Positive Selection on an Acrosomal Sperm Protein, M7 Lysin, in Three Species of the Mussel Genus *Mytilus*

*Cynthia Riginos* and *John H. McDonald*

*Department of Biology, Duke University; and †Department of Biological Sciences, University of Delaware*

Marine invertebrate sperm proteins are particularly interesting because they are characterized by positive selection and are likely to be involved in prezygotic isolation and, thus, speciation. Here, we present the first survey of interspecific and intraspecific variation of a bivalve sperm protein among a group of species that regularly hybridize in nature. M7 lysin is found in sperm acrosomes of mussels and dissolves the egg vitelline coat, permitting fertilization. We sequenced multiple protein-coding regions of M7 lysin from allopatric populations of mussel in the *Mytilus edulis* species group (*M. edulis, M. galloprovincialis*, and *M. trossulus*). A significant McDonald-Kreitman test showed an excess of fixed amino acid replacing substitutions between species, consistent with positive selection. In addition, Kolmogorov-Smirnov tests showed significant heterogeneity in polymorphism to divergence ratios for both synonymous variation and combined synonymous and nonsynonymous variation within *M. galloprovincialis*. These results indicate that there has been adaptive evolution at M7 lysin and, furthermore, show that positive selection on sperm proteins can occur even when postzygotic reproductive isolation is incomplete.

Introduction

Genes involved in sexual reproduction frequently show a strong signal of adaptive evolution at the molecular level (reviewed by Howard 1999; Swanson and Vacquier 2002), and some of the most extreme examples are proteins expressed on the sperm surfaces of marine invertebrates (Lee, Ota, and Vacquier 1995; Swanson and Vacquier 1995; Metz and Palumbi 1996; Vacquier, Swanson, and Lee 1997; Biermann 1998; Metz, Robles-Sikisaka, and Vacquier 1998; Hellberg and Vacquier 1999; Hellberg, Moy, and Vacquier 2000). Despite the repeated action of positive selection on sperm protein genes, the underlying cause of selection is not understood (reviewed by Vacquier, Swanson, and Lee 1997; Howard 1999; Swanson and Vacquier 2002 and references therein). However, considering the evolutionary patterns of such genes across a wide array of taxa may help elucidate causes of selection.

Although there are several examples of positive selection among marine invertebrate sperm proteins, the exact pattern of selection is somewhat idiosyncratic. In general, marine invertebrates reproduce by spawning their gametes into the water column. Sperm and eggs are drawn together by chemotaxis, the sperm acrosomal vesicle opens, the egg and sperm bind, the sperm lyses the egg vitelline envelope, and the gametes fuse (Vacquier 1998). Two sperm genes involved in the fertilization sequence have been extensively studied. Bindin is the gene causing sea urchin sperm to adhere to eggs. Positive selection in bindin (evidenced by a ratio of nonsynonymous to synonymous amino acid substitutions, dN/dS, between species exceeding 1) is restricted to a particular region of the gene, and considerable variation is observed both among and within species (Metz and Palumbi 1996; Biermann 1998). In contrast, positive selection in gastropod lysin (which lyses the egg vitelline envelope) is not restricted to one gene region. Many species pairs of abalone and teguline snails show substantially more nonsynonymous (dN) than synonymous (dS) substitutions across the entire gene, and within species, variation is nearly nonexistent (Lee, Ota, and Vacquier 1995; Metz, Robles-Sikisaka, and Vacquier 1998; Hellberg and Vacquier 1999; Yang, Swanson, and Vacquier 2000). Other gastropod sperm genes with unknown functions show similar patterns to lysin (e.g., many amino acid changes between species but no variation within species [Swanson and Vacquier 1995; Vacquier, Swanson, and Lee 1997; Hellberg, Moy, and Vacquier 2000]). The high ratios of dN/dS (>1) and the lack of intraspecific polymorphisms among gastropod sperm genes point to the action of selective sweeps that have rapidly fixed replacement substitutions within gastropod species and eliminated synonymous polymorphisms.

