J Mol Evol (2000) 51:481–490
DOI: 10.1007/s002390010111

JOURNAL OF MOLECULAR EVOLUTION

Evaluation of Sequence Variation and Selection in the Bindin Locus of the Red Sea Urchin, *Strongylocentrotus franciscanus*

Patty Debenham,*,* Mark A. Brzezinski,* Kathy R. Foltz*****

Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

Received: 19 June 1998 / Accepted: 2 August 2000

Abstract. Recent evidence suggests that gamete recognition proteins may be subjected to directed evolutionary pressure that enhances sequence variability. We evaluated whether diversity enhancing selection is operating on a marine invertebrate fertilization protein by examining the intraspecific DNA sequence variation of a 273-base pair region located at the 5′ end of the sperm bindin locus in 134 adult red sea urchins (*Strongylocentrotus franciscanus*). Bindin is a sperm recognition protein that mediates species-specific gamete interactions in sea urchins. The region of the bindin locus examined was found to be polymorphic with 14 alleles. Mean pairwise comparison of the 14 alleles indicates moderate sequence diversity (p-distance = 1.06). No evidence of diversity enhancing selection was found. It was not possible to reject the null hypothesis that the sequence variation observed in *S. franciscanus* bindin is a result of neutral evolution. Statistical evaluation of expected proportions of replacement and silent nucleotide substitutions, observed versus expected proportions of radical replacement substitutions, and conformance to the McDonald and Kreitman test of neutral evolution all indicate that random mutation followed by genetic drift created the polymorphisms observed in bindin. Observed frequencies were also highly similar to results expected for a neutrally evolving locus, suggesting that the polymorphism observed in the 5′ region of *S. franciscanus* bindin is a result of neutral evolution.

Key words: Sea urchin — Intraspecific variation — Bindin — Diversifying selection — Positive selection

Introduction

Among sexually reproducing organisms, many marine invertebrates exhibit unique life history and reproductive strategies dramatically different from those of most terrestrial animals. For organisms such as echinoderms, abalone, polychaetes, and bivalves, adults release gametes into the water column which then fuse species-specifically and develop into a motile larval phase (Strathmann 1987). In most cases, there is limited, if any, interaction between the adults. If any type of signal for assortative mating exists in animals using broadcast spawning as a reproductive strategy, it is most likely limited to interactions between the gametes themselves (Palumbi 1992, 1994; Metz and Palumbi 1996; Vacquier 1998).

Gamete recognition proteins play a critical role in mediating sperm–egg interactions (reviewed by Ward and Kopf 1993; Vacquier 1998). High levels of nucleotide variation in the genes encoding some of these recognition proteins has led to the hypothesis that these genes are subjected to directed (nonneutral) selective pressure (Vacquier and Lee 1993; Lee et al. 1995; Metz

* Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA 93106, USA
** *Current address:* National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, CA 93101, USA
*** Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, Santa Barbara, CA 93106, USA
Correspondence to: Dr. Patty Debenham; e-mail: pdebenham@seaweb.org

and Palumbi 1996; Palumbi 1995; Vacquier 1998). Unusual intraspecific sequence variation represented by high polymorphism and high proportions of replacement substitutions has been predicted for the gamete interaction molecules of free spawning invertebrates and specifically urchins (Metz and Palumbi 1996). The work presented here is a test of this hypothesis.

The neutral theory of evolution (Kimura 1983; Nei and Gojobori 1986) predicts that, assuming that synonymous substitutions are completely neutral, the rate of nonsynonymous substitutions per nonsynonomous site (D_n) should be less than or equal to but no greater than the rate of synonymous substitutions per synonymous site (D_s). This prediction recognizes the fact that all protein coding sequences are subject to a certain amount of selective constraint. When D_n is significantly greater than D_s in comparisons between species, the action of divergent evolution is suggested. When D_n is greater than D_s among individuals of the same species, diversity-enhancing selection is suggested (Nei and Gojobori 1986; Hughes and Nei 1988; Hughes et al. 1990). The term "positive selection" has been applied to both processes (Hughes and Nei 1988; Lee et al. 1995; Endo et al. 1996). This type of intraspecific selective pressure that generates sequence diversity in a gamete recognition molecule is particularly intriguing because it opposes the standard evolutionary theory that predicts that traits closely related to fitness (such as successful gamete interaction) are strictly conserved (Nei 1987, Clark et al. 1995).

High D_n-to-D_s ratios have been reported in two gamete interaction molecules of marine invertebrates, supporting the idea that selection results in rapid interspecific sequence divergence among related species (Vacquier and Lee 1993; Lee et al. 1995; Metz and Palumbi 1996, Vacquier et al. 1997). This rapid divergence has been observed in the bindin gene of species of *Echinometra* sea urchins, which encodes a sperm protein required for species-specific gamete interaction (for review, see Vacquier et al. 1995; Vacquier 1998). A significantly high D_n:D_s (based on a one-tailed t test) was observed in a 39-codon subset of a 276-codon region of the bindin gene, leading to the conclusion that rapid interspecific divergence occurred in three recently speciated *Echinometra* congenes [$p < 0.05$ (Metz and Palumbi 1996)]. Similarly, closely related species of abalone exhibit high D_n:D_s values in the sequence of another sperm protein, lysin, which is required for sperm penetration through the egg's vitelline coat (Vacquier and Lee 1993; Lee et al. 1995). In several pairwise species comparisons, D_n:D_s values were significantly greater than unity ($p < 0.001$), again suggesting that selection is operating to create rapid sequence divergence between species at this locus.

In addition to these examples of interspecific diversity enhancing selection, there are at least three examples of this process occurring intraspecifically. Intraspecific diversity enhancing selection has been observed in the major histocompatibility complex (MHC) immune recognition molecules in mice (and the analogous HLA system in humans) (Hughes and Nei 1988, 1989), in the merozoite surface antigen in the malarial parasite *Plasmodium falciparum* (Hughes 1992), and in the self-incompatibility locus (the S-allele system) in plants (Clark and Kao 1991; reviewed by Nasrallah 1997). All exhibit high intraspecific allelic polymorphism as a result of diversity enhancing selection. For the MHC and malarial parasite systems, there appears to be an adaptive value to increase the ability to recognize a wider range of proteins (Hughes and Nei 1989; Hughes 1992). For the S-allele system, allelic diversity is coupled to self-incompatibility to promote outcrossing and reduction of inbreeding depression (Haring et al. 1990; Nasrallah 1997). In all three systems there is an adaptive advantage to heterozygosity and high intraspecific polymorphism at these recognition loci. Although diversity enhancing selection, as evidenced by intraspecific comparisons, has not been reported in a marine invertebrate gamete interaction molecule, sequence variation at the bindin locus among *Echinometra* species shows higher than expected levels of polymorphism (Metz and Palumbi 1996), suggesting that diversity enhancing selection could be operating.

