographic isolation from other abalone species, these findings suggest that the evolution of lysin is not driven to create reproductive barriers to unfit hybrid formation with an overlapping species. Instead, sexual selection or sexual conflict acting during abalone fertilization could be responsible for the recent positive selection on this protein. The presence of multiple, rapidly evolving lysin genes in *H. tuberculata* presents an opportunity to study the early stages of diversification of a protein whose function is well understood.

Introduction

Gene duplication is a major evolutionary force capable of creating genes with novel functions (Lynch and Conery 2000; Lynch and Katju 2004). After a gene segment is duplicated, 3 outcomes are possible (Prince and Pickett 2002). Most commonly, deleterious mutations accumulate in 1 paralog, rendering that copy nonfunctional. A second possibility is neofunctionalization, where one paralog retains the gene's ancestral function and the other evolves a new function. Alternatively, subfunctionalization can occur, in which paralogs retain some aspect of their ancestral function but specialize in different parts of the same biological process. Neofunctionalization and subfunctionalization, although relatively rare, allow for the evolution of biological novelty.

Studies across diverse taxa have revealed that reproductive proteins often evolve rapidly; yet, the importance of gene duplication in the evolution of these proteins has received limited investigation (Armbrust and Galindo 2001; Frohlich et al. 2001; Holloway and Begun 2004; Torgerson and Singh 2004). Ancient gene duplication and subfunctionalization are evident in 2 acrosomal sperm proteins of the abalone (genus *Haliotis*), an externally fertilizing marine mollusk. One protein, sp18, is thought to mediate sperm–egg fusion by interacting with an unknown receptor on the egg plasma membrane (Swanson and Vacquier 1995b). The second acrosomal protein, lysin, interacts with the vitelline envelope receptor for lysin (VERL) to dissolve the vitelline envelope surrounding abalone eggs (Lewis et al. 1982; Swanson and Vacquier 1997). Although the lysin and sp18 proteins are unalignable at the level of their primary amino acid sequences, their similar molecular weights and 3-dimensional structures and conserved pattern of exons and introns indicate that they are paralogs (Vacquier et al. 1997; Kresge et al. 2001). Presumably, an ancestral sperm protein was responsible for both egg coat dissolution and egg membrane fusion. After duplication, each paralog specialized to perform one of these roles.

Sp18, lysin, and VERL evolve adaptively under positive selection, yet the selective pressures remain unknown (Lee et al. 1995; Metz et al. 1998; Galindo et al. 2003). Several hypotheses could explain this rapid evolution, including sperm competition, sexual conflict, pathogen avoidance, and reinforcement (Swanson and Vacquier 2002). Under reinforcement, fertilization proteins would be selected to diversify in order to prevent hybridization between species that overlap in geographic range and spawning time (Noor 1999; Marshall et al. 2002; Coyne and Orr 2004). This hypothesis is of particular interest for lysin and VERL because abalone lack courtship behavior and fertilize externally. Thus, gamete recognition proteins are largely responsible for mediating mate choice. If interspecific fertilization resulted in hybrids that were inviable or unfit inviable or unfit, diversifying selection may act on these proteins to minimize the production of hybrid offspring. Importantly, this hypothesis can only explain positive selection on the gamete proteins of sympatric species. If reinforcement were the sole force driving these proteins' evolution, we would not expect to see positive selection on these proteins in isolated species.

We studied the evolution of lysin in *Haliotis tuberculata*, an abalone species found in the eastern Atlantic and Mediterranean. While sequencing the gene, we discovered the first known case of lysin gene duplication. We experimentally verified the expression of each protein in testis tissue, and evolutionary analyses showed that both copies have been subject to positive selection. These diversification is consistent with subfunctionalization for different regions, alleles, or copies of VERL. To understand the selective forces that have shaped lysin's recent evolutionary history, we performed a polymorphism survey for one copy of lysin in an *H. tuberculata* subspecies. This population genetic data suggested a recent selective sweep. Because this subspecies is geographically isolated, a pressure other than reinforcement has likely driven the gene's evolution in this population.

Materials and Methods

Sampling and Genotyping

Eleven *Haliotis tuberculata coccinea* individuals were sampled from the Azores, Portugal. Three male individuals

[1] Present address: Box 355065, 1705 NE Pacific Street, Seattle, WA.

[2] Equally contributed to this work.

Key words: lysin, gene duplication, positive selection, *Haliotis tuberculata*, reinforcement, sexual conflict.

E-mail: clarknl@u.washington.edu; gfindlay@u.washington.edu.

Mol. Biol. Evol. 24(9):2081–2090. 2007
doi:10.1093/molbev/msm137
Advance Access publication July 13, 2007

of *Haliotis tuberculata tuberculata* were sampled off the Atlantic coast of Roscoff, France. Tissue for *Haliotis pustulata* was kindly provided by V. Vacquier. Total DNA was isolated using the Puregene DNA purification kit (Gentra Systems, Minneapolis, MN). Genotypes were determined by polymerase chain reaction (PCR) amplification from genomic DNA followed by either cloning or direct sequencing. Lysin PCR primers were designed from published transcripts (GenBank accession numbers L26280 and L26284) and then from empirically determined sequences. The entire lysin gene was sequenced, including all exons and introns. Primers and conditions for lysin PCR products are available from the authors upon request. The neutral locus *rpL5* is an intron from an abalone homolog of the oyster ribosomal protein L5 (GenBank accession number CAD91421.1), and it was amplified using the following primers based on sequence from a *Haliotis corrugata* ovary cDNA library (Aagaard et al. 2006)—Forward: GGCTGCATATTCCCATGAGT, Reverse: CTGGTTTGCCATCCTCATCT. Intron–exon structure was assumed identical to its vertebrate homolog, and primers were placed within the boundaries of conserved exons. A second neutral locus, the fifth intron of the *hemocyanin 1* gene, was amplified using the following primers designed from GenBank entry AJ252741—Forward: TAGTAGTGGGGGCGGGATAG, Reverse: CATGTTGGCAGCTCTTAACG. Mitochondrial *cytochrome oxidase 1* was amplified using primers specific for the species from GenBank entry AY377729—Forward: GGATCTGATCAGGGCTCCTT, Reverse: GCTGGGTCAAAGAATGAGGT. Single-band PCR products or cloned fragments were sequenced on an ABI 3100 using Big Dye v.3.1 (Applied Biosystems, Foster City, CA). Lysin sequences were analyzed using Sequencher 4.2 (Gene Codes, Ann Arbor, MI). Neutral locus sequences were aligned and analyzed using phred, phrap, and consed (Ewing et al. 1998; Gordon et al. 1998); neutral locus polymorphism analysis was aided by polyphred (Stephens et al. 2006). An additional neutral locus, *histone H3*, was used to estimate divergence between *H. tuberculata* and *H. pustulata*; sequences used were GenBank accession number AY070145 and AY923954.

