

Positive Selection and Propeptide Repeats Promote Rapid Interspecific Divergence of a Gastropod Sperm Protein

Michael E. Hellberg,*† Gary W. Moy,† and Victor D. Vacquier†

*Department of Biological Sciences, Louisiana State University at Baton Rouge; and †Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego

Male-specific proteins have increasingly been reported as targets of positive selection and are of special interest because of the role they may play in the evolution of reproductive isolation. We report the rapid interspecific divergence of cDNA encoding a major acrosomal protein of unknown function (TMAP) of sperm from five species of teguline gastropods. A mitochondrial DNA clock (calibrated by congeneric species divided by the Isthmus of Panama) estimates that these five species diverged 2–10 MYA. Inferred amino acid sequences reveal a propeptide that has diverged rapidly between species. The mature protein has diverged faster still due to high nonsynonymous substitution rates (>25 nonsynonymous substitutions per site per 10⁹ years). cDNA encoding the mature protein (89–100 residues) shows evidence of positive selection ($D_n/D_s > 1$) for 4 of 10 pairwise species comparisons. cDNA and predicted secondary-structure comparisons suggest that TMAP is neither orthologous nor paralogous to abalone lysin, and thus marks a second, phylogenetically independent, protein subject to strong positive selection in free-spawning marine gastropods. In addition, an internal repeat in one species (*Tegula aureotincta*) produces a duplicated cleavage site which results in two alternatively processed mature proteins differing by nine amino acid residues. Such alternative processing may provide a mechanism for introducing novel amino acid sequence variation at the amino-termini of proteins. Highly divergent TMAP N-termini from two other tegulines (*Tegula regina* and *Norrisia norrisii*) may have originated by such a mechanism.

Introduction

Although genes underlying cellular housekeeping functions commonly show little evolutionary change across phyla, genes intimately tied to mate recognition often diverge rapidly. cDNA sequences of mate recognition proteins have revealed that diversifying (positive) selection can promote their divergence. For example, in *Drosophila*, accessory gland proteins implicated in male mating success (Clark et al. 1995) display the signature of positive selection in an excess of nonsynonymous nucleotide substitutions relative to synonymous substitutions (Aguadé, Miyashita, and Langley 1992; Tsaur and Wu 1997; Tsaur, Ting, and Wu 1998; Aguadé 1999). Proteins that mediate interactions between single sexual cells seem especially likely to evince such rapid divergence promoted by positive selection. Such proteins occur on the surfaces of both unicellular sexual organisms (e.g., green algae [Ferris et al. 1997], yeast [Marsh and Herskowitz 1988], and basidiomycetous fungi [Bakkeren and Kronstad 1994]) and gametes (e.g., gastropods [Lee, Ota, and Vacquier 1995; Swanson and Vacquier 1995; Hellberg and Vacquier 1999] and echinoids [Metz and Palumbi 1996]). Reproductive proteins merit study because (1) they may offer clues as to how reproductive isolation evolves (Wu and Davis 1993; Palumbi 1998; Rice 1998), and (2) their accelerated rates of amino acid replacement facilitate observation of the signal of adaptive protein divergence in an arena relatively free of the neutral noise usually accompanying such change.

Proteins mediating the interaction between sperm and egg of free-spawning marine invertebrates are characterized by extensive interspecific divergence. In primitive marine gastropods, including abalone (*Haliotis*) and top snails (*Tegula*), the protein lysin performs a critical role in fertilization by dissolving a hole in a tough glycoproteinaceous envelope which surrounds the egg. Interspecific comparisons of lysin cDNA among closely related species of these gastropods reveal extensive divergence and rapid accumulation of nonsynonymous substitutions (Lee, Ota, and Vacquier 1995; Hellberg and Vacquier 1999). In abalone, a second major acrosomal protein also evolves extremely rapidly (Swanson and Vacquier 1995; Metz, Robles-Sikisaka, and Vacquier 1998).

Bindin, a sperm-egg attachment protein from sea urchins, likewise evolves rapidly (but see Metz, Gómez-Gutiérrez, and Vacquier 1998). Positive selection, however, is more localized within bindin than within gastropod acrosomal proteins (Metz and Palumbi 1996). Repetitive sequence elements play a substantial role in the interspecific divergence of bindin (Minor et al. 1991; Biermann 1998). To date, internal repeats have not been reported for gastropod acrosomal proteins.

Here, we report rapid interspecific divergence of cDNA encoding the major acrosomal protein (TMAP) from the sperm of five species of teguline gastropods: *Tegula aureotincta*, *Tegula brunnea*, *Tegula montereyi*, *Tegula regina*, and *Norrisia norrisii*. Previous work has established that *T. brunnea* and *T. montereyi* are sister taxa and that *T. regina* forms a monophyletic clade with this pair (Hellberg 1998). *Tegula aureotincta* and *N. norrisii* are relatively distant from this trio, and their relationships have not yet been resolved. A molecular clock that can be used to estimate divergence times (and, therefore, rates of substitution) between teguline species has also been calibrated (Hellberg and Vacquier 1999).

Abbreviation: TMAP, *Tegula* major acrosomal protein.

Key words: positive selection, fertilization, *Tegula*, sperm, prepro duplication, alternative processing.

Address for correspondence and reprints: Michael E. Hellberg, Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803. E-mail: mhellbe@lsu.edu.

Mol. Biol. Evol. 17(3):458–466. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

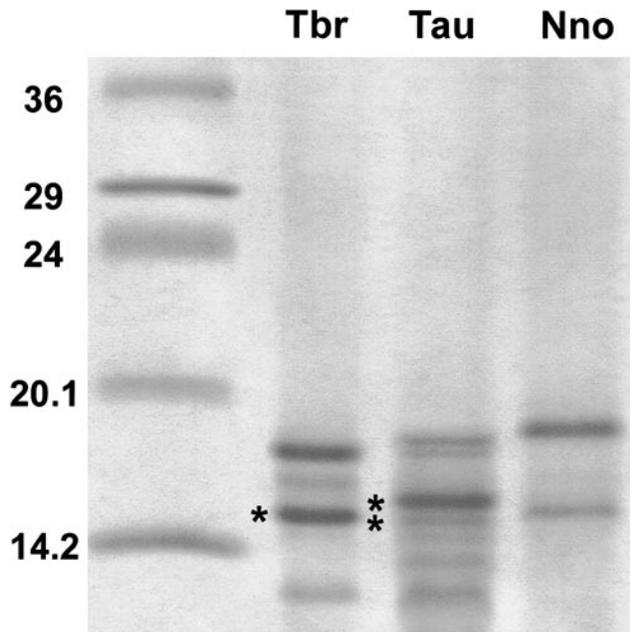


FIG. 1.—Coomassie-stained extracts of teguline gastropod spermatozoa separated by 17.5% SDS-PAGE. Twenty micrograms of protein were loaded into each lane. Lane 1, molecular weight marker; lane 2, *Tegula brunnea* (Tbr) extract; lane 3, *Tegula aureotincta* (Tau) extract; lane 4, *Norrissia norrisii* (Nno) extract. Asterisks mark the *Tegula* major acrosomal proteins (TMAPs) which were excised for direct amino acid sequencing.