Studies to date have found positive selection on marine invertebrate sperm proteins among closely related species with overlapping geographic ranges (sympathy). Only sympatric groups of sea urchins show signatures of positive selection on bindin (Metz and Palumbi 1996; Biermann 1998). In contrast, no evidence for positive selection has been found within genera characterized by allopatric species (Metz, Gómez-Gutiérrez, and Vacquier 1998; Zigler and Lessios 2003). In addition, the groups of gastropod mollusks (abalone and teguline snails) for which positive selection has been reported includes many sympatric species pairs.

Sympathy with closely related species also seems to be associated with increased sperm specificity. In sea urchins, allopatric species crosses require substantially less sperm for egg fertilization than crosses involving sympatric species pairs (Metz et al. 1994; Metz, Gómez-Gutiérrez, and Vacquier 1998). Similarly, interspecific fertilization assays of abalone (Leighton and Lewis 1982; Swanson and Vacquier 1995) and snails (Hellberg and Vacquier 1999) show that sperm are much more efficient in fertilizing conspecific eggs than heterospecific eggs. The only direct evidence for a sperm protein’s association with conspecific egg recognition comes from experimental crosses where sea urchin sperm were more likely to fertilize eggs with the same homozygous bindin genotype.
(Palumbi 1999). Both sea urchins and abalone can be experimentally crossed to produce some viable offspring (Leighton and Lewis 1982; Palumbi and Metz 1991). However, the occurrence of hybrids in nature is rare, with hybrid frequencies estimated as 0.1% of *Echinometra* sea urchin populations (Palumbi and Metz 1991) and 0.37% of *Haliotis* abalone populations (Owen, McLean, and Meyer 1971).

Here, we focus on a group of closely related taxa that extensively hybridize in nature. Mussels in the *Mytilus edulis* (Mollusca: Bivalvia) species group (*M. edulis*, *M. galloprovincialis*, and *M. trossulus*) are common in the intertidal and shallow subtidal areas of the temperate Northern and Southern hemispheres (McDonald, Seed, and Koehn 1991; Rawson and Hilbish 1995). In the Northern Hemisphere, *M. edulis* is present in the North Atlantic on both the North American and the European coastlines. *M. trossulus* is present in the North Pacific, Atlantic Canada, and the Baltic Sea. *M. galloprovincialis* is native to the Mediterranean and Black Seas and western Europe, and it has been recently introduced to California and Asia (McDonald and Koehn 1988; Heath, Rawson, and Hilbish 1996; Suchanek et al. 1997). Some hybrid individuals have been found in every area where these species are sympatric: *M. edulis* and *M. galloprovincialis* in western Europe (Skibinski and Beardmore 1979; Gardner 1996; Hilbish et al. 2002), *M. edulis* and *M. trossulus* in northern Europe (Vänionlää and Hvilson 1991; Riginos, Sukhdeo, and Cunningham 2002) and in Atlantic Canada (Saavedra et al. 1996; Comesaño et al. 1999), and *M. trossulus* and *M. galloprovincialis* in California (Suchanek et al. 1997; Rawson, Agrawal, and Hilbish 1999) and Asia (Skurikhina et al. 2001, but see Suchanek et al. 1997). In fact, hybridization is sufficiently frequent that hybrids comprise over 30% of the mussels in some regions of the major hybrid zones (Gardner 1996; Saavedra et al. 1996; Riginos, Sukhdeo, and Cunningham 2002).

Despite extensive hybridization, these three species remain genetically distinct throughout much of their species range; therefore, some prezygotic or postzygotic processes must contribute to reproductive isolation. The mechanisms of fertilization in mussels have not been well studied, although recent laboratory crosses show that there is some degree of reduced fertilization success between *M. edulis* and *M. galloprovincialis* (Bierne et al. 2002), suggesting that gamete recognition might play a role in isolating these species. In experimental crosses, F1 hybrid viability is also reduced for *M. edulis × M. galloprovincialis* (Bierne et al. 2002) and *M. edulis × M. trossulus* (Zouros et al. 1994a) crosses, demonstrating some level of postzygotic isolation among the three species. However, F2 and backcross-type hybrid individuals are found in mussel hybrid zones, indicating that at least some F1 hybrids are fertile (Gardner 1996; Rawson, Secor, and Hilbish; Saavedra et al. 1996; Riginos, Sukhdeo, and Cunningham 2002).