To test the hypothesis that diversity enhancing selection may be creating sequence variation at the bindin locus in urchins, a large-scale, intraspecific examination of the sequence spanning a 273-nucleotide region of the red sea urchin (*Strongylocentrotus franciscanus*) bindin gene was conducted. This region at the 5′ end of the bindin locus encodes a portion of the variable region of the protein and corresponds to the "hotspot" of interspecific divergence observed in the *Echinometra* bindin sequences (Metz and Palumbi 1996). In addition, this 273-base pair (bp) region corresponds to ~37% of the entire open reading frame and ~42% of the variable region and exhibits both interspecific and intraspecific sequence variation in *Echinometra* spp. (Vacquier et al. 1995; Metz and Palumbi 1996). The analyses indicate that although this 5′ region of the *S. franciscanus* bindin locus examined does exhibit relatively high allelic polymorphism, neutral evolution (and not diversifying selection) appears to explain the observed allelic polymorphisms.

Materials and Methods

Sample Collection and Preparation

Sample collection and methods used to obtain DNA sequence data are identical to those reported by Debenham et al. (2000). Briefly, gonadal tissue was dissected from 134 adult *S. franciscanus* collected from six locations along the Pacific coast of North America from Alaska to Baja California, Mexico. DNA extractions were performed using a modified phenol–chloroform extraction as described by Milligan (1992) and De-

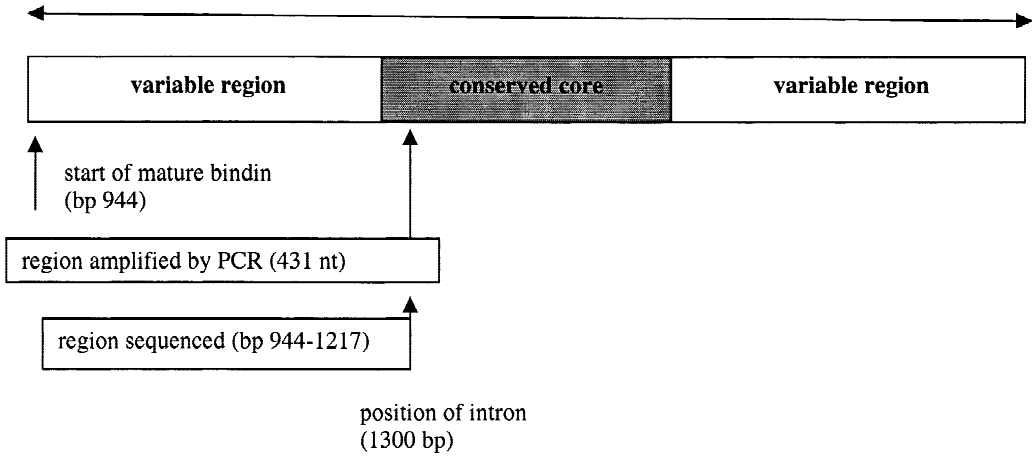

Fig. 1. Schematic diagram of the bindin gene and the regions analyzed by PCR amplification and sequence analysis. The numbering is according to that of Minor et al. (1991). The primers used in PCR amplification and in sequencing are given under Materials and Methods.

benham et al. (2000). It was imperative to conduct the extraction protocol on fresh ovary tissue. Ovary samples frozen for as little as 1 day showed a reduced yield of DNA and reduced success in subsequent PCR amplification.

Double-stranded PCR amplification of the bindin gene (Fig. 1) was followed by single-stranded asymmetric amplification to prepare a single-stranded template for sequencing. The primers used for amplification were FNbindin5′ (5′-AGTCGACGTTCGACAGACGAC-3′) and FNbindin3′ (5′-TTACATGGTCCATTATAGTATGCC-3′). The annealing temperature for the 5′ primer was 60°C and the final $MgCl_2$ concentration was 2 μM, using the primer at 0.5 μM. The PCR conditions used to amplify the 3′ single-stranded product required a lower annealing temperature (58°C) and lower final $MgCl_2$ concentration (1.2 μM). The concentration of the 3′ limiting primer was 2.5 μM (final). Some samples required additional adjustments to the annealing temperature (between 57 and 63°C) and limiting primer concentration (0.5–2.5 μM final concentrations). Internal sequencing primers were nested within the PCR product and were KTseq5′ (5′-GGAGCGC-GTAAHAAGCGTTAT-3′) and KTseq3′ (5′-ATACACACGATGGT-CAAG-3′). The Sanger dideoxy sequencing method was used (Sequenase version 2.0, Amersham).

To confirm each sequence, three separate PCR products (independent reactions) for each allele were cloned into the pBMKS vector and plasmid DNA from a minimum of four transformants was sequenced for each PCR product (Sambrook et al. 1989). PCR products derived from several individuals with the same genotype, but from different geographic locations, were cloned and sequenced to verify sequence consistency among alleles found at different geographic locations. The cloned products from four different PCR reactions were sequenced on both strands (Sambrook et al. 1989).