We find that, like other acrosomal proteins from marine gastropods, TMAP exhibits high rates of nonsynonymous nucleotide substitution and positive selection between closely related species. In addition, a region containing the propeptide's cleavage site has been duplicated in one species (*T. aureotincta*), resulting in two mature peptides, one of which incorporates a portion of the presumed ancestral propeptide into the mature peptide. Such alternative processing may have given rise to the highly divergent N-termini seen in TMAP for two other species (*T. regina* and *N. norrisii*).

Materials and Methods

Protein Purification and Peptide Sequencing

Gametes were isolated from mature gonads as described previously (Hellberg and Vacquier 1999). SDS-acrylamide gels revealed that a ~15-kDa protein was *Tegula brunnea*'s major acrosomal protein (TMAP; fig. 1). To isolate TMAP, sperm were extracted in Ca^{+2} -free seawater with 1% Triton X-100. The low-molecular-weight proteins released by such extraction of sperm from primitive marine gastropods are identical to those found in seawater after inducing the acrosome reaction (Lewis, Talbot, and Vacquier 1982). Debris was pelleted by centrifugation at $30,000 \times g$ (20 min, 4°C), and the supernatant was dialyzed against 100 volumes of 250 mM NaCl, 2 mM EDTA, and 10 mM MES (pH 6), then centrifuged to clarify. The supernatant was applied to a 50-ml carboxymethyl cellulose column and washed with the above dialysis buffer. A linear gradient of equilibration buffer versus 100 ml of 950 mM NaCl, 2 mM

EDTA, and 10 mM MES (pH 6) was used to elute the bound protein. Fractions containing purified TMAP were used to assay its ability to dissolve egg vitelline envelopes (as in Hellberg and Vacquier 1999). These fractions were also used for N-terminal amino acid sequencing.

Further peptide sequence was obtained by excising TMAP from 17.5% SDS-PAGE gels and cleaving with CNBr (Nikodem and Fresco 1979). The resulting fragments were separated by 17.5% SDS-PAGE, transferred to polyvinylidene difluoride membranes, and subjected to gas phase sequencing.

mRNA Isolation, cDNA Synthesis, and Sequencing

Total RNA was isolated from testes (either fresh or preserved in 70% ethanol) by homogenization in 4 M guanidinium isothiocyanate, 25 mM sodium acetate (pH 6), 0.5% β -mercaptoethanol (Chomczynski and Sacchi 1987). The resulting homogenate was layered over 5.7 M CsCl (in 25 mM sodium acetate (pH 6) with 0.5% β -mercaptoethanol) and centrifuged for 18 h in a Beckman SW41 rotor at 26,000 rpm (20°C). cDNA was produced by reverse transcription using oligo-dT (for *T. brunnea* and *T. montereyi*) or primer TB3END (for the others, see below). A *T. brunnea* testis cDNA library (Lambda ZAP II, Stratagene) was constructed following the manufacturer's instructions.

cDNA was initially amplified by PCR using the *T. brunnea* testis cDNA library as template and degenerate primers based on amino acid sequences of TMAP peptides. A primer based on internal amino acid sequence ENRMKN (5'-GARAARAAYMGATGAARAA-3') was paired with the T7 vector primer. DNA sequence obtained from the resulting amplicon was used to design a primer specific to the 3' untranslated region of TMAP cDNA (TB3END: 5'-TGAAGTGCAGGTTATTTATTTCA-3'), which was paired with the T3 vector primer to complete the *T. brunnea* cDNA. This TB3END primer, in combination with a degenerate primer based on the amino acid sequence EAKIDYDY (5'-GARGCNAARATHGAYTAYGAYTA-3'), was used to amplify the 3' end of *T. aureotincta* TMAP cDNA. A reverse primer (5'-TARTCRTARTCDATYTTNGCYTC-3') based on the same amino acid sequence was primed with oligo-dC₂₂ to amplify the 5' end from dG-tailed *T. aureotincta* cDNA. Finally, TMAP from the remaining three species was amplified using TB3END and a forward primer based on signal sequence shared between *T. brunnea* and *T. aureotincta* (TMAPSIGSEQ: 5'-TGATGTTGTGTGTCGATCATATGG-3').

PCR reactions contained each primer at 0.5 μM (except that concentrations of degenerate primers were increased in direct proportion to their degeneracy), *Taq* polymerase at 10 U/ml, *Taq*Extender (Stratagene) at 10 U/ml, 1 \times *Taq*Extender buffer, 0.2 mM of each dNTP, and 1–2 μl of template DNA in a total volume of 50 μl . Thermal profiles consisted of 35 cycles of 40 s at 94°C, 2 min at 46°C, and 1.5 min at 72°C.

PCR products were either sequenced directly (*T. aureotincta*, *T. brunnea*) using amplification primers or

blunt-end cloned (*T. montereyi*, *T. regina*, *N. norrisii*) into pBlueScript, which was then used to transfect DH5 α -competent *Escherichia coli* cells. Both strands were sequenced using ABI Prism FS or BigDye chemistry. The five new sequences presented here have been assigned GenBank accession numbers AF190895–AF190899.

Southern Blotting

Southern blotting was used to determine gene copy number for *T. brunnea* TMAP. Genomic DNA digested with *Bgl*III, *Cla*I, *Eco*RI, *Eco*RV, *Hind*III, or *Xba*I was separated on a 0.6% TBE agarose gel and blotted onto Hybond N filters using the manufacturer's instructions. After UV cross-linking, the filter was probed with a radiolabeled *T. brunnea* TMAPSIGSEQ/TB3END PCR product.