Takagi et al. (1994) found three proteins in high concentration on the acrosomal surface of *Mytilus* sperm. One of these three proteins showed strong lysogenic activity, nonenzymatically dissolving the egg vitelline coat. This 20 kDa protein, which Takagi and colleagues named M7 lysin, also reinitiates the suspended meiosis by releasing the egg’s first polar body. M7 lysin does not exhibit significant sequence similarity to lysins from abalone or snails, rather is classified by its function. Given the rapid rate of protein evolution observed for lysins in these gastropods, the lack of obvious similarity is not surprising and does not necessarily suggest independent origins of the different lysins.

Understanding how selection drives the evolution of sperm proteins and understanding the role of sperm proteins in speciation will be aided by studying such genes in diverse taxa, particularly among species with incomplete reproductive isolation. Consequently, we have sequenced multiple alleles of the mature protein coding region of M7 lysin in allopatric populations of *M. edulis*, *M. galloprovincialis*, and *M. trossulus*. By examining the patterns of polymorphism and divergence, we demonstrate that *Mytilus* M7 lysin has evolved in a manner consistent with positive selection, as with other marine invertebrate sperm proteins. This study provides a new phylogenetic perspective by demonstrating that positive selection of gamete recognition genes also occurs in bivalve mussels. This is also the first evidence for positive selection on a marine invertebrate gamete recognition gene in a group characterized by extensive hybridization among species.

**Methods**

**RNA Isolation and Sequencing**

*Mytilus edulis* was collected from Woods Hole, Mass., and *M. galloprovincialis* was collected from Samos, Greece. These regions are known to have pure (nonhybrid) populations of each species. *M. trossulus* was collected from Penn Cove, Wash. Because imported *M. galloprovincialis* are cultivated in Washington alongside the native *M. trossulus*, the species identity of the *M. trossulus* individuals was confirmed using sequences of the mitochondrial COIII gene (Zouros et al. 1994b).

Total RNA was extracted from mantle tissue using Tri Reagent (Sigma). Fresh tissue was homogenized in 0.5 ml Tri Reagent and vortexed briefly after a 0.2 ml addition of chloroform. The tissue-Tri Reagent-chloroform mixture was centrifuged at 10,000 RPM for 15 min at 4°C. The top aqueous phase was pipetted into a fresh tube, and RNA was precipitated with 0.5 ml isopropanol. The RNA pellet was washed with 75% ethanol and dried at room temperature. The RNA pellet was resuspended in RNase-free water.

Takagi et al. (1994) sequenced M7 lysin from one individual of *M. galloprovincialis*. To identify a part of the 3’ untranslated region that was conserved among all three species and would thus be a suitable primer site, 3’ RACE was used to amplify and sequence the 3’ region from several individuals of *M. edulis* and *M. trossulus*. Total mRNA was reverse transcribed with SUPERSCRIPT RNase H Reverse Transcriptase (Life Technologies) using an oligo-dT adaptor (5’-AAG GAT CCG TCG ACA TCG ATA AT A GA C G A T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T T
and a primer matching part of the adaptor sequence (5’-AAG GAT CCG TCG ACA TC-3’). A second, nested reaction was performed using a small amount of the first PCR product and the primers MEL-F6 (5’-AAC AAG GAT CAT GCA AAC -3’) and a second primer matching the adaptor sequence (5’-GAC ATC GAT AAT ACG AC-3’), which amplified fragments from a number of M. edulis and M. trossulus individuals. These sequences of the 3’ region, along with the Takagi et al. sequence of the 5’ region, were used to design primers targeting the untranslated regions (UTRs) that would amplify the entire coding region of M7 lysin: MEL-F7, 5’-TGG AAG CCA AAT CTT TAC AT-3’, and MEL-R5, 5’-GTC TTT AGT TCT TTA TTT TAT ATT GAC-3’. PCR was performed with Promega Taq DNA polymerase (Madison, Wis.) for 35 cycles of 20 s at 94°C, 30 s at 54°C, and 80 s at 72°C and amplified an approximately 770 bp fragment containing the entire 540 bp coding sequence for the mature peptide and 26 bp of the 5’ UTR. The length of the amplified 3’ UTR varied among alleles and was approximately 200 bp long.