Expression Analysis

Total RNA was isolated from ethanol-preserved testis tissue from *H. t. tuberculata* using TRIzol (Invitrogen, Carlsbad, CA). First-strand cDNA was synthesized with the SuperScript III reverse transcriptase–polymerase chain reaction (RT–PCR) kit (Invitrogen) according to the manufacturer's instructions. This cDNA was used as template DNA in standard PCR reactions, and the products were TA cloned into the TOPO pCR 2.1 vector (Invitrogen). Plasmid DNA was isolated from several positive clones and sequenced with M13 primers, and sequences were analyzed in Sequencher 4.2.

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

Haliotis tuberculata tuberculata testis tissue was homogenized in 1% sodium dodecyl sulfate at 70 °C. Several dilutions of saturated protein sample were loaded onto a 15% polyacrylamide gel and run for 8 h at 24 mA. The gel was stained with Coomassie Blue, and 3 prominent bands migrating between 10–20 kDa were excised and sent to the Stanford University Protein and Nucleic Acid Biotechnology Facility, where mass mapping was used to confirm their identities (Jensen et al. 1997). Internal Edman sequencing of tryptic peptides was performed for the band at 17.5 kDa, yielding 2 sequences, EIAQDFKTDLR and EKYDLTPSQAK, which show partial homology to the rapidly evolving sp18 (Swanson and Vacquier 1995a). The gel presented in figure 2*A* was run with the same samples and conditions, but proteins were visualized by silver staining.

Mass Spectrometry

Soluble proteins from the same *H. t. tuberculata* testis sample that showed expression of both copies of lysin were obtained by homogenizing ~50 mg of tissue in 50 mM ammonium bicarbonate and keeping the supernatant after centrifugation. Protein concentration was assessed using a bicinchoninic acid protein assay kit (Pierce, Rockford, IL). After the protein assay, 200 μg of protein were denatured and digested with trypsin as previously described (Aagaard et al. 2006). A 25-μg aliquot of the protein digest was analyzed by microcapillary liquid chromatography–tandem mass spectrometry. The peptides were separated using a 75-μm internal diameter capillary column packed with 40 cm of reversed phase chromatography material (C12, 4 μm, 90 Å; Phenomonex, Torrance, CA). The effluent from the column was ionized, and peptides were directed through the atmospheric pressure interface of an LTQ ion trap mass spectrometer using a distal voltage of 2.4 kV applied directly to the solvent using a gold wire. The mass spectrometer was configured to acquire MS/MS spectra for 9 *m*/*z* regions cyclically throughout the chromatographic separation. The 9 *m*/*z* regions were calculated in silico from tryptic peptides predicted from the inferred protein sequences for each copy of lysin (4 peptides from copy 1 and 5 peptides from copy 2). The SEQUEST program (Eng et al. 1994) was then used to search the acquired MS/MS spectra against a database containing the protein sequences deduced for *H. t. tuberculata* lysin, a 6-frame translation of *Haliotis*-expressed sequence tag sequences obtained in the Swanson laboratory, all *Haliotis* sequences present in GenBank, common contaminants, and a shuffled decoy database.

Lysin Phylogeny

Divergent lysin coding sequences were aligned using a translated amino acid alignment as a guide. The amino acid alignment was made in ClustalW (Thompson et al. 1994) and codons were aligned in Se-Al v2.0a11 (Chenna et al. 2003). The gene phylogeny was inferred using maximum likelihood in "dnaml" of the PHYLIP package (Felsenstein 1989). Branch support was calculated from 1,000 bootstrap replicates of the data.

Structural Modeling and Hypotheses

We tested various structural hypotheses on threaded 3-dimensional structural models made using SwissModel

(Schwede et al. 2003). Images in figure 1*C* were made using the solved crystal structure for *Haliotis rufescens* lysin (Protein Data Bank ID 2LIS) so that the N-terminal deletions could be visualized. Images were created with MacPyMOL (DeLano 2002). The degree of solvent exposure was determined using GETAREA 1.1, and buried core residues were inferred as those with 10% or less solvent exposure (Franzkiewicz and Braun 1998). Amino acid changes were mapped to phylogenetic branches using "codeml" of the PAML package (Yang 1997). Hypotheses involving categories of sites were tested using Fisher's exact test. Hypotheses involving measured structural statistics (solvent exposure and neighbor tests) were tested by random permutations in a program written by N.L.C. For the neighbor test, it was important to restrict analysis to surface residues because earlier tests in this study found a nonrandom pattern of divergence between surface and core sites. Therefore, the observed statistic and random sets of residues involved only surface sites.

Selective Pressure Acting on Lineages

Estimates of d_N/d_S on lineages were made using "codeml" of PAML (Yang 1997; Yang et al. 2000). To detect positive selection on the various lineages for lysin in figure 1*A*, a branch model in which each branch d_N/d_S was uniquely estimated was compared with a model with one ratio for all branches using a likelihood ratio test (LRT). In this test, 2 times the negative difference in likelihoods between the null and alternative models is approximated by the chi-square distribution, with the number of degrees of freedom (df) in the test equaling the difference in the number of parameters between the 2 models. To detect positive selection specifically after gene duplication, a d_N/d_S estimate for all postduplication branches together was made by constraining those branches to have the same value. This postduplication model was compared with a null model in which those branch values were set equal to 1, the neutral expectation for d_N/d_S. To test for specific sites in the lysin protein that have evolved under positive selection since duplication, the branch-site model was used to partition the codons into classes that adopt different d_N/d_S values (Zhang et al. 2005). Two of these codon classes can adopt $d_N/d_S > 1$, indicating positive selection. This model is then compared with an LRT to a nested null model in which d_N/d_S is constrained to values between 0 and 1 for all codon classes. To assess significance of the LRT for the branch-site models, we used a chi-square distribution with one df, a conservative measure (Zhang et al. 2005).

Polymorphism and Tests of Selection

We used DnaSP to calculate summary statistics, estimate population parameters, and conduct polymorphism-based tests of selection (Rozas et al. 2003). We used sequences from *H. pustulata* for tests and statistics requiring an outgroup. *P* values for Tajima's and Fu and Li's tests were determined from coalescent simulations of a neutral model under locus-specific parameter estimates. Simulations were also performed under the conservative assumption of no recombination. Tests that rejected neutrality under recombination also did so under no recombination. For *histone H3*, the maximum likelihood estimate of d_S between *H. tuberculata* and *H. pustulata* was made by codeml (Yang 1997). Because the alignment was short (109 aligned codons), codon equilibrium frequencies were estimated from the product of nucleotide frequencies at each codon position (option F3 × 4). An estimate using the method of Nei and Gojobori (1986) was essentially the same. F_{ST} was estimated with the pairwise difference method in Arlequin, and a *P* value was computed using 100,000 permutations (Schneider et al. 2000).