Analysis of Protein and cDNA Sequences

Searches for proteins with sequences similar to the TMAPs were performed using BLASTp. Molecular weights and isoelectric points were calculated using MacVector. MacVector was also used to identify repeated sequence elements.

cDNA sequences were aligned by eye. Proportions of nonsynonymous (D_n) and synonymous (D_s) substitutions per site were calculated by method one of Ina (1995) using FENS (de Koning et al. 1998). Indels were dropped in pairwise fashion. The N-termini of *T. regina* and *N. norrisii* could not be aligned with any certainty and were excluded from the analysis. *t*-tests determined whether nonsynonymous substitutions were statistically more frequent than synonymous ones.

The scaled χ^2 method was used to assess codon usage bias (Shields et al. 1988). Nucleotide biases were calculated following Irwin, Kocher, and Wilson (1991). Because the purpose of these tests was to determine whether nucleotide or codon biases could have produced high D_n values, nonalignable sites (the N-termini of *T. regina*, *N. norrisii*, and the larger form of *T. aureotincta*) were excluded from these analyses.

Divergence times between pairs of teguline species were estimated using a molecular clock based on a 639-bp fragment of mitochondrial cytochrome oxidase I (mtCOI). This clock was previously calibrated at one silent transversion per million years using a pair of *Tegula* species (*T. verrucosa* and *T. viridula*) isolated by the rise of the Isthmus of Panama (Hellberg and Vacquier 1999). Although species presently separated by the Isthmus may have diverged long before the Isthmus' rise (Knowlton and Weight 1998), this particular pair belongs to a subgenus that arose 4 MYA and likely split 3 MYA (Coates and Obando 1996). Times of divergence were estimated conservatively using silent transversions (Irwin, Kocher, and Wilson 1991) instead of Kimura (1980) two-parameter distances, because the latter consistently gave more recent estimates of divergence (Hellberg and Vacquier 1999).

Secondary structure may reveal homologies between distantly related proteins even when DNA se-

quence comparisons cannot. Such was the case for lysin and an 18-kDa protein in abalone (Swanson and Vacquier 1995; Metz, Robles-Sikisaka, and Vacquier 1998). TMAP secondary structure was analyzed using tools available from PredictProtein (http://dodo.cpmc.columbia.edu/pp/submit_adv.html).

Secondary structure was inferred using PHDsec (Rost and Sander 1993, 1994). PHDsec employs neural networks trained on observed position-specific replacements to make predictions for secondary structure at individual sites in a target protein of unknown secondary structure. These initial predictions are refined by observed replacements in aligned input reference proteins. Each of the five teguline TMAPs was used in turn as a target protein, with the remaining four serving as reference proteins. PROSITE (Bairoch, Bucher, and Hofmann 1997) and ProDom (Corpet, Gouzy, and Kahn 1998) were used to search for functional motifs and putative domains, respectively.

Results

Abundance, Activity, and Sequencing of Proteins

The protein components of the acrosomal extracts of *T. aureotincta*, *T. brunnea*, and *N. norrisii* are shown in figure 1. The TMAP in the total extracts has an M_r of approximately 15 kDa and is the most abundant component of the acrosomal extract of the two *Tegula* species, although not of *N. norrisii*. Purified TMAP from *T. brunnea* did not dissolve conspecific egg vitelline envelopes even at concentrations as high as 1.4 mg/ml; thus, it does not possess lysin activity and its function remains unknown.

Gas phase sequencing of purified *T. brunnea* TMAP yielded 14 amino-terminal residues, beginning with Gly¹ (fig. 2). A CNBr fragment yielded an additional 45 contiguous residues. In *T. aureotincta*, TMAP resolved as two bands (fig. 1), both of which were subjected to gas phase sequencing. We also obtained an additional 28 internal residues of *T. aureotincta* sequence from a CNBr fragment (fig. 2).

Protein Divergence

Degenerate primers were used to amplify the full-length cDNA sequences of TMAP from *T. brunnea* and *T. aureotincta*. cDNA sequences for the pro- and mature peptides (no signal sequence due to primer position) of *T. montereyi*, *T. regina*, and *N. norrisii* were also obtained. Deduced amino acid sequences of these five TMAPs are shown in figure 2. Residues M⁻⁶⁰ to A⁻⁴⁴ of *T. aureotincta* represent a typical eukaryotic signal sequence, as do the corresponding residues of *T. brunnea*. M⁻⁶⁰ marks the first AUG in the mRNA of *T. aureotincta* and *T. brunnea*, and an adenine lies three bases upstream, consistent with this codon marking the start of translation (Kozak 1991).

In *T. aureotincta*, the sequence M⁻⁴³ to R⁻¹ represents a prepro sequence element of 43 residues. However, direct sequencing indicated that the N-terminus of the longer (and more abundant) of the two TMAPs appears within these 43 residues. This suggests that the