PCR products were cloned (TA Cloning: Invitrogen). PCR with M13 primers was performed on individual colonies. These M13 PCR reactions were cleaned with a Qiaquick PCR Purification Kit (Qiagen) and then sequenced with M13 primers using the ABI Prisim BigDye Terminators vers. 2.5 Cycle Sequencing Kit. Sequencing reactions were visualized with an ABI 3700 DNA Analyzer. For each individual mussel, multiple colonies were sequenced in order to discriminate between polymorphism and polymerase error. Thus, two alleles were inferred for each individual, and each allele was represented by a minimum of three sequenced clones. Sequence chromatograms were visualized and edited using Sequencher ver. 4.1 software.

Descriptive Statistics and Sequence Similarity

Genetic variation within each of the three species was described using a number of measures. These measures do not test for selection but are useful for comparing M7 lysin to other genes. DNASp ver. 2.5 (Rozas and Rozas 1999) was used to estimate species heterozygosity \((\theta = 4N_{\mu \mu},\) where \(N\) is the effective population size and \(\mu\) is the neutral mutation rate) based on the number of segregating sites (Watterson 1975) and the average number of pairwise differences (Nei and Li 1979). These two estimators of \(\theta\) capture different elements of the frequency distribution of polymorphisms. Recombination rates were estimated by the methods of Hudson and Kaplan (1985) in the program SITES (Hey and Wakeley 1997).

Although M7 lysin is one of the most abundant proteins in mussel sperm acrosomes, Takagi et al. (1994) found no similarity between M7 and any other previously published sperm protein. Since their original description of M7, however, many more protein sequences have been deposited in public databases. To verify their results, we performed a protein BLAST search against the SwissProt and TrEMBL databases of proteins using the default BLOSUM matrix. In addition, we explicitly tried BLASTing the M7 lysin protein sequence against abalone lysin (Vacquier, Carner, and Stout 1990), teguline lysin (Hellberg and Vacquier 1999), and gastropod TMAP (Hellberg, Moy, and Vacquier 2000), the other described sperm proteins for mollusks. Protein secondary structure was predicted for M. edulis using the PHDsec program (Rost and Sander 1993) implemented by PredictProtein (http://cubic.bioc.columbia.edu/predictprotein/) for comparison to gastropod lysins. The charge and isoelectric static point of M7 lysin was determined by EMBL Isoelectric Point Service (http://www.embl-heidelberg.de/cgi/pi-wrapper.pl).

Tests of Selection

We analyzed patterns of genetic variation within and among species in a number of ways to determine whether positive selection has shaped genetic variation at the mussel M7 lysin gene. The observed frequency spectra of polymorphisms for each species were compared with neutral expectations using Tajima’s \(D\) statistic (Tajima 1989), as implemented in DNASp. Heterogeneity in the polymorphism to divergence ratios across the gene was tested using the runs test, sliding \(G\)-test, and Kolmogorov-Smirnov test of McDonald (1998). Because there was no \(a\) priori estimate of the whole-sequence recombination parameter, \(R\), the program DNA Slider was run for values of \(R = 2, 4, 8, 16, etc.,\) until \(P\), the probability of obtaining the observed heterogeneity by chance, reached a peak and started to decline. For each value of \(R\), 1,000 replicates were run; if the peak value of \(P\) was less than 0.10 for any test, 10,000 replicates were run. This was done for synonymous variation alone and for synonymous and amino acid replacement variation combined. One M. trossulus sequence (T765A) was used as the outgroup for testing heterogeneity within M. edulis and M. galloprovincialis, and one M. edulis sequence (E390B) was used as the outgroup for M. trossulus.