Sequence and Statistical Analyses

Sequences were aligned using the Seq-Ap version 1.9a multiple sequence alignment program for the Macintosh (Gilbert, 1994). Molecular Evolutionary Genetic Analysis version 1.01 [MEGA (Kumar et al. 1993)] was used to calculate nucleotide sequence divergence. Average proportions of replacement substitutions per replacement site (D_n) and silent substitutions per silent site (D_s) (Nei and Gojobori 1986), with standard errors (Nei and Jin 1989), were calculated using the MEGA program (Kumar et al. 1993). Corrections for multiple hits were not used because of the close relationship of the taxa (i.e., all data are alleles from the same species). PSWIN, a program available from T. Whittam (Pennsylvania State University), was used to conduct the sliding window D_n:D_s analysis on the sequence data. Tajima's *T* and Fu's *F* tests of neutrality were performed for all six distinct populations as well as the six populations combined (http://www.hgc.sph.uth.tmc.edu/fu/genealogy/test2/welcome.html). All amino acids encoded by each sequence were classified into categories of charge and hydrophobicity according to Lewin (1994; see Weir 1996). The cladistic analysis was performed by PAUP 3.1.1 (Swofford and Begle 1993). The search for the shortest tree was made by the exact branch-and-bound algorithm, which is guaranteed to find all optimal trees (data not shown).

Individual genotypes were coded as paired alphabetical characters and analyzed with BIOSYS (Swofford 1989) to obtain estimates of the following: allele frequencies, conformance to Hardy–Weinberg equilibrium, Wright's (1978) *F* statistics, and Nei's (1972) minimum genetic distance in pairwise comparisons. Conformance to Hardy–Weinberg proportions were estimated in three ways: (1) contingency chi-square analysis with Levene's (1949) correction for small sample size, (2) chi-square with pooling of rare and common categories, and (3) significance test with exact probabilities.

The *S. franciscanus* bindin polymorphic data were compared to the number of fixed differences in the bindin gene of four other species following the McDonald and Kreitman (1991) test for neutral evolution. cDNA sequence was available for the bindin locus of *S. purpuratus* [(Gao et al. 1986) GenBank accession number M14487], *S. droebachiensis* (C Biermann and W Eanes, State University of New York—Stony Brook, unpublished data; Bierman 1998), and *Lytechinus variegatus* [(Minor et al. 1991) GenBank accession number M59489]. Replacement and silent sites were counted as described by McDonald and Kreitman (1991) and a *G* test of independence was used with the Williams correction for continuity (McDonald and Kreitman 1991; Sokal and Rohlf 1981).

Comparisons with Simulated Models of Neutral Evolution

To assess if the amino acid substitutions observed in the 273-bp region of bindin were more common than would be expected by chance, a simulated data set was constructed that mimics the process of neutral evolution [DNA Evolve (Metz and Palumbi 1996); program available from SR Palumbi, Harvard University, at www.oeb.harvard.edu/cceg]. This new data set provided neutral expectations of random nucleotide substitutions. All data analyses were repeated on the simulated sequences to compare the results of observed bindin sequence variation to random (neutral) expectations. A more desirable approach to this analysis would be to preserve the genealogical relationship of the observed data set in the simulated data set. Unfortunately, results from our phy-

logenetic analysis show similar genealogical relationships among alleles, and further, the bootstrap values show that there is a high degree of uncertainty in the genealogy. The range (9–63) and the mean (35.5) of the bootstrap values are quite low (data not shown), indicating that there are many alternate trees possible. This uncertainty also is reflected in the polytomies of the sequence. Therefore, the most appropriate approach at this time is to eliminate all assumptions about allele genealogy in generating the simulated data set.

Results

Allelic Variation

DNA sequence analysis of a 273-bp region of the bindin locus (Fig. 1) from 134 individual *S. franciscanus* collected from Alaska to Baja California, Mexico, revealed the presence of 14 different alleles and 21 genotypes (Fig. 2). The four most common alleles exhibited frequencies of 0.51, 0.20, 0.16, and 0.08 (Fig. 2). These frequencies indicate that the bindin locus is polymorphic according to the criterion of Hartl (1988); i.e., polymorphism is indicated when the frequency of the most common allele of a gene is less than 0.95. There is a total of 13 variable positions among the 14 alleles; all other positions are identical (Fig. 2). The number of nucleotide differences between any two given sequences is relatively low, ranging between one (0.35%) and six (2.1%) substitutions in the 273-bp region analyzed. On average, there are 2.9 variable positions between any two alleles, resulting in average nucleotide sequence diversity of 1.06% [p-distance calculated using MEGA (Kumar et al. 1993)].

Assessment of Purifying Selection

Using published data, we examined the 5′, 91-codon subset of the 273-bp region of the bindin loci among three *Strongylocentrotus* congeners (*S. franciscanus, S. purpuratus,* and *S. droebachiensis*) and a more distantly related species, *Lytechinus variegatus.* The results indicate that although the nucleotide sequences vary to some degree, there are 25 amino acid positions (27%) that are identical among all four species, suggesting that the amino acid sequence is significantly constrained. This interspecific comparison, together with the observed polymorphism in *S. franciscanus* (Figs. 2 and 3), indicates that while interspecific conservation exists between species, the sequence of this region of the bindin locus is not entirely constrained by purifying selection.

Tests of Neutral Evolution and Diversity Enhancing Selection

Probability of Replacement Changes

It is possible for selection to act on a subset of positions in a gene. For example, the strong signal of positive selection seen in the human MHC was based on the 57 codons in the antigen recognition site (ARS) of the 398 total codons in the MHC molecule (Hughes and Nei 1988). We explored the possibility that selection is acting on the region of the bindin gene examined here. The polymorphic sites in *S. franciscanus* bindin appear to have a high proportion of replacement substitutions. Nine of the 13 polymorphic sites (69%) in the bindin region analyzed are replacement substitutions. If a nucleotide substitution results in an amino acid change, the new amino acid can have similar physiochemical properties (i.e., a conserved amino acid class change) or change in charge and/or polarity from the original amino acid (i.e., a radical change). Seven of nine (77%) replacement substitutions in the observed data set result in radical amino acid substitutions that change either charge or polarity. Radical amino acid substitutions are often considered to have a greater impact on protein function than conservative changes (Hughes et al. 1990) If this is so, then the different bindin alleles could have slightly different properties.

As a comparison to the observed bindin data, we calculated the number and type of amino acid changes in the DNA Evolve 2% sequences as a measure of the proportion of conserved and radical amino acid changes that would be expected to occur at random. The 10 DNA Evolve 2% data sets (140 sequences total) had on average 74% replacement substitutions, and 63% of these were radical amino acid changes. Those values are quite similar to the observed values of 69% replacement and 77% radical changes in our observed bindin data set. The mean values for the simulated data set for the proportion of replacement substitutions and the proportion of radical replacement substitutions are, respectively, within 1.19 and 0.61 standard deviations away from the observed values. This nonsignificant result suggests that the measured sequence variation does not deviate from the neutral expectation.