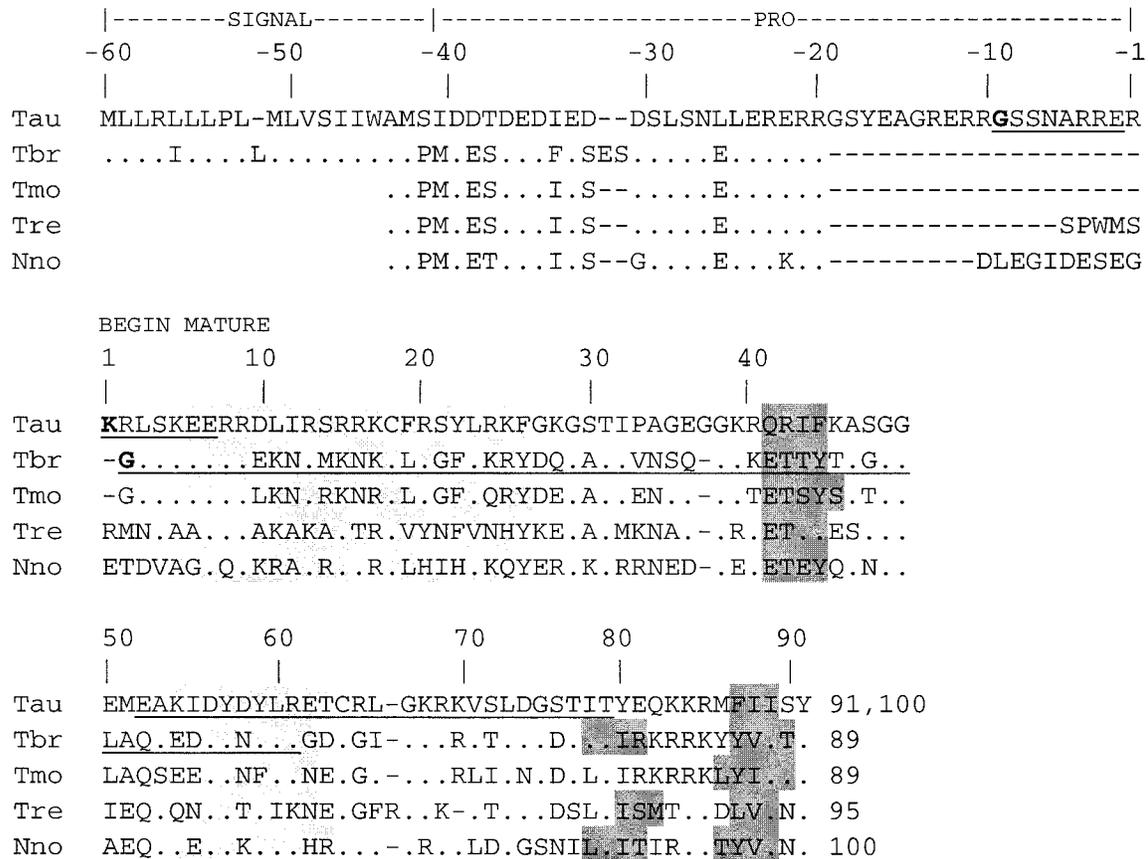


FIG. 2.—Deduced amino acid sequence for TMAP from five teguline gastropods. Position numbers refer to the *Tegula aureotincta* (Tau) sequence. Positions -60 to -44 represent the signal sequence. Underlined residues were determined by gas phase sequencing, with bold residues marking N-terminal residues. Light shaded regions indicate predicted α -helices; dark shaded regions indicate predicted β -sheets. Only those predictions for which three consecutive residues have a reliability of $\geq 70\%$ are shown. Dots denote identity to the Tau TMAP; dashes are inserted for alignment. Tbr = *Tegula brunnea*; Tmo = *Tegula montereyi*; Tre = *Tegula regina*; Nno = *Norrisia norrisii*.

two forms of *T. aureotincta* TMAP are alternatively processed mature forms of the same translation product, with the N-terminus of the longer form being G⁻⁹ and the N-terminus of the shorter form being K¹. The terminus of the longer form is adjacent to a furin cleavage site (R-X-R/K-R), typical of proteins with prepro regions. All species have additional furin sites occurring at R⁻²³. Adjacent to the N-terminal K⁻¹ of the shorter form is the sequence R⁻⁴-R-E-R⁻¹, which must be cleaved by another protease. The predicted starts of the propeptides of the four other species all align with position -42 of *T. aureotincta* and were either 26 (*T. brunnea*) or 24 residues long. The propeptides of the three *Tegula* species which mitochondrial sequences suggest are monophyletic (*T. brunnea*, *T. montereyi*, and *T. regina*; Hellberg 1998) are nearly identical, differing by only a two-amino acid insertion and an I→F replacement in *T. brunnea*. The presumed N-termini of mature TMAPs of *T. regina* and *N. norrisii* (-10 to +6 in fig. 2) are highly divergent (only 1 residue in 11 shared) and do not obviously align with any region of the other species.

The longer prepro peptide of *T. aureotincta* contains three imperfect repeats which align with the prepro region of the other four species (fig. 3). The identity

between *T. aureotincta* and the other species is greatest for the first *T. aureotincta* repeat (positions -31 to -20). The two other elements apparently arose by duplication of this 12-amino acid segment.

The mature lengths of TMAPs from the five species vary between 100 (in *T. aureotincta* and *N. norrisii*) and 89 (in *T. brunnea* and *T. montereyi*) residues. All are highly basic, with computed isoelectric points between 9.7 and 10.9. In the larger form of *T. aureotincta* TMAP, 40 of 100 residues are charged. Computed molecular weights are smaller than those estimated by PAGE, varying from 10.3 to 11.9 kDa.

The alignable portions of the mature proteins from the five species vary in amino acid identity from 36% to 72% (table 1). Both cysteine residues do not vary among the five species, nor do most positions occupied by aromatic residues (Y and F). There were no significant matches to GenBank or to any recognized functional motifs or domains. The TMAP signal sequence matches the lysin signal sequence at only 2 of 18 alignable residues (one of them being the start methionine) in *T. brunnea*, the only species for which complete signal sequences are available for both molecules. In contrast, the signal sequences of the distant homologs lysin

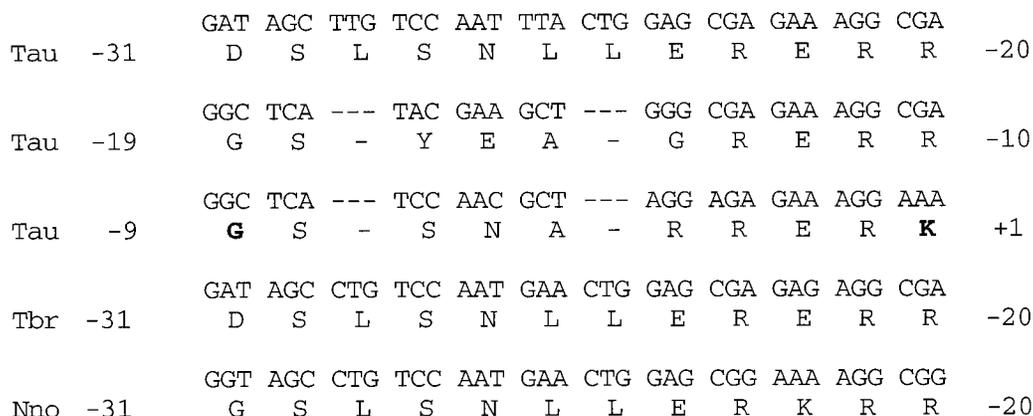


FIG. 3.—Alignment of the three imperfect repeats of the propeptide of *Tegula aureotincta* (Tau) TMAP. Homologous positions from *Tegula brunnea* (Tb) and *Norrisia norrisii* (Nn) are shown for comparison. Bold residues mark N-termini as indicated by gas phase sequencing. The first *T. aureotincta* repeat is most similar to sequences from other species. Direct sequencing suggests that endolytic cleavage takes place at Arg⁻¹ and Arg⁻¹¹ in *T. aureotincta*, but at the Arg homologous to position -21 in *T. brunnea*. Position numbers refer to alignment in figure 2.

and 18-kDa of *Haliotis rufescens* match at 8 of 16 alignable residues.