Relative frequencies of synonymous and amino acid replacement substitutions between species were also used to test for selection. McDonald-Kreitman tests (McDonald and Kreitman 1991) compare the number of synonymous and nonsynonymous fixations (between species) with polymorphisms (within species). Standard McDonald-Kreitman tests were conducted on the whole data set. We also tested for an excess of radical amino acid fixations between species in the McDonald-Kreitman test framework using radical and conservative amino acid change definitions (charge, polarity, and polarity with volume) based on Zhang (2000).

In addition, we looked at ratios of synonymous and nonsynonymous divergence along the three species branches, where a ratio of nonsynonymous to synonymous substitutions greater than unity is taken as evidence for positive selection. In MEGA 2.0 (Kumar et al. 2001), the proportions of nonsynonymous substitutions per nonsynonymous site \(\left(\frac{d}{S}\right)\) and synonymous substitutions per synonymous site \(\left(\frac{d}{S}\right)\), based on the modified Nei and Gojobori (1986) method, were calculated across the whole gene and for a 30-codon sliding window across the gene. Where \(\frac{d}{S}\) was greater than \(\frac{d}{S}\), the significance of this difference was tested with Fisher’s exact test (Zhang, ...
Kumar, and Nei 1997). Yang and others have developed a series of nested ML tests that test for differences among \(d_N/d_S\) among codons in a gene (Nielsen and Yang 1998; Yang et al. 2000) and along branches of a tree (Yang and Nielsen 1998; Yang, Swanson, and Vacquier 2000). These methods were implemented with the CODEML program in PAML (Yang 1997), based on a maximum likelihood tree generated in PAUP* (under an HKY with gamma model) (Swofford 1998).

### Results

**Descriptive Statistics and Sequence Similarity**

Most of the sequenced fragments were M7 lysin, based on similarity to the sequence of Takagi et al. (1994); a few M6 fragments from \(M. edulis\) were also sequenced but are not analyzed here (GenBank accession number AY131183). No more than two distinct alleles were found per individual, consistent with M7 being a single copy gene. The sequenced fragments of M7 lysin included less than 26 bp of the 5' UTR, the entire 540-bp coding region, and 173 to 220 bp of the 3' UTR. The 3' UTR had a large amount of indel variation within and between species, making it difficult to align accurately; therefore only the coding region is analyzed here.

It is unclear whether M7 lysin shares a common origin with gastropod lysins, as protein BLAST searches and analyses of protein structure failed to reveal any significant similarities between \textit{Mytilus} M7 lysin and other proteins, including gastropod sperm proteins. Two \(\alpha\)-helices were predicted, at amino acid positions 77 to 86 and 97 to 109, and several short (<10 a.a.) \(\beta\)-sheets were predicted interspersed along the protein. The two predicted \(\alpha\)-helices fall in the central region of the protein, a region containing \(\alpha\)-helices in abalone (based on crystallography: Shaw et al. 1993; Kresge, Vacquier, and Stout 2001). However, many of the other \(\alpha\)-helical regions in abalone lysin (and predicted for teguline lysin [Hellberg and Vacquier 1999]) are not predicted for mussel lysin. M7 lysin is inferred to have a charge of –0.02 and an isoelectric point at 8.5, less than that for gastropods (10.0 to 11.6) (Vacquier, Swanson, and Lee 1997; Hellberg and Vacquier 1998). Although a common origin for gastropod and bivalve lysins cannot be rejected, there is no clear evidence to substantiate homology.

In all, we sequenced eight alleles from \(M. trossulus\) (GenBank accession numbers AY131175 to AY131182), 10 alleles from \(M. edulis\) (GenBank accession numbers AY131165 to AY131174), and six alleles from \(M. galloprovincialis\) (GenBank accession numbers AY131159 to AY131164). Although Takagi et al. (1994) described their sequence from a Japanese mussel as being from an “\(M. edulis\)” individual, we include it with \(M. galloprovincialis\) because it is most similar to our Mediterranean \(M. galloprovincialis\) sequences and because introduced populations of \(M. galloprovincialis\) are the only “\(edulis\)” mussel in most of Japan (Suchanek et al. 1997). Thus, our sample size for \(M. galloprovincialis\) includes seven alleles.