The analysis described above compared the proportions of radical and conservative replacement substitutions in the observed data set to those expected by chance over the entire 91 codon bindin region examined. Examination of only the nine codons that have replacement substitutions also indicates that the observed proportion of radical substitutions could have arisen by chance. The probability that the nine radical substitutions resulted in a new amino acid of a different class (charge or polarity) was calculated taking into account that each polymorphic site observed had only one variable nucleotide position. We calculated the probability that, given that a nucleotide is going to change and result in a new amino acid, this new amino acid will be a radical replacement change. For example, nucleotide position 85 is polymorphic (see Fig. 2). Some alleles contain the sequence GGG (Gly), while others are AGG (Arg). This is a radical replacement substitution (Gly is neutral hydrophobic

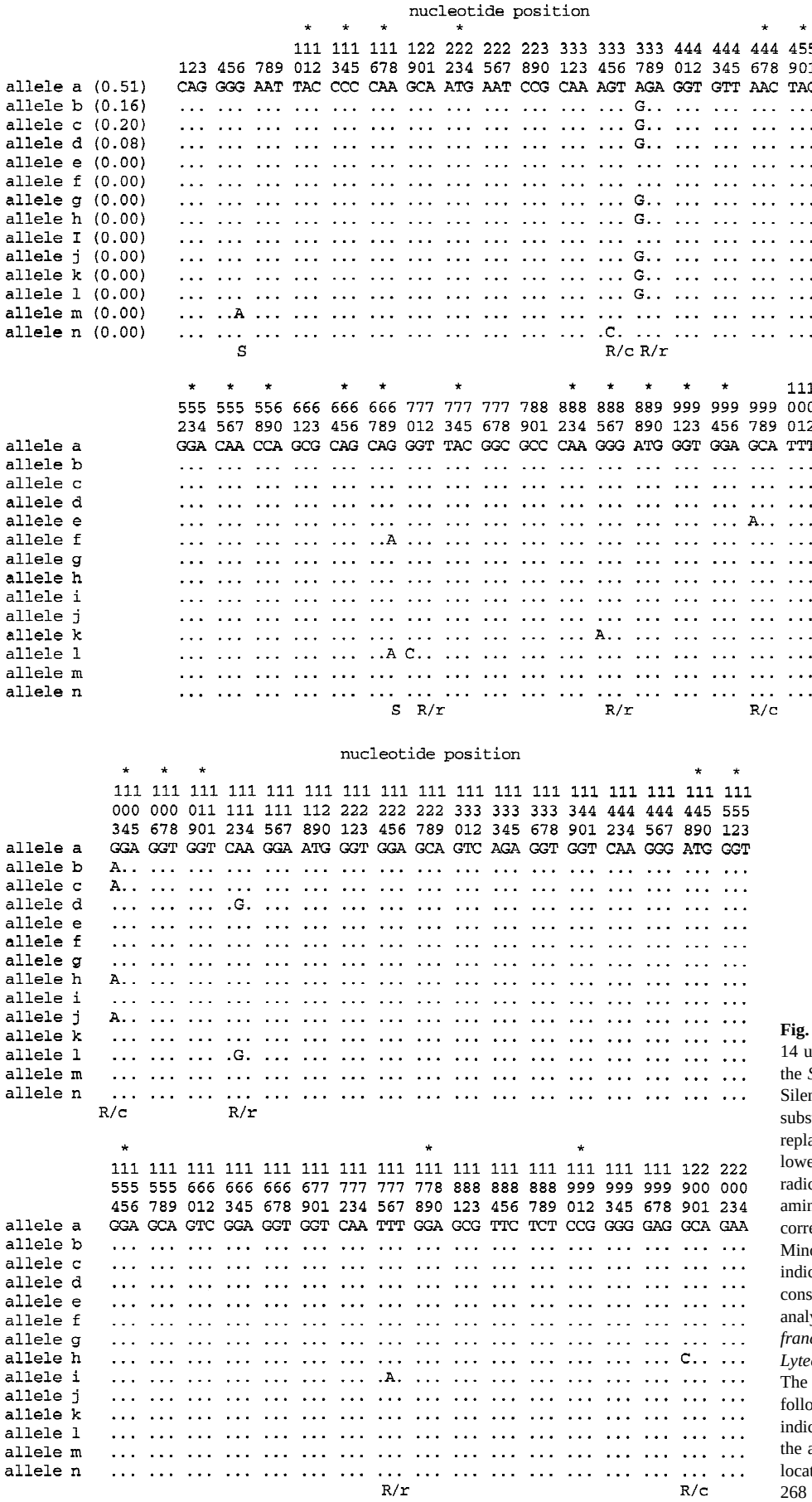

Fig. 2. Nucleotide sequence of the 14 unique alleles in the 5′ region of the *S. franciscanus* bindin gene. Silent (S) and replacement (R) substitutions are indicated. For replacement (R) substitutions, the lowercase letter denotes either a radical (r) or a conservative (c) amino acid change. Position 1 corresponds to position 944 of Minor et al. (1991). An *asterisk* indicates interspecific amino acid conservation in all four species analyzed (*S. purpuratus, S. franciscanus, S. droebachiensis,* and *Lytechinus variegatus;* see text). The *number in parentheses* following each allele (a through n) indicates the proportion in which the allele was found in all sampling locations (n = 134 individuals with 268 alleles).