The predicted secondary structure of TMAP differs from those of other known gastropod acrosomal proteins, which are strongly α -helical (Shaw et al. 1993; Swanson and Vacquier 1995; Hellberg and Vacquier 1999). Two α -helices are predicted for TMAP (fig. 2): one near the N-terminus and the other in the middle of the protein. All TMAPs are predicted to have lower proportions of α -helices than any lysin or 18-kDa acrosomal protein. Short β -sheets are predicted to occur between the α -helices and toward the C-terminus of the protein. No β -sheets are predicted for any lysin or 18-kDa acrosomal protein.

Rates of Nucleotide Divergence

Using the cytochrome oxidase I silent-transversion molecular clock, estimated times of divergence for the

five species range from 4 to 20 Myr (table 2). Nonsynonymous substitution rates for the mature protein based on these times of divergence are high: between 26.3 and 60.5 per site per billion years (table 2). Synonymous substitution rates are similarly high for most comparisons but are less than one third the nonsynonymous rate for the contrast between the sister species *T. brunnea* and *T. montereyi* (Table 2). The Kimura two-parameter clock (not shown) estimated shorter times of divergence than did the silent-transversion clock and, hence, higher TMAP substitution rates than those presented here.

Positive Selection

D_n and D_s were calculated for full-length TMAP cDNAs. D_n values greater than D_s suggest that positive selection promotes sequence divergence. Table 1 shows that D_n is significantly greater than D_s for 4 of the 10 possible pairwise comparisons: the one involving the

Table 1
Pairwise Amino Acid Identities and Estimates of Per-Site Proportions of Nonsynonymous and Synonymous Nucleotide Substitutions (calculated using Method 1 of Ina 1995) for Alignable^a Regions of the Propeptide and Mature Protein Encoded by TMAP cDNA

SPECIES COMPARISON	PROPEPTIDE				MATURE PROTEIN			
	% Amino Acid Identity	D_n (\pm SE)	D_s (\pm SE)	D_n/D_s	% Amino Acid Identity	D_n (\pm SE)	D_s (\pm SE)	D_n/D_s
Tau-Tbr	70.8	0.187 (0.064)	0.209 (0.136)	0.89	42.7	0.487 (0.65)	0.450 (0.108)	0.95
Tau-Tmo	70.8	0.154 (0.056)	0.161 (0.125)	0.95	41.6	0.543 (0.071)	0.459 (0.108)	1.18
Tau-Tre	70.8	0.148 (0.054)	0.296 (0.190)	0.50	36.3	0.655 (0.085)	0.640 (0.154)	1.02
Tau-Nno	62.5	0.213 (0.071)	0.297 (0.178)	0.72	37.1	0.557 (0.075)	0.569 (0.137)	0.98
Tbr-Tmo	95.8	— ^b	0.0	—	71.9	0.176 (0.033)	0.050 (0.027)	3.51*
Tbr-Tre	95.8	0.018 (0.018)	0.130 (0.095)	0.14	42.0	0.471 (0.064)	0.325 (0.093)	1.45
Tbr-Nno	83.3	0.105 (0.106)	0.167 (0.048)	0.63	40.4	0.508 (0.071)	0.234 (0.071)	2.17*
Tmo-Tre	100.0	0.000 (0.000)	0.128 (0.133)	0.00	40.9	0.484 (0.066)	0.356 (0.092)	1.36
Tmo-Nno	87.5	0.081 (0.041)	0.113 (0.085)	0.72	40.4	0.473 (0.068)	0.268 (0.076)	1.76**
Tre-Nno	87.5	0.073 (0.037)	0.156 (0.114)	0.47	38.2	0.474 (0.066)	0.268 (0.081)	1.77**

NOTE.—Tau = *Tegula aureotincta*; Tbr = *Tegula brunnea*; Tmo = *Tegula montereyi*; Tre = *Tegula regina*; Nno = *Norrisia norrisii*.

^a Alignments begin with residue 9 of Tre (+4 in fig. 2) and residue 17 of Nno (+7 in fig. 2).

^b No calculation owing to zero synonymous substitutions.

* Significant at the $P < 0.005$ level.

** Significant at the $P < 0.05$ level.

Table 2
Pairwise Estimates of Rates of Nonsynonymous and Synonymous Substitutions for Alignable^a Regions of TMAP cDNA Based on Divergence Times Estimated Using a COI Molecular Clock

Species Comparison	Estimated Divergence ^b	Nonsynonymous Rate ^c	Synonymous Rate ^c
Tau-Tbr.....	18	27.0	25.0
Tau-Tmo.....	18	30.2	25.5
Tau-Tre.....	18	36.4	35.5
Tau-Nno.....	20	27.8	28.4
Tbr-Tmo.....	4	44.0	12.5
Tbr-Tre.....	8	58.9	40.6
Tbr-Nno.....	18	28.2	13.0
Tmo-Tre.....	8	60.5	44.5
Tmo-Nno.....	18	26.3	14.9
Tre-Nno.....	18	26.3	14.9

NOTE.—Tau = *Tegula aureotincta*; Tbr = *Tegula brunnea*; Tmo = *Tegula montereyi*; Tre = *Tegula regina*; Nno = *Norrissia norrisii*.

^a Alignments begin with position +7 in figure 2.

^b Time of divergence (in Myr) estimated using a cytochrome oxidase I clock calibrated by the number of synonymous transversions (3) between *Tegula viridula* and *Tegula verrucosa*, two species separated by the rise of the Isthmus of Panama 3 MYA.

^c Rates in substitutions per site per 10⁹ years based on D_n and D_s calculated using method 1 of Ina (1995).

sister species *T. montereyi* and *T. brunnea* and three involving *N. norrisii*. All other values are close to unity, with the exception of a few low values calculated for propeptide comparisons (none of these were significantly less than 1).

Bias in nucleotide usage at silent sites or in codon usage could result in underestimates of D_s , leading to inaccurate conclusions of positive selection (Ticher and Graur 1989). The percentage of G+C shows little bias over the full coding region (45.2%–48.0%) or at third positions of codons (46.7%–53.6%; table 3). The percentage of C at third positions varies from 21.1% to 28.1%. Nucleotide usage falls toward the low end of the theoretical range (from 0 = no bias to 1 = maximum bias; Irwin, Kocher, and Wilson 1991). Codon usage biases are also low (table 3): all values are lower than that of chymotrypsin from the abalone *H. rufescens* (Lee 1994). These data suggest that neither nucleotide nor codon usage bias can account for the significant excess of D_n relative to D_s seen for four of the TMAP interspecific comparisons.