There was variation at M7 lysin both between and within species. Estimates of heterozygosity (\(h\)) were fairly similar across the three species (table 1). Allelic variation within species was characterized by recombination. The Hudson and Kaplan (1985) method described considerable recombination within \(M. edulis\) and \(M. trossulus\) (table 1), and our limited survey of within-species diversity found instances of all four possible recombinant gametes for \(M. galloprovincialis\) (positions 52 and 292; fig. 1) and for \(M. edulis\) (positions 252 and 510; fig. 1). A direct consequence of recombination is that there is no single genealogy for the alleles sampled. However, each species was defined by unique substitutions. There were 28 fixed differences (11 synonymous and 17 amino acid replacements) between the \(M. trossulus\) alleles and \(M. edulis\) and \(M. galloprovincialis\) alleles (figs. 1 and 2). \(M. edulis\) and \(M. galloprovincialis\) were each defined by one nonsynonymous substitution (sites 56 and 144; fig. 2). Thus, the species relationships inferred from lysin are the same as those inferred from mtDNA, where the \(M. trossulus\) species branch is the longest and there is less divergence between \(M. edulis\) and \(M. galloprovincialis\) (Rawson and Hilbish 1995).

### Tests of Selection

Our tests of selection revealed an excess of fixed amino acid replacement differences between species, consistent with positive selection on the mussel lysin gene. A McDonald-Krietman test of the entire data set was statistically significant (\(P < 0.0167\)) (table 2), with nonsynonymous fixations exceeding synonymous fixations between species and synonymous polymorphisms exceeding nonsynonymous polymorphisms within species. McDonald-Krietman tests based on ratios of conservative to radical amino acid substitutions did not show an excess of radical fixed differences. In fact, the proportion of fixed differences was greater for conservative changes than for radical changes considering charge, polarity, and polarity with volume change.

While the McDonald-Krietman test demonstrated a pattern consistent with positive selection, the overall ratio of \(d_S/d_S\) was less than 1 (\(M. edulis\) vs. \(M. galloprovincialis\)).

### Table 1

<table>
<thead>
<tr>
<th>Species</th>
<th>No. Alleles</th>
<th>No. Unique Alleles</th>
<th>Segregating Sites</th>
<th>0 % (±SD)*</th>
<th>(\pi) % (±SD)*</th>
<th>Tajima’s (D)</th>
<th>4Ne/bp %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M. edulis)</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>0.651 (0.322)</td>
<td>0.524 (0.340)</td>
<td>-0.866</td>
<td>1.84</td>
</tr>
<tr>
<td>(M. galloprovincialis)</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>0.677 (0.363)</td>
<td>0.561 (0.380)</td>
<td>-0.906</td>
<td>—</td>
</tr>
<tr>
<td>(M. trossulus)</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>0.568 (0.304)</td>
<td>0.579 (0.381)</td>
<td>0.090</td>
<td>7.00</td>
</tr>
</tbody>
</table>

* Estimators of heterozygosity: \(0\) based on segregating sites; \(\pi\) based on average pairwise differences.

b Could not be estimated.
galloprovincialis: 0.25; M. edulis vs. M. trossulus: 0.48; M. galloprovincialis vs. M. trossulus: 0.48). A dN/dS less than 1 is expected if most replacement mutations are deleterious, which is likely to be the case for most protein coding genes. Sliding window analyses of dN/dS ratios located a region in the 3' end of the gene where dN/dS is greater than 1, but Fisher's exact test revealed that dN was not significantly greater than dS for any of the sliding window partitions. None of the maximum likelihood methods for identifying selection at specific sites or along specific branches yielded significant results. However, these methods assume no recombination and are not sensitive to changes along short branches such as those among closely related species (Anisimova, Bielawski, and Yang 2001).