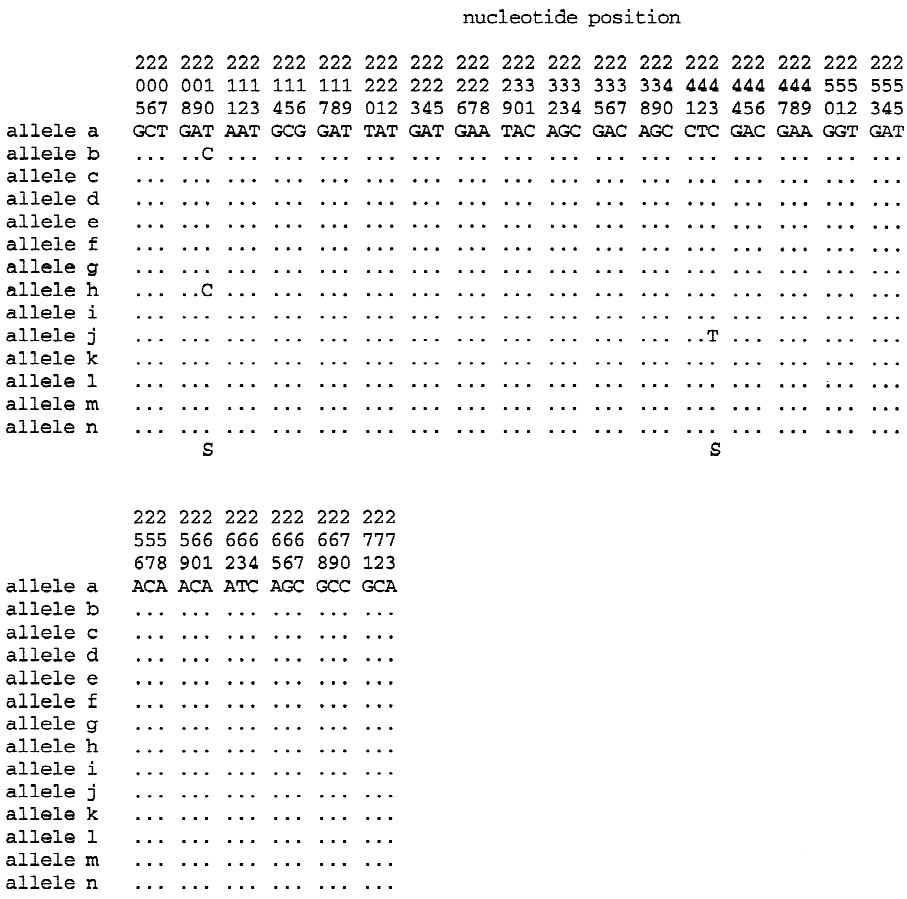

Fig. 2. Continued.

and Arg is basic). Yet any other nucleotide substitution (e.g., T or a C resulting in Trp and Arg, respectively) would also generate radical replacement substitutions. Therefore if nucleotide position 85 varies, the probability of a radical substitution at this codon is 1.0, and the probability of a conservative substitution is 0. We used sequence data of *S. purpuratus* bindin to determine the ancestral state of the sequence and thus the direction of the mutation in the *S. franciscanus* polymorphism. We then averaged the probabilities over the nine replacement polymorphic positions. In the DNA Evolve 2% bindin data set, the average probability of a replacement change being radical is 0.85, compared to the observed proportion of radical replacement substitutions of 0.77. The observed value is within 0.68 standard deviation from the calculated probability for an amino acid substitution being radical. Nucleotide substitutions at these polymorphic positions will result in a radical amino acid substitution over 80% of the time by chance. It is, however, important to remember that the power of the statistic is low because of the small number of polymorphic positions evaluated.

D_n:D_s Analysis

To evaluate the possibility of diversity-enhancing selection, the D_n:D_s values (Nei and Gojobori 1986) were calculated. For the 273-bp region of the *S. franciscanus* binding locus examined, the values of D_n:D_s for each geographic location range from 0.64 to 1.45, when evaluating just the unique bindin alleles in a specific population (Table 1). Similarly nonsignificant results were obtained for analyses of all alleles in each population (data not shown). Except for Alaska, all geographic locations have D_n:D_s values greater than 1.0. The overall D_n:D_s for the 14 unique alleles present in the species is 0.82. All measures of D_n:D_s are statistically indistinguishable from 1.0 ($p > 0.05$). A sliding window analysis (e.g., Metz and Palumbi 1996; Ina 1996) of a variety of window sizes (20, 25, 30, 40 codons) did not reveal any regions containing D_n:D_s ratios significantly different from 1.0. Analysis of the simulated DNA Evolve 2% data set displays similar results.

Tajima's *T* and Fu's *F* Tests

Table 1 includes the results from two additional tests for neutral evolution. Tajima's *T* test and Fu's *F* test of all alleles in each individual population, as well as the alleles from all the populations combined, do not reject the null hypothesis of neutral evolution. Similar nonsignificant results that support neutral evolution were obtained from Fu and Lis's *D*, Fu and Lis's *F*, Fu and Li's

```
          *** *        ** ** * ** *  ** ***  ***              ***
                1 1111111112 2222222223 3333333334 4444444445 5555555556
       1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
fran.  QGNYPQAMNP QSRGVNY-GQ PAQQGYGAQG MGGAFGGGQ- GMGGAVRGGQ GMGGAVGGGQ
purp.  TMG.....S. .MG..... .......... ...PV...P. ---------- -...P---P.
lyt.   NM....P..Q PMG.G..P.. .P..N.AP.. ...PV...G. -.....-.AG A...P....G
droe.  TMG..... S..MG..... .......... ........PV ......A... ..........

        *   *                                  1 1111111111 1111111111
       6666666667 7777777778 8888888889 9999999990 0000000001 1111111112
       1234567890 1234567890 1234567890 1234567890 1234567890 1234567890
fran.  FGAFSPGEAE ADNADYDEY- SDSLD----- -EGDTTISAA VMDDIKAVLG ATKIDLPVDI
purp.  ...LP..Q-- -ADT.FGSS. .S.V...... .G.......R .......... ..........
lyt.   G.MGG.VGGA NGIGESV.D. EM.V.SDYSS LG.E.....K .IQ...NL.. ..........
droe.  sequence currently not available in this region for S. droebach.

       **** ***** ********** ****** *** ********** **** *** *
       1111111111 1111111111 1111111111 1111111111 1111111111
       2222222223 3333333334 4444444445 5555555556 6666666667
       1234567890 1234567890 1234567890 1234567890 1234567890
fran.  NDPYDLGLLL RHLRHHSNLL ANIGDPAVRE QVLSAMQEEE EEEEEDAANG
purp.  .......... .......... .......... .......... ........T.
lyt.   ....Y..... .......... ......E... .......... ....N.....
droe.  .......... .......... .......... .......... ........T.
```