Results

Lysin Has Duplicated in *H. tuberculata*

While sequencing cloned PCR products from genomic lysin for a polymorphism survey, we discovered that all *H. t. coccinea* individuals in our sample had 4 lysin alleles. The *H. tuberculata* karyotype is diploid with 14 pairs of homologous chromosomes (Arai and Wilkins 1986) and 2 other nuclear loci, *rpL5* and *hemocyanin 1*, each showed 2 alleles per individual. For each individual sequenced for lysin, the 4 alleles grouped into related pairs. Pairs were distinguished from each other by coding polymorphisms and intronic features such as indels and the presence or absence of polymorphic microsatellite repeats. Paired sequences were assumed to be allelic variants of the same gene because they showed 99.94% identity and no coding polymorphism. Identity between paralogs was 88% in coding regions. Thus, *H. tuberculata* harbors the first known case of lysin gene duplication.

We next investigated whether both lysin loci were transcribed and could encode full-length lysin proteins. Lacking testis tissue for *H. t. coccinea*, we isolated RNA from ethanol-preserved testis samples from a closely related subspecies from Roscoff, France, *H. t. tuberculata*. Primers located at both ends of the gene were used to amplify both copies of the gene from first-strand cDNA, which were then cloned and sequenced. Sequences were assigned to a locus based on similarity to the *H. t. coccinea* genomic sequences, and for one individual, we found at least 4 clones for each locus. No polymorphisms were observed between clones for these cDNA sequences, so each locus is represented by one coding sequence. Conceptual translation revealed that both loci encode properly spliced messages without premature stop codons and containing predicted signal sequences (fig. 1*A*). We defined copy 1 as the previously identified lysin in *H. tuberculata*, which has a 9-amino acid deletion just after the signal sequence (Lee et al. 1995). The novel copy 2 has a 2-residue deletion in this same region. The full coding sequence of copy 1 was determined in both subspecies, *H. t. coccinea* and *H. t. tuberculata*. Because RNA was available only for *H. t. tuberculata*, the coding sequence for copy 2 was determined in this subspecies. We sequenced over 2 kb of genomic DNA for this copy in *H. t. coccinea*, confirming that duplication preceded subspeciation (fig. 1*B*).

Comparing the paralogs of *H. t. tuberculata* revealed that among 143 aligned sites, the predicted proteins are 83% identical; of the 24 amino acid differences, 11 are

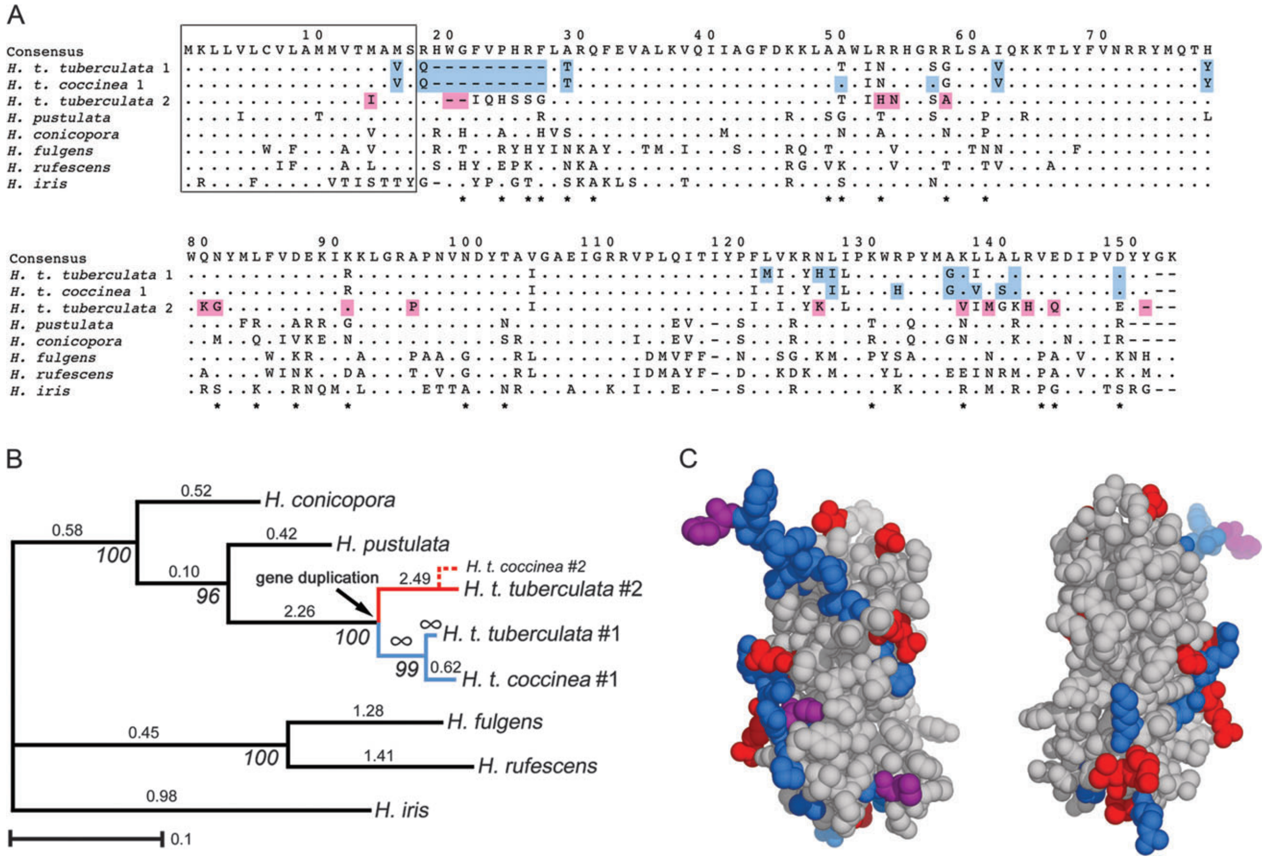
A
10 20 30 40 50 60 70
Consensus MKLLVLCVLAMMVTMAMSRHWGFVPHRFLARQFEVALKVQIIAGFDKKLAAWLRRHGRRLSAIQKKTLYFVNRRYMQTH
H. t. tuberculata 1 ................V.Q---------.T....................T.IN...SG...V...............Y
H. t. coccinea 1 ................V.Q---------.T......................IN....G...V...............Y
H. t. tuberculata 2 ..............I.....--IQHSSG......................T.IHN..SA....................
H. pustulata ....I.....T................R..................R..SG..T...S...P..R.............L
H. conicopora ..............V....R.H..A..HVS...........M........N..A....N..P.................
H. fulgens ......W.F...A.V....R.T..RYHYINKAY..TM.I...S...RQ.T....V......TNN....F..........
H. rufescens .......IF...A.L....S.HY.EPK..NKA..............RG.VK...V...T..TV...A............
H. iris .R...F.....VTISTTYG-..YP.GT..SKAKLS...T.......R...S......N.....................
* * ** * * ** * * *
80 90 100 110 120 130 140 150
Consensus WQNYMLFVDEKIKKLGRAPNVNDYTAVGAEIGRRVPLQITIYPFLVKRNLIPKWRPYMAKLLALRVEDIPVDYYGK
H. t. tuberculata 1 ............R.............I................IMI.YHIL.......G.I.............--
H. t. coccinea 1 ............R.............I................I.I.Y.IL...H...G.V.S...........--
H. t. tuberculata 2 .KG..............P........I................I.I.YK.L........VIMGKH.Q....E.---
H. pustulata ....FR..ARR.G...........N............EV.-.S...R.....T..Q...N...R.......R----
H. conicopora ..M..Q.IVKE.N...........SR........I..EV.-.S...R.....R..Q..GN...K...N..IR----
H. fulgens ......W.KR...A...PAA.G..RL.........DMVFF-.N..SG.KM..PYSA.....N...PA.V..KNH..
H. rufescens .A....WINK..DA...T.V.G..RL........IDMAYF-.D..KDK.M...YL...EEINRM.PA.V..K.M..
H. iris .RS..K..RNQM.L....ETTA..NR...A..K.I..E..-.S...R.......K....R.M.R.PG...TSRG--
* * * * * * * * ** *
B
0.52 H. conicopora
0.58
100
0.42 H. pustulata
0.10
96
gene duplication
2.49 H. t. coccinea #2
H. t. tuberculata #2
2.26
100
∞ H. t. tuberculata #1
∞
99
0.62 H. t. coccinea #1
1.28 H. fulgens
0.45
100
1.41 H. rufescens
0.98 H. iris
0.1
C