Tajima's D statistic test of selection based on the frequency spectrum of polymorphisms did not show any strong deviations from neutrality (table 1). The runs test, sliding G-test, and Kolmogorov-Smirnov test (McDonald 1998) did not reveal significant heterogeneity in the polymorphism to divergence ratio within either M. edulis or M. trossulus. However, along the length of the M. galloprovincialis gene, the Kolmogorov-Smirnov test revealed significant heterogeneity, both for synonymous variation ($P = 0.026$) and for synonymous and non-synonymous variation combined ($P = 0.025$), with the

![FIG. 1.—Polymorphic nucleotide sites in the mature M7 lysin protein. Dots show identity to the consensus sequence. Species identities of alleles are described by the letters G, E, or T, where G is M. galloprovincialis, E is M. edulis, and T is M. trossulus, and A and B refer to two alleles within a single individual. For example, T776B is the B allele from M. trossulus individual 776. On the bottom row, S and N indicate whether polymorphisms are synonymous or non-synonymous (e.g., causing amino acid substitutions). In two cases where there are three substitutions at the same site (positions 321 and 489), there is a synonymous polymorphism within M. edulis and a fixed non-synonymous difference between M. trossulus and M. edulis/galloprovincialis.](image1)

![FIG. 2.—Alignment of M7 protein sequences from three mussel species. Dots show identity to the consensus sequence. Amino acid sites that are polymorphic within species are underlined. At two amino acid sites (positions 96 and 108, italicized) Mytilus trossulus is distinguished by two separate nucleotide substitutions, both of which are non-synonymous. Substitutions at sites 56 and 144 are unique to M. edulis and M. galloprovincialis, respectively, and are shown in boldface.](image2)
beginning of the sequence having a higher ratio of polymorphism to divergence than the latter portion (fig. 3).

**Discussion**

Positive Selection in Mussel Lysin

In this first survey of molecular variation of a bivalve sperm protein, we find evidence for positive selection shaping patterns of genetic variation between closely related species that frequently hybridize in nature. The strongest evidence for positive selection comes from the significant ($P < 0.0167$) McDonald-Kreitman test. If we assume that ratios of replacement to synonymous polymorphism within species approximate a neutral distribution, then the McDonald-Kreitman test points to an excess of fixed replacement substitutions between species. The proportion of adaptive substitutions can be estimated as $z = 1 - d_S P_N / d_R P_S$ (Smith and Eyre-Walker 2002), where $D_S$ and $D_N$ are the number of synonymous and nonsynonymous differences between species, and $P_S$ and $P_N$ are the number of synonymous and nonsynonymous polymorphisms. For this data set we estimate that 75% of the nonsynonymous differences between species are adaptive. Thus, the pattern of selection observed at mussel lysin confirms the emerging conclusion that marine invertebrate sperm proteins are frequently characterized by increased rates of amino acid substitutions.

Other studies have found $d_S/d_S$ greater than 1 for gastropod lysins. However, many genes undergoing adaptive evolution are likely to have $d_S/d_S$ less than 1, because the majority of nonsynonymous substitutions are probably deleterious even if a small fraction are adaptive. Here we find evidence for positive selection at a reproductively active protein even though $d_S/d_S$ is less than 1. This observation emphasizes the value of examining within-species polymorphism in addition to between-species divergence.

In addition to the significant McDonald-Kreitman test, significant heterogeneity in the polymorphism to divergence ratio along M7 lysin in $M. galloprovincialis$ suggests that adaptive evolution has occurred. One nonsynonymous nucleotide substitution occurred in the $M. galloprovincialis$ lineage after it split from $M. edulis$. This substitution is at amino acid site 144, near the 3′ end of the coding sequence, which has a lower polymorphism/divergence ratio than the 5′ end of the gene (fig. 3). Such a pattern is expected if a rapid, recent substitution at site 144 swept away the polymorphism in the 3′ end of the gene, while, because of recombination, the 5′ end of the gene was unaffected by the sweep. At least one intron of unknown length occurs in the lysin gene of *Mytilus* (Riginos, unpublished data), so it is not implausible that different parts of the coding region would have different evolutionary histories.