Fig. 3. Amino acid alignment of sea urchin bindin (after Vacquier et al. 1995), including a portion of *S. droebachiensis* bindin. The single-letter code is used. A *dot* signifies identity with the *S. franciscanus* sequence, while a *dash* indicates the absence of nucleotide at the position. *Asterisks* signify amino acid conservation across all four species. The position of the first amino acid corresponds to nucleotide 944 based on the numbering in Minor et al. (1991). fran., *S. franciscanus;* purp., *S. purpuratus;* droe., *S. droebachiensis;* lyt., *Lytechinus variegatus.*

Table 1. Tests for neutral evolution for the *S. franciscanus* 5′ bindin gene[a]

	D_n (±SE)	D_s (±SE)	D_n:D_s *(DNA Evolve Result)*	**Tajima's *T* test** *(5% significance value)*	**Fu's *F* test** *(5% significance value)*
AK	0.0091	0.0141	**0.82**	**−0.296**	**−1.330**
	(±0.0042)	(±0.0084)	*(0.80)*	(−1.39)	(−4.06)
WA	0.0068	0.0061	**1.11**	**1.107**	**0.549**
	(±0.0040)	(±0.0060)	*0.69*	(−1.40)	(−3.68)
OR	0.0081	0.0076	**1.06**	**1.148**	**1.719**
	(±0.0047)	(±0.0075)	*1.58*	(−1.36)	(−3.91)
No. CA	0.0083	0.0130	**0.63**	**−0.086**	**−0.604**
	(±0.0044)	(±0.0074)	*0.96*	(−1.41)	(−4.25)
So. CA	0.0083	0.0076	**1.09**	**0.595**	**0.750**
	(±0.0039)	(±0.0054)	*0.79*	(−1.40)	(−3.79)
Baja CA, Mexico	0.0097	0.0060	**1.61**	**0.694**	**−3.720**
	(±0.0049)	(±0.0060)	*0.95*	(1.40)	(−7.30)
Data for all sites				**0.036**	**−1.272**
				(−1.40)	(−4.53)

[a] Values for the proportion of nonsynonymous substitutions per nonsynonymous site (D_n) and the proportion of synonymous substitutions per synonymous site (D_s) calculated using MEGA (Kumar et al. 1993) as per Nei and Gojobori (1986). Values for D_n:D_s calculated based on pairwise comparisons of all unique alleles in each population. DNA Evolve Result (in parentheses) represents the D_n:D_s result for each population based on a randomly mutated data set (see text). Tajima's *T* test and Fu's *F* test results followed by critical value at 5% significance level in parentheses. Values for all tests are not significantly different from 1.0

*D**, and Fu and Li's *F** (data not shown). For these tests to be completely informative, it is important to distinguish between the derived and the ancestral alleles. Phylogenetic analysis of the sequences suggests many alternative relationships among the 14 unique alleles. Consequently, we present these results only as a side note.

McDonald–Kreitman Test for Neutral Evolution

The McDonald–Kreitman method is a more powerful test than the D_n:D_s method for assessing neutral evolution because purifying selection in portions of a molecule will not dampen the signal of directional selection (see debate by Li and Graur 1991; Whittam and Nei 1991; McDonald and Kreitman 1991). Therefore, the ratio of

Table 2. Summary of the number of replacement and synonymous substitutions for fixed differences between and polymorphisms within species for a 123-base pair region of the bindin gene[a]

	Fixed differences		Polymorphic differences		Ratio	
	Number of replacement substitutions	Number of silent substitutions	Number of replacement substitutions	Number of silent substitutions	Replacement:silent fixed substitutions	Replacement:silent polymorphisms
S. fran. vs *S. purp.*	14	6	8	1	2.3:1	8:1
S. fran. vs *S. droe.*	14	8	9	1	1.8:1	9:1
S. fran. vs *L. var.*	19	7	8	1	2.7:1	8:1

[a] cDNA sequence data for three species (*S. purpuratus, S. droebachiensis,* and *L. variegatus*) compared to intraspecific polymorphic data of *S. franciscanus* bindin. None of the relationships are statistically significant ($p > 0.05$)

replacement-to-synonymous fixed differences in the bindin sequence within species was compared to that between species. To enumerate fixed differences between species, bindin cDNA sequences from *S. purpuratus* (Gao et al. 1986), *S. droebachiensis* (Biermann 1998), and *Lytechinus variegatus* (Minor et al. 1991) were used (amino acid alignment presented in Fig. 3). for all four species, the McDonald–Kreitman test was used to evaluate selection in a 117-bp region of bindin, a region where alignment was unambiguous. This region initiates at the beginning of mature bindin and corresponds to base pairs 944 to 1061 in Minor et al. (1991) (see Fig. 1). These DNA sequence alignments represent at least one-third of the entire region examined within the *S. franciscanus* bindin gene.

In all comparisons, the ratio of replacement to silent differences in fixed versus polymorphic sites are not equal and differ by a factor of three to five. This suggests a proclivity for replacement polymorphic substitutions; however, none of the comparisons are statistically significant (see Table 2). In fact, for the comparison of the ratio of replacement-to-silent differences *S. franciscanus* to *S. purpuratus* to become significant, there would need to be an additional 6 (14 total) replacement polymorphisms (nearly twice as many as exist). In addition, analysis of simulated data produced by DNA Evolve also produced similar, nonsignificant results (data not shown).

Discussion

Our results indicate that the 273-bp region of the *S. franciscanus* bindin locus analyzed in this study is not subject to diversity enhancing selection. All tests are consistent with the conclusion that purifying selection constrains approximately 30% of the bindin sequence and that neutral evolution results in a small number of polymorphic positions. Under purifying selection, mutations resulting in amino acid substitutions occur periodically but are selected out of the population because of functional inferiority (Nei 1987). Silent substitutions that have no impact on fitness are ultimately fixed or lost through drift. Based on interspecific comparison of amino acid conservation in four urchin species, purifying selection seems to be constraining replacement substitutions in approximately 30% of the bindin region analyzed. Purifying selection has been shown to operate at other strongylocentrotid loci not involved in mating, such as the mitochondrial gene cytochrome oxidase I [CO1 (Edmands et al. 1996; Palumbi and Kessing 1991)].