FIG. 1.—Duplication and divergence of *Haliotis tuberculata* lysin. (*A*) Aligned protein sequences for *H. tuberculata* lysins and those of several other species. Blue and red shading indicates lineage-specific substitutions inferred to have occurred after gene duplication of the ancestral *H. tuberculata* protein. The box indicates the cleaved signal sequence. Asterisks indicate sites found to be under selection in other abalone species (Yang et al. 2000). (*B*) Maximum likelihood phylogeny of lysin-coding DNA sequences. Duplication predates the subspeciation event in *H. tuberculata* and is indicated by an arrow. The position of *Haliotis tuberculata coccinea* copy 2 is shown with a dashed line because its entire sequence was not determined. Bootstrap support is shown in italics beneath each node. Branch-specific d_N/d_S ratios are shown above each branch; values of ∞ occur when no synonymous substitutions are inferred. (*C*) Front and rear views of a structural model of lysin. Lineage-specific substitutions are shown for each copy of lysin: blue resides, substitution in copy 1; red residues, substitution in copy 2; and purple residues, substitution in both copies.

nonconservative. The divergence between the 2 copies is particularly pronounced around the C-terminus, a region of lysin shown to be involved in species-specific interactions in other *Haliotis* species (Lyon and Vacquier 1999). Therefore, both copies of lysin are expressed as mRNA in abalone testis and encode full-length messages.

The predicted proteins encoded by the 2 copies of lysin differ in molecular weight by only 500 Da and were not resolved by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE; fig. 2*A*). Thus, we used MS to test whether both copies of lysin are present as proteins in the same sample of *H. t. tuberculata* testis tissue from which the transcripts were found. Because the 2 predicted proteins differ at various sites (fig. 1*A*), we reasoned that digestion with trypsin would produce unique sets of diagnostic peptides that could be used to identify each protein. Soluble testis proteins were digested with trypsin and analyzed with an LTQ ion trap mass spectrometer. The instrument was configured to acquire tandem mass spectra for 9 precursor ion *m*/*z* values, each for an anticipated tryptic peptide of lysin. Of these 9 *m*/*z* values, 4 were predicted for doubly charged peptides unique to copy 1 and 5 were predicted to be doubly charged peptides unique to copy 2. From the resulting tandem mass spectra, we found 4 out of 4 diagnostic peptides for copy 1 and 2 out of 5 diagnostic peptides for copy 2 (fig. 2*B*). In each case, the peptides were unique when compared with all of the *Haliotis* proteins currently in GenBank, confirming them as copy-specific lysin peptides.

Although it is difficult to determine the relative abundance of each copy of the lysin protein, several lines of evidence suggest that copy 1 is more abundant. First, copy 1 was the only copy of *H. tuberculata* lysin identified from cDNA in a prior study (Lee et al. 1995). Second, copy 1 transcripts were cloned more frequently in our RT–PCR experiment. Finally, copy 1 peptides were more abundant in our MS analysis: the mass spectrometer identified 111 spectra (representing 4 unique peptides) for copy 1, compared with only 20 spectra (representing 2 peptides) for copy 2. Because comparing spectral counts is an approximation for protein quantities in MS experiments (Liu et al. 2004), these data, taken together with the transcript data, suggest that copy 1 is more abundant. This difference is unlikely due to a loss of function of copy 2 because its structural divergence is consistent with continued function (see below).

The above expression analysis was conducted by necessity in *H. t. tuberculata*. We confirmed the presence of