Comparison of nonorthologous gene regions could likewise create misleading D_n/D_s ratios. The duplication of entire genes necessary for such an explanation should leave multiple copies in the genome, especially if the comparison involves species that diverged recently (as have *T. brunnea* and *T. montereyi*). Southern blot analysis (fig. 4) suggests that this is not the case for TMAP. *Tegula brunnea* TMAP probes produced a single band of hybridization, suggesting that TMAP is a single-copy gene in this species.

Discussion

Positive Selection

The acrosomal protein (TMAP) of unknown function studied here evolves at extremely high rates: over

Table 3
Nucleotide and Codon Bias in TMAP

Species	% G+C	Third-Position % G+C	Third-Position % C	Nucleotide Bias	Third-Position Nucleotide Bias	Codon Bias
<i>Tegula aureotincta</i> ..	45.2	46.7	21.1	0.178	0.148	0.140
<i>Tegula brunnea</i>	46.8	52.8	28.1	0.198	0.113	0.125
<i>Tegula montereyi</i> . . .	47.2	52.8	24.7	0.177	0.112	0.244
<i>Tegula regina</i>	46.7	52.9	26.4	0.176	0.038	0.196
<i>Norrissia norrisii</i>	48.0	53.6	26.2	0.178	0.148	0.286

25 nonsynonymous changes per nonsynonymous site per billion years (table 2). These rates are over four times as great as those for any full-length protein tabulated by Li (1997, p. 191) for *Drosophila*, an organism with generation times far shorter than those of *Tegula*

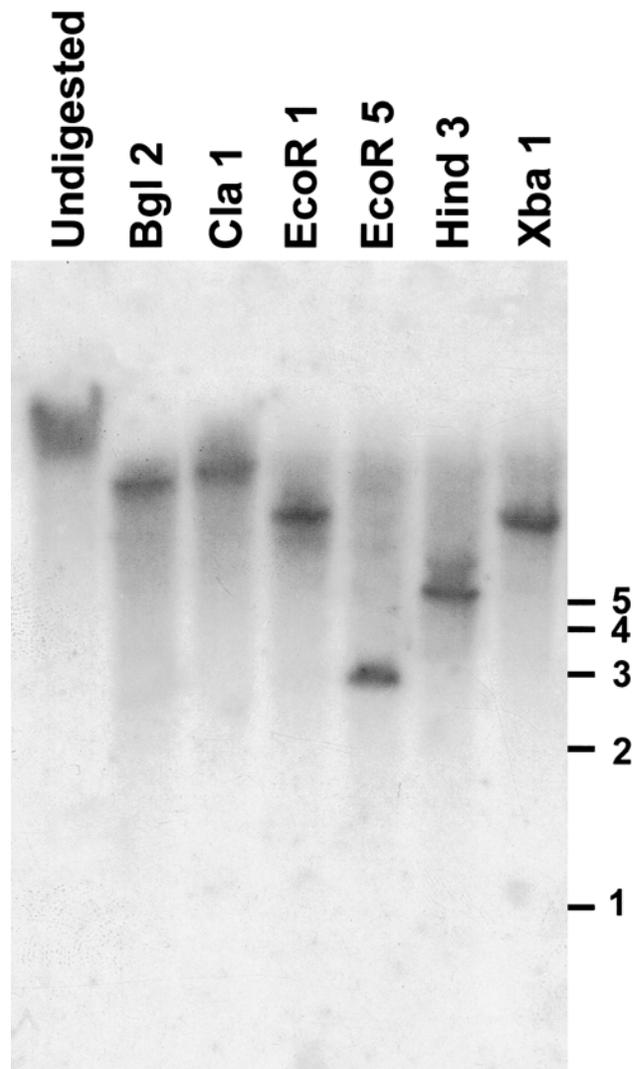


FIG. 4.—Southern blot hybridization of radiolabeled *Tegula brunnea* TMAP cDNA to restriction enzyme-digested genomic DNA from the same species, indicating that the TMAP gene occurs as a single copy. Positions of kilobase ladder size standards are indicated at right.

species (Paine 1971; Horikawa and Yamakawa 1982). Such rapid rates have also been found for other acrosomal proteins from *Haliotis* (abalone) and *Tegula* (Metz, Robles-Sikisaka, and Vacquier 1998; Hellberg and Vacquier 1999) and for channel-blocking toxins produced by the predatory gastropod *Conus* (Duda and Palumbi 1999).

Four comparisons of the mature TMAP reveal a significant excess of nonsynonymous substitutions relative to synonymous substitutions (table 1). Several observations suggest that this significant excess of D_n relative to D_s results from selection for amino acid change in TMAP. First, relatively high values of D_n are restricted to the mature protein (table 1). D_n/D_s values for the immediately adjacent propeptide are below unity (although not significantly so). Signal sequences from *T. aureotincta* and *T. brunnea* (not shown) evolve still more slowly than the propeptide. Similar relative rates of nonsynonymous change (conserved signal, moderate propeptide, rapid mature protein) have been reported for interlocus divergence of three mating pheromones from the ciliate *Euplotes raikovi*, another instance of reproductive protein radiation marked by extensive amino acid replacements (Miceli et al. 1991). Second, the relatively high value of D_n is not likely due to either nucleotide or codon bias (table 3; Lee 1994), although the species with the highest codon bias (*N. norrisii*) is involved in three of the four significant D_n/D_s comparisons.

Finally, the D_n/D_s value of >1 is not likely to be due to the comparison of nonorthologous loci, at least for the species pair yielding the highest ratio (*T. brunnea/T. montereyi*). Southern analysis of *T. brunnea* shows TMAP to be a single-copy gene in this species (fig. 4), and the probability of duplication and subsequent extinction of one copy of the gene during the brief time (≤ 2 Myr) separating *T. brunnea* and *T. montereyi* seems low. TMAP PCR products were directly sequenced in *T. aureotincta*, suggesting that TMAP occurs as a single mRNA and is probably a single-copy gene in this species as well. Results for *T. regina* and *N. norrisii* (which were cloned) are less clear; ultimately, Southern analysis of all of these species will be needed to ascertain copy number.

Six of 10 interspecific TMAP comparisons did not show $D_n/D_s > 1$. Comparisons of gastropod acrosomal proteins generally do not exceed unity when $D_s > 0.2$ (see Hellberg and Vacquier 1999), regardless of whether species co-occur (Lee, Ota, and Vacquier 1995). Here, $D_s > 0.2$ for 9 of the 10 pairwise comparisons (the *T. brunnea/T. montereyi* pair, with the highest D_n/D_s value, being the sole exception). Thus, relatively low D_n/D_s values probably result from the downward bias of estimators of D_n/D_s when divergence is great (Ina 1995).