The proportion of replacement-to-silent substitutions and the probability values for amino acid class changes both suggest that changes in the 5′ region of the *S. franciscanus* bindin gene appear to be the result of neutral evolution. Results from Tajima's *T* and Fu's *F* tests also support the existence of neutral evolution. All analyses performed on the DNA Evolve data set (created to mimic bindin variation resulting from neutral evolution) gave results nearly identical to those of the observed bindin data set. Although this simulated DNA Evolve data set is not ideal, it does indicate that random mutations will result in a high proportion of replacement changes and a majority of these changes will be radical amino acid substitutions. Many of those changes would be preserved under neutral evolution, giving rise to an apparent, although statistically insignificant, abundance of radical replacements. Yet neutral evolution would not necessarily predict the relatively high number of alleles observed in the bindin locus. Following a neutral mutation, new polymorphisms should either go to fixation or be removed from the population by genetic drift (Nei 1987). However, because population size can influence the action of drift (Ohta 1992), in large populations it will be a long time before the neutral polymorphisms are either fixed or lost through genetic drift. As very fecund broadcast spawners, urchins are known to have large populations (Morris et al. 1980). For *S. franciscanus,* in particular, the entire Pacific coast of North America appears to be one extended interbreeding population with no genetic subdivision (Debenham et al. 2000).

To date there is no example of diversity enhancing selection operating intraspecifically in a marine invertebrate gamete interaction molecule. Intraspecific sequence data within each *Echinometra* species (Metz and

Palumbi 1996) are similar to those observed for *S. franciscanus* bindin. For example, *E. mathaei* has 17 polymorphic sites in 252 bp, compared to 13 of 273 in *S. franciscanus.* Like *S. franciscanus,* a high proportion (76%) of these sites is replacement substitutions. The bindin sequences in both genera have comparable intraspecific D_n:D_s values that are not significantly different from 1.0. Finally, McDonald–Kreitman analysis on both genera revealed nonsignificant results that initially appear to have an excess of replacement polymorphisms (Metz and Palumbi, 1996; this work).

The major discrepancy between the two genera is that bindin sequences in *Echinometra* spp. have insertions and deletions (indels) both within and between species that range from 1 to 10 codons in length, while there are no indels in the *S. franciscanus* bindin sequence. The function, if any, of these indels is unknown, but it does suggest that, compared to *S. franciscanus, Echinometra* spp. bindin can tolerate relatively large alterations to sequence structure conceivably without impairment to function, a feature that would be consistent with the neutral theory of evolution.

Although diversity enhancing selection does not appear to be operating on the 5′ region of the bindin locus, the picture of neutral evolution seen here is still markedly different from that seen in most proteins. The coding regions of most genes are highly conserved such that only silent substitutions are observed within a species. Yet for bindin, the majority of substitutions is replacement substitutions. Neutral evolution present in a crucial gamete interaction molecule is contrary to expectations for a trait presumably closely related to fitness. The explanation for this contradiction is not clear.

If the bindin locus is evolving according to a neutral model, it would suggest that this 91-codon region of bindin is nonfunctional. For several reasons, it is difficult to accept that this region of the bindin protein has no necessary functional/structural role. The fact that 27% of this region is conserved between four species suggests that there is a functional need to maintain these amino acids. In addition, there is clearly a cost to maintaining this region of bindin; one example is the cost of potential reduction in an individual's fitness resulting from deleterious mutations that impair bindin's ability successfully to bind its cognate receptor on the egg surface (Palumbi 1994; Metz and Palumbi 1996). If there is a cost associated with the variation in this region of bindin, it is reasonable to assume that there is a concomitant benefit to maintain this region.

Other work also indicates that this region of the bindin gene product has an important function (Lopez et al. 1993; Palumbi 1999). Lopez et al. (1993) suggested that the variable 5′ (N-terminal) and/or 3′ (C-terminal) region is responsible for the observed species specific recognition properties of the bindin protein, as they determined that either the 5′ or the 3′ region was required for species-specific agglutination of urchin eggs. Palumbi (1999) examined the intraspecific functional role of bindin in *Echinometra mathaei.* Exposure of sea urchin eggs to experimental sperm mixtures showed that different females produce eggs that nonrandomly select sperm from different males with a bindin genotype similar to their own. The region of bindin that was characterized to determine genotype was the 5′ (N-terminal) region. The results appear to indicate that "a bindin allele has a functional advantage only in the context of a particular female genotype. Because different bindin alleles function best with different mating partners, fertility variation within populations is not necessarily eliminated by rapid selective sweeps" (Palumbi 1999).

Unusually fast rates of amino acid substitutions, indicative of positive natural selection, would predict very low polymorphism within a species as all newly adapted alleles become fixed. The moderate level of polymorphism observed here in *S. franciscanus* does not support the existence of a recent selective sweep. However, based on Palumibi's work (1999) it is possible that the strong linkage between female choice and male trait loci maintains the observed, yet unexpected, polymorphism in *S. franciscanus.*

Bindin evolution presumably is coupled with that of a cognate egg binding protein. Any changes in bindin sequence must be tolerated by a protein(s) on the egg surface that interacts with it (Orr 1990; Vacquier 1998). This interaction is dynamic, with both bindin and its cognate coevolving. Therefore, these polymorphic positions represent the few sites where variation is tolerated in both sperm bindin and the egg cognate. Perhaps these positions could be viewed as highly informative regarding the nonfunctional regions of bindin or perhaps there is an even more interesting, yet currently not understood mechanism operating to maintain polymorphisms in a gamete interaction molecule.

Although it seems unlikely, it is possible that simultaneous action of diversity enhancing and purifying selection could combine to appear as neutral evolution. Regardless of the mode of evolution, it is clear that bindin is a divergent protein that functions in gamete interactions (Vacquier et al. 1995; Vacquier 1998). Further, it is important to note that other as yet unexamined regions of the bindin locus may be important indicators of evolutionary mechanisms.