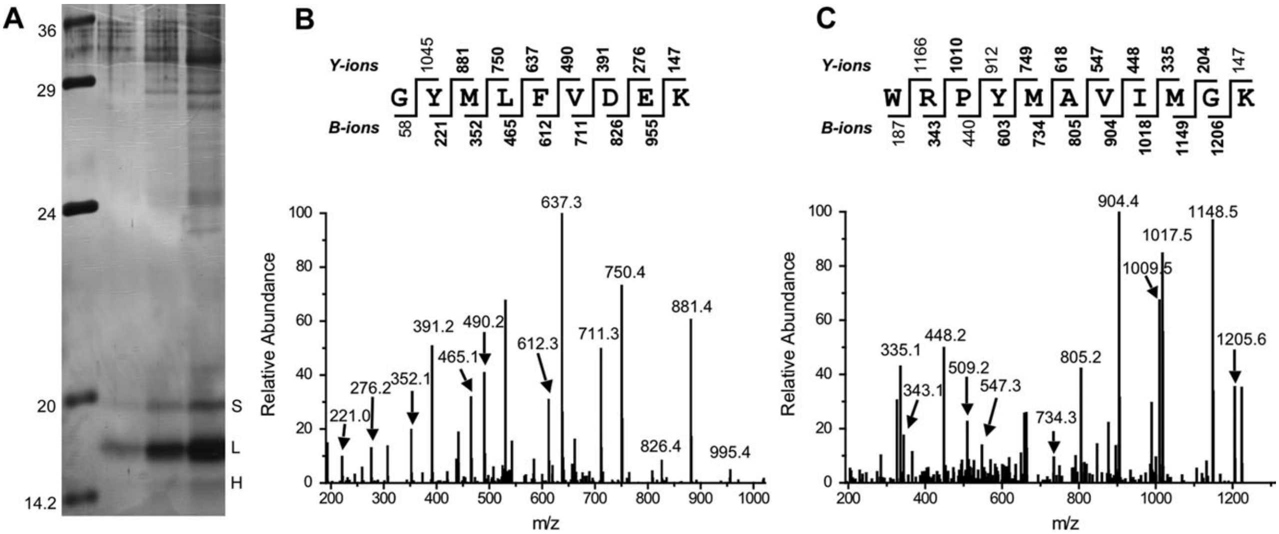

FIG. 2.—Analysis of *Haliotis tuberculata tuberculata* testis proteins. (*A*) Silver-stained SDS-PAGE gel of testis proteins at 3 concentrations. Bands are indicated as S (sp18), L (lysin), and H (histone H4). L and H bands were identified by mass mapping; S was identified by its mobility relative to lysin, its abundance, and 2 internally sequenced peptides that show partial identity to sp18 sequences in other species. Molecular weight markers (in kDa) are shown at the left. (*B* and *C*) MS showing the peptide fragmentation patterns used to identify 2 peptides specific to lysin copy 2. The peptide sequence is shown above, with masses (in Daltons) for each possible peptide fragment indicated between each residue. Masses that are highlighted in bold numbers on the peptide sequence were identified by the instrument and are indicated by number on the spectra. Four peptides unique to copy 1 were identified in the same manner (data not shown).

both copies in *H. t. coccinea* by amplifying genomic lysin sequence from total DNA. Each copy of lysin was present in all 11 individuals sampled, suggesting that both copies of lysin are fixed in both subspecies. We then determined the full gene sequence for copy 1 in 9 individuals (see the polymorphism survey results below).

Patterns of Divergence of Duplicated Lysins

A coding sequence phylogeny of lysin shows with 100% bootstrap support that the duplication event occurred before divergence of the *H. tuberculata* subspecies (fig. 1*B*). To study potential functional divergence, we compared the lineages of each copy, highlighted in blue or red in figure 1. We first addressed whether either copy could be a nonfunctional pseudogene. During evolution, amino acid residues in the hydrophobic core of a protein change slowly to maintain proper folding (Li 1997). Hence, a nonrandom distribution of substitutions between core and surface sites could reflect negative selection to maintain function. We used maximum likelihood to infer all changes leading to each lysin copy since duplication, and these changes were mapped onto a threaded structural model of *H. tuberculata* lysin. Copy 1 showed a deficiency of core residue substitutions, with 0.037 core substitutions per core site versus 0.152 substitutions per surface site. To test whether this difference in substitution rate was significant, we compared the mean solvent exposure of substituted residues (59%) to the mean solvent exposures of 100,000 sets of randomly selected sites. The average mean solvent exposure for these sets was 39.5%, and it was very rare for a randomly generated set to meet or exceed the observed 59% ($P = 0.00315$). Therefore, post-duplication substitutions on copy 1 show a bias toward solvent-exposed sites, which is consistent with selection for maintaining the inner core of a functional protein. Copy 2 showed a similar deficiency of core substitution, 0 versus 0.119, and the mean solvent exposure (68.6%) was also significantly higher than expected by chance, again judged by 100,000 permutations ($P = 0.00125$). The observed patterns of recent divergence show that different levels of selective constraint have acted on the core and surface residues in both lysin copies, suggesting that their protein folds are maintained by purifying selection.

We then asked whether the 2 copies had diverged at similar regions of the folded protein. Using a structural model of lysin, we found 18 changes along the lineage leading to copy 1 whose neighboring residues changed in the lineage leading to copy 2. Two residues were designated neighbors if their Van der Waals' surface areas were found within 1 Å of each other. We judged the observed number of neighbors to be nonrandom by comparison to sets of randomly chosen surface residues ($P = 0.04003$; 100,000 permutations). Thus, copies 1 and 2 are changing at similar 3-dimensional sites on the protein surface.

Finally, we asked whether these duplicate copies continued to diverge at sites previously found under positive selection in distantly related abalone species (Yang and Swanson 2002). Copy 1 showed an excess of substitutions at positively selected sites but was not significant (Fisher's exact $P = 0.27$; supplementary table S1, Supplementary Material online). Copy 2 showed a statistically significant excess of substitutions at positively selected sites, even when analysis was restricted to surface exposed sites (Fisher's exact $P = 0.016$; supplementary table S1, Supplementary Material online). Taken together, these evolutionary lines of evidence suggest that both *H. tuberculata* lysin proteins continue to diverge at structurally similar amino acid sites driven by forces similar to those observed in other species.

Table 1
Tests of Selection Using Likelihood Models

Hypothesis	Null Model	Alternative Model	LRT
Branch model variation among branches	$\omega = 0.76$ lnL = −2442.63 (np = 15)	independent ω values in figure 1*B* lnL = −2424.55 (np = 27)	$\chi^2 = 36.16$, 12 df, $P = 0.0003$
Branch model postduplication $\omega > 1$	$\omega = 1$ lnL = −2427.83 (np = 23)	$\omega = 2.12$ lnL = −2426.36 (np = 24)	$\chi^2 = 2.94$, 1 df, $P = 0.086$
Branch-site model copy 1 lineage	$\omega_2 = 1$ lnL = −2387.32 (np = 17)	$\omega_2 = \infty$ for 9.2% of sites lnL = −2380.63 (np = 18)	$\chi^2 = 13.39$, 1 df, $P = 0.00025$
Branch-site model copy 2 lineage	$\omega_2 = 1$ lnL = −2388.22 (np = 17)	$\omega_2 = 70$ for 4.8% of sites lnL = −2384.74 (np = 18)	$\chi^2 = 6.97$, 1 df, $P = 0.0083$

NOTE.—ω = estimated d_N/d_S ratio; np = number of parameters; lnL = log likelihood