The function of TMAP remains unknown; purified TMAP did not dissolve vitelline envelopes. Previous assays of lysin activity (Hellberg and Vacquier 1999) used whole acrosomal extracts enriched for lysin, so the possibility remains that TMAP serves some role with lysin in dissolving vitelline envelopes. However, observed dissolution activity in those experiments varied directly

with the proportion of lysin in the preparation, suggesting little role for TMAP. The localization of TMAP within the acrosome strongly suggests some role in fertilization.

Strong interspecific positive selection has previously been reported for lysins from *Haliotis* (Lee, Ota, and Vacquier 1995) and *Tegula* (Hellberg and Vacquier 1999) and for an 18-kDa acrosomal protein from *Haliotis* (Swanson and Vacquier 1995). Loci encoding these other gastropod fertilization proteins are either orthologous (the two lysins) or paralogous (the two *Haliotis* proteins; see Metz, Robles-Sikisaka, and Vacquier 1998) to each other. Comparisons of these to cDNA and predicted secondary structure of TMAP do not suggest any obvious relationship. Thus, TMAP and lysin appear to be two historically independent, male-specific sex proteins, both experiencing strong diversifying selection between species.

One possible explanation for positive selection on fertilization proteins is to avoid heterospecific fertilization. The high D_n/D_s value for the *T. brunnea/T. montereyi* comparison is striking in this light, as these are co-occurring sister species with significant overlap in microhabitat (Riedman, Hines, and Pearse 1981) and spawning season (Watanabe 1982). However, in gastropod sperm proteins, D_n/D_s ratios are generally highest for closely related species (Lee, Ota, and Vacquier 1995; Swanson and Vacquier 1995), and closely related species tend to be sympatric among these taxa (Hellberg 1998), so this single observation can be regarded as merely consistent with a role for reinforcement.

Propeptide Repeats and Alternative Processing

The propeptide of *T. aureotincta* is 20 residues longer than those of *T. brunnea* and *T. montereyi* (fig. 2). *Tegula aureotincta* has two ~ 12 -residue repeats showing 60% amino acid identity and 77% nucleotide identity to each other (fig. 3). The first of the three *T. aureotincta* repeats is more similar to presumed homologous sites in the other four species analyzed. The two repeats must have originated by duplication of the first 12-residue region (fig. 3).

Most interestingly, the duplication includes at its C-terminal end a dibasic repeat (RR or RK), the usual recognition sequence for the endolytic cleavage which separates propeptide regions from mature peptides (Bond and Butler 1987). Direct sequencing of two gel-purified acrosomal proteins from *T. aureotincta* confirmed that two different mature proteins, one corresponding to each of the duplicated cleavage sites (fig. 2), are expressed.

The net result of the duplication of cleavage sites is that some amino acid residues previously restricted to the propeptide are, under one alternative processing, incorporated into the mature peptide. As with intron capture (Golding, Tsao, and Pearlman 1994), and unlike exon shuffling or duplications of regions already encoding mature peptides, such propeptide capture should have the effect of introducing truly novel sequence into a mature protein. The nonalignable N-terminal residues

of *T. regina* and *N. norrisii* (fig. 2) may have been introduced initially in such a fashion, with subsequent deletions and substitutions leaving no trace of the duplication.

Incorporation of residues that alter protein structure might be expected to have negative selective consequences. Such consequences, however, may be limited for TMAP. The N-terminal differences would not alter the distance between the two conserved cysteines; thus, forms both with and without the N-terminal residues would be expected to have similar folds. Furthermore, gastropod acrosomal proteins often show interspecific length variation of several residues at their amino- and carboxy-termini (Lee, Ota, and Vacquier 1995; Swanson and Vacquier 1995). In addition, overcoming the potential selective barrier of replacing an ancestral *T. aureotincta* 90-residue TMAP with one 10 residues larger may have been facilitated by the fact that both mature proteins would initially be produced (Smith, Patton, and Nadal-Ginard 1989). Alternative processing may thus provide another genetic mechanism, along with positive selection on point mutations, for promoting diversification of reproductive proteins.

Acknowledgments

We thank W. Swanson and E. Metz for many helpful discussions and for reviewing the manuscript, P. Arbour-Reily, R. Bouchard, and D. Taranek for technical assistance, J. Leichter and L. Tomanek for collecting samples from Pacific Grove, and W. Swanson and B. Rogers for assistance with computer programs. This work was supported by NIH grant HD12986 to V.D.V., by an NSF Marine Biotechnology Postdoctoral Fellowship (OCE 9321243) to M.E.H., and by startup funds for M.E.H. from Louisiana State University.