Comparisons of Mussel Lysin to Other Marine Invertebrate Sperm Proteins

Like other lysins, *Mytilus* M7 lysin is found on the sperm acrosome and dissolves the egg vitelline envelope (Takagi et al. 1994). However, its structure is not demonstrably similar to gastropod lysins, the only other molluscan lysins that have been described. Lysin proteins evolve quite rapidly, and so the lack of similarity is not surprising. For example, the rate of nonsynonymous substitutions per site per 10⁹ years is estimated at 6 to 84 for abalone (Metz, Robles-Sikisaka, and Vacquier 1998) and 21 to 76 for snails (Hellberg and Vacquier 1999). We estimate a rate of 6.5 for mussels, assuming a 3.5 million year divergence between *M. trossulus* and *M. edulis/galloprovincialis* (Vermeij 1991). Given these high substitution rates and a Cambrian gastropod-bivalve divergence (570 MYA; Benton 1993), recognizable sequence similarity between gastropod and bivalve lysins is not necessarily expected.

All marine invertebrate sperm proteins studied to date show evidence of adaptive evolution. However, the patterns of intraspecific polymorphism and rates of divergence are substantially different. Gastropod sperm proteins have virtually no polymorphism and divergence is extremely rapid (Lee, Ota, and Vacquier 1995; Metz, Robles-Sikisaka and Vacquier 1998; Hellberg and Vacquier 1999; Yang, Swanson, and Vacquier 2000), suggesting that recent selective sweeps have removed any intraspecific variation. Sea urchin bindin genes, in constrast, are characterized by an excess of nonsynonymous variation between and within species (Metz and

Table 2

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Fixed</th>
<th>Polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonymous</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>19</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note:* Fisher’s exact test, $P = 0.0167$.  

![FIG. 3.—Sliding window plot of the polymorphism to divergence ratio in *M. galloprovincialis*, with *M. trossulus* used for divergence. For each window of 10 variable sites, the proportion of variable sites that are polymorphic (combined synonymous and nonsynonymous) was calculated. Each value is plotted at the midpoint of the window. G indicates the position of the nonsynonymous substitution defining the *M. galloprovincialis* lineage at bp position 430. Kolmogorov-Smirnov tests revealed significant heterogeneity in the polymorphism to divergence ratio both for synonymous variation ($P = 0.026$) and for combined synonymous and nonsynonymous variation ($P = 0.025$).*
Palumbi 1996; Biermann 1998), which may result from positive selection on new amino acid–changing mutations or positive assortative mating (Palumbi 1999).

In mussels, we find a third pattern. As in gastropod lysins and urchin bindins, there is evidence that the majority of replacement substitutions between species are adaptive. However, unlike gastropod lysins, adaptive substitutions have not occurred so frequently that polymorphism has been swept away. Unlike urchin bindins, there is no significant evidence from the dN/dS ratio that amino acid polymorphisms are adaptive.

Clearly, more extensive surveys are needed to fully characterize the nature of intraspecific variation in *Mytilus* lysin. If lysin is directly involved in prezygotic isolation, then specific selection pressures may differ among populations, particularly those where hybridization does and does not occur. It will also be particularly interesting to examine this gene in populations of uncertain origin, such as those in the Baltic Sea (Rigosín, Sukdheo, and Cunningham 2002) and the Southern Hemisphere (Hilbish et al. 2000).

Given that all three *Mytilus* species surveyed here hybridize extensively in nature, our results imply that the action of selection on mussel lysin precedes the attainment of complete postzygotic reproductive isolation. Our results substantiate the emerging pattern that marine invertebrate sperm proteins are characterized by positive selection, particularly when ranges of closely related species overlap. It remains to be seen whether a common explanation underlies the general observation of adaptive evolution on genes involved in sexual reproduction.

**Acknowledgments**

We thank C. W. Cunningham, A. J. Abrams, B. Y. Kong, K. Zigler, and S. R. Palumbi for their assistance and comments throughout this project. Suggestions from three anonymous reviewers substantially improved this manuscript. This project was funded by the Duke University program in Molecular Evolution and Comparative Genomics, the PADI Foundation, and the National Science Foundation.

**Literature Cited**


Stephen Palumbi, Associate Editor

Accepted October 9, 2002