To compare the selective pressure acting along each lineage, we used maximum likelihood to estimate the d_N/d_S ratio for each branch of the lysin phylogeny. The d_N/d_S ratio is a measure of the rate of amino acid change normalized by the rate of synonymous change. A ratio greater than one indicates an excess of amino acid substitution (positive selection), whereas a ratio less than one indicates conservation of amino acid sequence (purifying selection). A branch model allowing each branch to have its own d_N/d_S ratio fit the lysin data significantly better than a 1-ratio model ($P = 0.0003$). As expected from previous studies of lysin, elevated rates of evolution were detected on branches leading to multiple species (fig. 1*B*; Yang et al. 2000). Importantly, we also found evidence for positive selection acting in the *H. tuberculata* lineage both before and after gene duplication (fig. 1*B*; table 1). Considering the divergence of the paralogs, all postduplication branches experienced a total of 31 nonsynonymous and 7 synonymous substitutions, yielding a d_N/d_S estimate of 2.12 across all codons. This ratio is marginally different from the neutral expectation of $d_N/d_S = 1$ ($P = 0.086$, table 1). However, this test for positive selection over the entire protein is conservative because some codons are likely to be conserved for structural purposes, as demonstrated above. An improvement over estimating one ratio for all codons is to accommodate variation between them. Branch-site models allow classes of codons with different d_N/d_S ratios while testing along designated branches (Zhang et al. 2005). These models allow sets of codons to experience purifying selection, whereas others can evolve neutrally or under positive selection. Branch-site analysis showed significant signs of positive selection acting on a subset of codons along each postduplication lineage ($P = 0.00025$ for copy 1 and $P = 0.0083$ for copy 2; table 1). These patterns of substitution are consistent with positive selection and suggest that the lysin paralogs have evolved rapidly and adaptively.

Polymorphism in Allopatric *H. t. coccinea*

We surveyed polymorphism at lysin and 3 neutral loci in a population of the subspecies *H. t. coccinea* collected in the Azores archipelago. This mid-Atlantic subspecies has diverged substantially from the greater European subspecies, *H. t. tuberculata*, as shown by allele frequency differences between the 2 populations (F_{ST} in table 2). This population is of interest for studying selection on gamete

Table 2
Polymorphism and Selection in *Haliotis tuberculata coccinea*, an Island Subspecies

	Lysin Copy 1	*rpL5*	*hemocyanin 1*	*cytochrome oxidase I*
Location	Nucleus	Nucleus	Nucleus	Mitochondrion
Aligned nt	5,597	491	357	435
Coding nt	459	0	0	435
Chromosomes	18	20	22	11
S	20	8	14	3
Singletons	16	4	4	2
π	0.0006	0.00371	0.01373	0.00159
$\pi a/\pi s$	0	n/a	n/a	0
F_{ST} between subspecies	0.96 ($P = 0.00002$)	0.52 ($P = 0.015$)	0.19 ($P = 0.042$)	0.91 ($P = 0.0028$)
Jukes–Cantor *D*	0.25	0.24	n/a	0.19
Tajima's *D*	−1.63 ($P = 0.012$)	−0.65 (NS)	0.42 (NS)	−1.11 (NS)
Fu and Li's *D*	−2.29 (P = 0.008)	−1.01 (NS)	−0.47 (NS)[a]	0.97 (NS)
Fu and Li's *F*	−2.55 ($P = 0.005$)	−1.15 (NS)	−0.24 (NS)[a]	0.64 (NS)

NOTE.—*rpL5* = intron of a conserved gene encoding ribosomal protein L5; nt = nucleotides; S = segregating sites; π = nucleotide diversity; n/a = not applicable; Jukes–Cantor *D* = distance to outgroup species *Haliotis pustulata*; NS = not significant ($P > 0.05$).

[a] Because an outgroup species was not sequenced for *hemocyanin 1*, the values listed for the Fu and Li's tests are D^* and F^*.

Table 3
HKA Test Between Lysin and a Neutral Locus

HKA Test	Lysin Copy 1	*rpL5*
Intraspecific segregating sites	20	8
Intraspecific aligned sites	5597	491
Interspecific differences	236.5	86
Interspecific aligned sites	1096	413
HKA Test	$\chi^2 = 10.472$, $P = 0.0012$	

recognition: by characterizing lysin in a species outside of the North Pacific clade, we can ask whether patterns of evolution are different in this distinct phylogenetic clade (Coleman and Vacquier 2002). Also, lysin polymorphism has been characterized only in sympatric species. By testing for selection in this geographically isolated, allopatric population, we asked whether a selective force other than reinforcement is acting in the population. In this survey, we focused exclusively on copy 1.

The pattern of polymorphism in *H. t. coccinea* suggests a recent selective sweep at lysin. No amino acid–changing polymorphisms and only 2 synonymous substitutions were found among 18 sequenced chromosomes (table 2). To ask whether this coding monomorphism was caused by recent selection, we also examined polymorphisms in lysin's introns. The entire gene region yielded 5,597 contiguous aligned base pairs and contained 20 polymorphic sites within the population sample. The *H. t. coccinea* population shows reduced levels of polymorphism at lysin ($\pi = 0.06\%$) compared with neutral loci, *rpL5*, and *hemocyanin 1* ($\pi = 0.37\%$ and 1.37%, respectively). This reduction cannot be attributed to a different mutation rate (table 3, Hudson–Kreitman–Aguade [HKA] test $P = 0.0012$). The HKA test (Hudson et al. 1987) shows nonconcordance of polymorphism and divergence between lysin and an intron of *rpL5*. Assuming *rpL5* has undergone neutral evolution, this departure from neutrality could be due to either a deficiency of polymorphism at lysin or an excess of divergence in lysin's introns. Because estimates of interspecific divergence are similar for introns of lysin and *rpL5* (table 2, Jukes–Cantor *D*) and are consistent with synonymous divergence at *histone H3* (estimated $d_S = 0.17$), polymorphism is likely reduced in the lysin gene region.

Recent positive selection probably swept an advantageous allele to high frequency in this population, removing polymorphism from the gene region. Under this scenario, recovery from a selective sweep is expected to produce an excess of low-frequency polymorphisms, which can be detected by the Tajima's *D* statistic (Tajima 1989). As shown in table 2, Tajima's *D* is significantly negative for lysin, indicating just such an excess of low-frequency polymorphisms. Similarly, *D* and *F* statistics of Fu and Li (1993) show a significant excess of rare polymorphisms within the sample; 16 of 20 polymorphisms are singletons. Observed values of these statistics differ significantly from the results of neutral coalescent simulations (table 2). The simulations used parameter estimates from the data, including the recombination rate. Even when simulations were performed under no recombination, a conservative assumption (Przeworski 2002), all tests remained significant at the $\alpha = 0.05$ level. Because demographic effects can also produce departures from neutrality across the entire genome, we tested for this effect at neutral loci, *rpL5*, *hemocyanin 1*, and *cytochrome oxidase I*. The same statistics at these loci were consistent with neutral expectations, showing the result at lysin to be locus specific (table 2). Overall, polymorphism-based evidence is consistent with recent positive selection sweeping the copy 1 lysin coding allele to high frequency in the allopatric Azores population of *H. t. coccinea*.