Acknowledgments. Thanks go to S. Palumbi and E. Metz, who provided valuable feedback to many versions of the manuscript. Thanks are also due to R. Jovine, K. Thomas, E. DeLong, L. Wooninck, B. Phillips, S. Poole, D. Mederios-Bergen, and S. Karl for technical advice; J. Holcomb, S. Anderson, F. DeWitt, A. Rodriquez, D. Woodby, R. Kraxberger, B. Marthaler, D. Kushner, K. Duran, M. Ellis, and T. Minicucci for assistance in sample collection; and Y. Abassi, S. Adams, S. Aiken, R. Feldan, A. Giusti, students at Fort Bragg High School, B. Hinkle, K. Hoang, M. Kelley, L. O'Bryan, and R. Shipe for assistance with sample preparation. This work was supported in part by a Saltonstall–Kennedy grant from the National Oceanic and Atmospheric Administration to M. Brzezinski and a National Science Foundation grant to K. Foltz.

References

Biermann CH (1998) The molecular evolution of sperm binding in six species of sea urchins (Echinoida: Strongylocentrotidae). Mol Biol Evol 15:1761–1771

Clark AG, Kao T-H (1991) Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of *Solanaceae.* Proc Natl Acad Sci USA 88:9823–9827

Clark AG, Aguadé M, Prout T, Harshman LG, Langley CH (1995) Variation in sperm displacement and its association with accessory gland. Genetics 139:189–201

Debenham P, Foltz KR, Brzezinski MA, Gaines SR (2000) Population genetics of the red sea urchin, *Strongylocentrotus franciscanus,* along the west coast of North America. JEMBE (in press)

Edmands S, Moberg PE, Burton RS (1996) Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus.* Mar Biol 126:443–450

Endo T, Ikeo K, Gojobori T (1996) Large-scale search for genes on which positive selection may operate. Mol Biol Evol 13:685–690

Gao B, Klein LE, Britten RJ, Davidson EH (1986) Sequence of mRNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization. Proc Natl Acad Sci USA 83:8634–8638

Gilbert D (1994) SeqApp. Multiple sequence alignment program. Indiana State University, Evansville

Haring V, Gray JE, McClure BA, Anderson MA, Clarke AE (1990) Self-incompatibility: A self-recognition system in plants. Science 250:937–941

Hartl DL (1988) A primer of population genetics, 2nd ed. Sinauer Associates, Sunderland, MA

Hughes AL (1992) Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum.* Mol Biol Evol 9:381–393

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–168

Hughes AL, Nei M (1989) Major histocompatibility complex class II loci: Evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of Class 1 major-histocompatibility-complex molecules. Mol Biol Evol 7: 515–524

Ina Y (1996) Pattern of synonymous and nonsynonymous substitutions: An indicator of mechanisms of molecular evolution. J Genet 75: 91–115

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis. Pennsylvania State University, University Park

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

Levene H (1949) On a matching problem arising in genetics. Ann Math Stat 20:91–94

Lewin B (1994) Genes V. Wiley, New York

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA

Lopez A, Miraglia SJ, Glabe CG (1993) Structure/function analysis of the sea urchin sperm adhesive protein bindin. Dev Biol 156:24–33

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila.* Nature 351:652–654

Metz EC, Palumbi SR (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol Biol Evol 13:397–406

Milligan BG (1992) Plant DNA Isolation. In: Hoelzel AR (ed) Molecular genetic analysis of populations. IRL Press, Oxford, pp 71–74

Minor JE, Fromson DR, Britten RJ, Davidson EH (1991) Comparison of the bindin proteins of *Strongylocentrotus franciscanus, S. purpuratus,* and *Lytechinus variegatus:* Sequences involved in the species specificity of fertilization. Mol Biol Evol 8:781–795

Morris RH, Abbott DP, Haderlie EC (1980) Intertidal invertebrates of California. Stanford University Press, Stanford, CA

Nasrallah JB (1997) Evolution of the *Brassica* self-incompatibility locus: A look into S-locus gene polymorphisms. Proc Natl Acad Sci USA 94:9516–9519

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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

Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6: 290–300

Ohta T (1992) The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst 23:263–286

Orr HA (1990) Is single-gene speciation possible. Evolution 45:764–769

Palumbi SR (1992) Marine speciation on a small planet. Trends Ecol Evol 7:114–117

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25:547–572

Palumbi SR (1995) Using genetics as an indirect estimator of larval dispersal. In: McEdward LR (ed) The ecology of marine invertebrate larvae. CRC Press, Boca Raton, FL, pp 369–387

Palumbi SR (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in se urchins Proc Natl Acad Sci USA 96:12632–12637

Palumbi SR, Kessing BD (1991) Population biology of the trans-Arctic exchange: mtDNA sequence similarity between Pacific and Atlantic sea urchins. Evolution 45:1790–1805

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY

Sokal RR, Rohlf FJ (1981) Biometry: The principles and practice of statistics in biological research, 2nd ed. W.H. Freeman, New York

Strathmann MF (1987) Reproduction and development of marine invertebrates of North Pacific Coast: Data and methods for the study of eggs, embryos, and larvae. University of Washington Press, Seattle, p 657

Swofford DL (1989) BIOSYS-1: A computer program for the analysis of allelic variation in population genetics and biochemical systematics. University of Illinois at Urbana, Champaign

Swofford DL, Begle DP (1993) PAUP: Phylogenetic analysis using parsimony. Laboratory of Molecular Systematics, Smithsonian Institution, Washington, DC

Vacquier VD (1998) Evolution of gamete recognition proteins. Science 281:1995–1998

Vacquier VD, Lee Y-H (1993) Abalone sperm lysin: Unusual mode of evolution of a gamete recognition protein. Zygote 1:181–196

Vacquier VD, Swanson WJ, Hellberg ME (1995) What have we learned about sea urchin sperm bindin? Dev Growth Diff 37:1–10

Vacquier VD, Swanson WJ, Lee Y-H (1997) Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure for their divergence? J Mol Evol 44 (Suppl 1):S15–S22

Ward CR, Kopf GS (1993) Molecular events mediating sperm activation. Dev Biol 158:9–34

Weir BS (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Associates, Sunderland, MA

Whittam TS, Nei M (1991) Neutral mutation hypothesis test. Nature 354:115–116

Wright S (1978) Evolution and the genetics of populations, Vol 4. Variability within and among natural populations. University of Chicago Press, Chicago