LITERATURE CITED

- AGUADÉ, M. 1999. Positive selection drives the evolution of Acp29AB accessory gland protein in *Drosophila*. *Genetics* **152**:543–551.
- AGUADÉ, M., N. MIYASHITA, and C. H. LANGLEY. 1992. Polymorphism and divergence in the Mst26A male accessory gland gene region in *Drosophila*. *Genetics* **132**:755–770.
- BAIROCH, A., P. BUCHER, and K. HOFMANN. 1997. The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**: 217–221.
- BAKKEREN, G., and J. W. KRONSTAD. 1994. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous fungi. *Proc. Natl. Acad. Sci. USA* **91**: 7085–7089.
- BIERMANN, C. H. 1998. The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Mol. Biol. Evol.* **15**:1761–1771.
- BOND, J. S., and P. E. BUTLER. 1987. Intracellular proteases. *Annu. Rev. Biochem.* **56**:333–364.
- CHOMCZYNSKI, P., and N. SACCHI. 1987. Single-step method of RNA isolation by guanidine-thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**:156–159.
- CLARK, A. G., M. AGUADÉ, T. PROUT, L. G. HARSHMAN, and C. H. LANGLEY. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**:189–201.
- COATES, A. G., and J. A. OBANDO. 1996. The geological evolution of the Central American Isthmus. Pp. 21–56 in J. B. C. JACKSON, A. F. BUDD, and A. G. COATES, eds. *Evolution and environment in tropical America*. University of Chicago Press, Chicago.
- CORPET, F., J. GOUZY, and D. KAHN. 1998. The ProDom database of protein domain families. *Nucleic Acids Res.* **26**: 323–326.
- DE KONING, J., M. PALUMBO, W. MESSIER, and C.-B. STEWART. 1998. FENS, facilitated estimates of nucleotide substitutions. Version 0.9. Distributed by C.-B. Stewart, Department of Biological Sciences, State University of New York, Albany.
- DUDA, T. F. JR., and S. R. PALUMBI. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **96**:6820–6823.
- FERRIS, P. J., C. PAVOLVIC, S. FABRY, and U. W. GOODENOUGH. 1997. Rapid evolution of sex-related genes in *Chlamydomonas*. *Proc. Natl. Acad. Sci. USA* **94**:8634–8639.
- GOLDING, G. B., N. TSAO, and R. E. PEARLMAN. 1994. Evidence for intron capture: an unusual path for the evolution of proteins. *Proc. Natl. Acad. Sci. USA* **91**:7506–7509.
- HELLBERG, M. E. 1998. Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropod *Tegula*. *Evolution* **52**:1311–1324.
- HELLBERG, M. E., and V. D. VACQUIER. 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Mol. Biol. Evol.* **16**:839–848.
- HORIKAWA, H., and H. YAMAKAWA. 1982. Ecological study of *Omphalius pfeifferi* Philippe (Gastropoda: Prosobranchia). *Bull. Nansei Reg. Fish. Res. Lab.* **14**:71–81.
- INA, Y. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J. Mol. Evol.* **40**:190–226.
- IRWIN, D. M., T. D. KOCHER, and A. C. WILSON. 1991. Evolution of cytochrome *b* gene in mammals. *J. Mol. Evol.* **32**: 128–144.
- KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- KNOWLTON, N., and L. A. WEIGHT. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc. R. Soc. Lond. B Biol. Sci.* **265**:2257–2263.
- KOZAK, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**: 19867–19870.
- LEE, Y.-H. 1994. Abalone sperm lysin: molecular evolution of a fertilization protein, implications concerning the species-specificity of fertilization and speciation in marine invertebrates. Ph.D. dissertation, University of California, San Diego.
- LEE, Y.-H., T. OTA, and V. D. VACQUIER. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol. Biol. Evol.* **12**:231–238.
- LEWIS, C. A., C. F. TALBOT, and V. D. VACQUIER. 1982. A protein from abalone sperm dissolves the egg vitelline layer by a non-enzymatic mechanism. *Dev. Biol.* **92**:227–239.
- LI, W. 1997. *Molecular evolution*. Sinauer, Sunderland, Mass.
- MARSH, L., and I. HERSKOWITZ. 1998. STE2 protein of *Saccharomyces kluyveri* is a member of the rhodopsin/adrenergic receptor family and is responsible for recognition of the peptide ligand factor. *Proc. Natl. Acad. Sci. USA* **85**: 3855–3859.
- METZ, E. C., G. GÓMEZ-GUTIÉRREZ, and V. D. VACQUIER. 1998. Mitochondrial DNA and bindin gene sequence evolution of the Central American Isthmus. Pp. 21–56 in J. B. C. JACKSON, A. F. BUDD, and A. G. COATES, eds. *Evolution and environment in tropical America*. University of Chicago Press, Chicago.

- lution among allopatric species of the sea urchin genus *Arbacia*. *Mol. Biol. Evol.* **15**:185–195.
- METZ, E. C., and S. R. PALUMBI. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**:397–406.
- METZ, E. C., R. ROBLES-SIKISAKA, and V. D. VACQUIER. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **95**:10676–10681.
- MICELI, C., A. LATERZA, R. A. BRADSHAW, and P. LUPORINI. 1991. Structural characterization of mating pheromone precursors of the ciliate protozoan *Euplotes raikovi*: high conservation of pre and pro regions versus high variability of secreted regions. *Eur. J. Biochem.* **202**:759–764.
- MINOR, J. E., D. R. FROMSON, R. J. BRITTEN, and E. H. DAVIDSON. 1991. Comparison of the bindin proteins of *Strongylocentrotus franciscanus*, *S. pupuratus*, and *Lytechinus variegatus*: sequences involved in the species specificity of fertilization. *Mol. Biol. Evol.* **8**:781–795.
- NIKODEM, V., and J. R. FRESCO. 1979. Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr. *Anal. Biochem.* **97**:382–386.
- PAINE, R. T. 1971. Energy flow in a natural population of the herbivorous gastropod *Tegula funebralis*. *Limnol. Oceanogr.* **16**:86–98.
- PALUMBI, S. R. 1998. Species formation and the evolution of gamete recognition loci. Pp. 271–278 in D. HOWARD and S. H. BERLOCHER, eds. *Endless forms: species and speciation*. Oxford University Press, Oxford, England.
- RICE, W. R. 1998. Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. Pp. 261–270 in D. HOWARD and S. H. BERLOCHER, eds. *Endless forms: species and speciation*. Oxford University Press, Oxford, England.
- RIEDMAN, M. L., A. H. HINES, and J. S. PEARSE. 1981. Spatial segregation of four species of turban snails (Gastropoda: *Tegula*) in central California. *Veliger* **24**:97–102.
- ROST, B., and C. SANDER. 1993. Improved prediction of protein structure at better than 70% accuracy. *J. Mol. Biol.* **232**:584–599.
- . 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**:55–72.
- SHAW, A., D. E. MCREE, V. D. VACQUIER, and C. D. STOUT. 1993. The crystal structure of lysin, a fertilization protein. *Science* **262**:1864–1867.
- SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and W. WRIGHT. 1988. “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**:704–716.
- SMITH, C. W. J., J. G. PATTON, and B. NADAL-GINARD. 1989. Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**:527–577.
- SWANSON, W. J., and V. D. VACQUIER. 1995. 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.
- TICHER, A., and D. GRAUR. 1989. Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. *J. Mol. Evol.* **28**:286–298.
- TSAUR, S.-C., C.-T. TING, and C.-I. WU. 1998. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. *Mol. Biol. Evol.* **15**:1040–1046.
- TSAUR, S.-C., and C.-I. WU. 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. *Mol. Biol. Evol.* **14**:544–549.
- WATANABE, J. M. 1982. Aspects of community organization in a temperate kelp forest habitat: factors influencing the bathymetric segregation of three species of herbivorous gastropods. Ph.D. dissertation, University of California, Berkeley.
- WU, C.-I., and A. W. DAVIS. 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane’s rule and its genetic basis. *Am. Nat.* **142**:187–212.

SHOZO YOKOYAMA, reviewing editor

Accepted December 3, 1999