Discussion

We have presented the first known case of lysin gene duplication in the geographically isolated abalone, *H. tuberculata*. This duplication occurred before *H. t. tuberculata* and *H. t. coccinea* diverged and appears to be fixed in both populations. Both paralogs are expressed as proteins in testis tissue, and each copy has evolved under positive selection. Additionally, copy 1 shows signs of a recent selective sweep in an isolated population of *H. t. coccinea*.

Because the lysin duplicates have diverged considerably, one copy might have become a pseudogene or acquired a function other than vitelline envelope dissolution. Our data do not support either possibility. Both copies probably remain functional because each shows negative selection at core structural sites. We predict that each is still able to dissolve vitelline envelopes by comparing them to the functionally characterized lysins of *H. corrugata* and *H. rufescens*. These lysins show a slightly higher level of divergence but are still able to dissolve vitelline envelopes of the opposite species, albeit with lower efficiency (Swanson and Vacquier 1997). Furthermore, the patterns of evolution of copies 1 and 2 reflect those observed for lysins of other abalone species. Both copies evolve under positive selection, and many of their amino acid substitutions are found at sites shown to be under selection in other abalone lineages. Finally, both lysins were found as proteins in testis tissue, suggesting a continued role in fertilization. Taken together, these data suggest that both *H. tuberculata* lysins continue to dissolve vitelline envelopes and are perhaps undergoing subfunctionalization.

Given the coevolution of lysin and VERL, 2 hypotheses could explain the fixation of and selection on the lysin paralogs. The first hypothesis is that lysin paralogs have specialized on divergent regions of VERL, its binding partner on the egg coated. VERL is a large, multivalent protein containing 22 tandem repeat units in *H. rufescens* (Galindo et al. 2002). Binding stoichiometry indicates that lysin binds to each repeat (Swanson and Vacquier 1997), yet the repeats evolve under variable selective pressures. The 2 N-terminal repeats evolve independently from the remaining repeats and from each other, and their divergence is driven by positive selection (Galindo et al. 2003; Clark NL, Springer S, Swanson WJ, in preparation). The remaining 20 repeats evolve neutrally and are homogenized by concerted evolution. If the structure of VERL in *H. tuberculata* were similarly complex, lysin molecules would encounter a heterogeneous set of motifs when dissolving the vitelline envelope. Diversification and subfunctionalization of the duplicate lysins could allow more efficient interactions with the egg coat, thus allowing competing sperm to more quickly penetrate the egg. Future studies quantifying each lysin's binding affinity to different regions of VERL can directly test this hypothesis.

A second hypothesis is that lysin paralogs have subfunctionalized to bind divergent alleles of VERL. VERL shows substantial polymorphism in *H. corrugata*, which may result from sexual conflict (Clark NL, Springer S, Swanson WJ, in preparation). In the abalone fertilization system, sexual conflict is hypothesized to occur over the rate of sperm entry into the egg (Swanson and Vacquier 2002). Sperm competition continually selects for male sperm proteins capable of an increased rate of egg penetration. However, if an increased rate results in high levels of polyspermy, female gamete recognition proteins are predicted to evolve to reduce the efficiency of sperm–egg interactions. This conflict over fertilization rate can result in a coevolutionary chase, which is observed between lysin and VERL (Clark NL, Springer S, Swanson WJ, in preparation). Mathematical models of sexual conflict have shown that frequency-dependent selection can act on the female locus, resulting in 2 divergent alleles (reviewed in Gavrilets and Hayashi 2005; Hayashi et al. 2007). In this regime, the interacting male locus is unable to optimize to either female allele, reducing male fitness. In a species in which the male locus underwent gene duplication, each male paralog could optimize to a female allele. Thus, the fixation of gene duplicates could be a novel result of sexual conflict. Under this hypothesis, the 2 *H. tuberculata* lysin loci are driven by sexual conflict to adapt to distinct VERL alleles.

Lysin gene duplication is also of interest because of the ancient duplication that created lysin and sp18 (Vacquier et al. 1997; Kresge et al. 2001). This subfunctionalization is one of the best-documented cases of gene duplication of a fertilization protein, but because it predates the formation of the *Haliotis* genus, it is impossible to study the patterns of divergence that occurred immediately after the event. The duplication of lysin in *H. tuberculata* presents an opportunity to follow the functional consequences of gene duplicates in the early stages of divergence.

We found signs of recent positive selection at lysin copy 1 in an island population of *H. t. coccinea*. Because population-based tests of selection reflect recent evolutionary history, their results have interesting implications for hypotheses explaining the rapid evolution of gamete recognition proteins. The reinforcement hypothesis requires hybridization between populations for which a hybrid is less fit than either parental species. Because the sampled population of *H. t. coccinea* is found in apparent allopatry in the Azores archipelago, a different selective pressure was likely responsible for the recent selective sweep and coding monomorphism. Reinforcement remains a formal possibility because it is difficult to rule out complex demographic explanations, such as historical range changes. However, *H. t. coccinea* is the only reported abalone in the Azores archipelago (Geiger 2000), which is ~900 km from the nearest abalone population off Madeira Island and over 1,300 km from the nearest continental population. A detailed phylogeographic study is required to determine the degree of allopatry for this island population and to more completely assess the likelihood of scenarios allowing reinforcement.

In contrast to reinforcement, several processes of sexual selection can explain lysin's adaptive evolution among both sympatric and allopatric species. First, sperm competition is well documented in other external fertilizers (Levitan 2004; Levitan and Ferrell 2006) and could exert continual selection on male gamete proteins. Second, as described above, sexual conflict over the rate of polyspermy is predicted to drive coevolution of male and female fertilization proteins (Gavrilets 2000). Third, female eggs may select for specific male alleles in a classical process of sexual selection (Palumbi 1999). Finally, egg vitelline envelopes may be subject to pathogen attack while awaiting fertilization, which would create a selective pressure on VERL, and subsequently on lysin, to evolve (Swanson and Vacquier 2002). While one or more of these forces may be at play, none of them depend on the presence or absence of other species in the same area. Thus, they are able to explain the rapid evolution of lysin in all species of abalone. If 2 groups of abalone undertook different evolutionary trajectories in response to any of these pressures, speciation could occur solely due to a selective pressure intrinsic to the mating system and regardless of any spatial relationships with other populations (Gavrilets and Waxman 2002; Hayashi et al. 2007).

We have discovered the first case of lysin gene duplication and shown that divergent selection continues to act on both paralogs, even though the species harboring the duplication is allopatric. Both copies are expressed as proteins and are likely undergoing subfunctionalization to different regions or alleles of the lysin receptor. This finding of recent lysin duplication, particularly given the protein's well-documented interaction with VERL, provides a unique opportunity to investigate the functional divergence of gene duplicates. Our results suggest that gene duplication may increase reproductive fitness and implicate a process of sexual selection or coevolution as the driving force behind lysin's rapid evolution.

Supplementary Material

Supplementary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). Sequence data have been deposited in GenBank under the accession numbers: EF660344–EF660434, EF673712, and EF681891.

Acknowledgments

We thank Mike Hellberg, Jeff Kidd, and members of the Swanson lab for valuable input. Vic Vacquier generously donated several abalone samples. This work was funded by National Institute of Health (NIH) grants HD42563 (to W.J.S.) and P41 RR011823 (to M.J.M.) and National Science Foundation grant DEB-0716761 (to W.J.S.). N.L.C. and G.D.F. were funded by NIH departmental training grants (T32 GM07735 and T32 HG00035).

Literature Cited

Aagaard JE, Yi X, MacCoss MJ, Swanson WJ. 2006. Rapidly evolving zona pellucida proteins are a major component of the vitelline envelope of abalone eggs. Proc Natl Acad Sci USA. 103:17302–17307.

Arai K, Wilkins NP. 1986. Chromosomes of *Haliotis tuberculata* L. Aquaculture. 58:305–308.

Armbrust EV, Galindo HM. 2001. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus *Thalassiosira*. Appl Environ Microbiol. 67:3501–3513.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31:3497–3500.

Coleman AW, Vacquier VD. 2002. Exploring the phylogenetic utility of its sequences for animals: a test case for abalone (*Haliotis*). J Mol Evol. 54:246–257.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

DeLano WL. 2002. The PyMOL molecular graphics system. San Carlos (CA): DeLano Scientific.

Eng JK, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 5:976–989.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Felsenstein J. 1989. PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics. 5:164–166.

Franzkiewicz R, Braun W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. J Comput Chem. 19:319–333.

Frohlich O, Po C, Young LG. 2001. Organization of the human gene encoding the epididymis-specific EP2 protein variants and its relationship to defensin genes. Biol Reprod. 64:1072–1079.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133:693–709.

Galindo BE, Moy GW, Swanson WJ, Vacquier VD. 2002. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. Gene. 288:111–117.

Galindo BE, Vacquier VD, Swanson WJ. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci USA. 100:4639–4643.

Gavrilets S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature. 403:886–889.

Gavrilets S, Hayashi TI. 2005. Speciation and sexual conflict. Evol Ecol. 19:167–198.

Gavrilets S, Waxman D. 2002. Sympatric speciation by sexual conflict. Proc Nat Acad Sci USA. 99:10533–10538.

Geiger DL. 2000. Distribution and biogeography of Haliotidae (Gastropoda: Vetigastropoda) worldwide. Boll Malacol. 35:57–120.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Hayashi TI, Vose M, Gavrilets S. 2007. Genetic differentiation by sexual conflict. Evolution. 61:516–529.

Holloway AK, Begun DJ. 2004. Molecular evolution and population genetics of duplicated accessory gland protein genes in *Drosophila*. Mol Biol Evol. 21:1625–1628.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

Jensen ON, Podtelejnikov AV, Mann M. 1997. Identification of the components of simple protein mixtures by high accuracy peptide mass mapping and database searching. Anal Chem. 69:4741–4750.

Kresge N, Vacquier VD, Stout CD. 2001. The crystal structure of a fusagenic sperm protein reveals extreme surface properties. Biochemistry. 40:5407–5413.

Lee YH, Ota T, Vacquier VD. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol. 12:231–238.

Levitan DR. 2004. Density-dependent sexual selection in external fertilizers: variances in male and female fertilization success along the continuum from sperm limitation to sexual conflict in the sea urchin *Strongylocentrotus franciscanus*. Am Nat. 164:298–309.

Levitan DR, Ferrell DL. 2006. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science. 312:267–269.

Lewis CA, Talbot CF, Vacquier VD. 1982. A protein from abalone sperm dissolves the egg vitelline layer by a non-enzymatic mechanism. Dev Biol. 92:227–239.

Li W-H. 1997. Molecular evolution. Sunderland (MA): Sinauer Associates.

Liu HB, Sadygov RG, Yates JR. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 76:4193–4201.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155.

Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. Trends Genet. 20:544–549.

Lyon JD, Vacquier VD. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev Biol. 214: 151–159.

Marshall JL, Arnold ML, Howard DJ. 2002. Reinforcement: the road not taken. Trends Ecol Evol. 17:558–563.

Metz EC, Robles-Sikisaka R, Vacquier VD. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci USA. 95:10676–10681.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.

Noor MA. 1999. Reinforcement and other consequences of sympatry. Heredity. 83:503–508.

Palumbi SR. 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. Proc Natl Acad Sci USA. 96:12632–12637.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet. 3:827–837.

Przeworski M. 2002. The signature of positive selection at randomly chosen loci. Genetics. 160:1179–1189.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin: a software for population genetics data analysis. Version 3. Geneva (Switzerland): Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31:3381–3385.

Stephens M, Sloan JS, Robertson PD, Scheet P, Nickerson DA. 2006. Automating sequence-based detection and genotyping of SNPs from diploid samples. Nat Genet. 38:375–381.

Swanson WJ, Vacquier VD. 1995a. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc Natl Acad Sci USA. 92:4957–4961.

Swanson WJ, Vacquier VD. 1995b. Liposome fusion induced by a M(r) 18000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoa. Biochemistry. 34:14202–14208.

Swanson WJ, Vacquier VD. 1997. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci USA. 94:6724–6729.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3:137–144.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Torgerson DG, Singh RS. 2004. Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha4 proteasome subunits in *Drosophila*. Genetics. 168:1421–1432.

Vacquier VD, Swanson WJ, Lee YH. 1997. Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J Mol Evol. 44:S15–S22.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.

Yang Z, Swanson WJ. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol. 19:49–57.

Yang Z, Swanson WJ, Vacquier VD. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol. 17:1446–1455.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22:2472–2479.

Ziheng Yang, Associate Editor

Accepted July 3, 2007