Positive Selection and Sequence Rearrangements Generate Extensive Polymorphism in the Gamete Recognition Protein Bindin

Edward C. Metz and Stephen R. Palumbi

Department of Zoology and Kewalo Marine Laboratory, University of Hawaii

Bindin is a gamete recognition protein of sea urchins that mediates species-specific attachment of sperm to an egg-surface receptor during fertilization. Sequences of bindin from closely related urchins show fixed species-specific differences. Within species, highly polymorphic bindin alleles result from point substitution, insertion/deletion, and recombination. Since speciation, positive selection favoring allelic variants has generated diversity in bindin polypeptides. Intraspecific bindin variation can be tolerated by the egg receptor, which suggests functional parallels between this system and other flexible recognition systems, including immune recognition. These results show that polymorphism in mate recognition loci required for rapid evolution of sexual isolation can arise within natural populations.

Downloaded from https://academic.oup.com/mbe/article/13/2/397/993332 by guest on 22 August 2025

Introduction

Intercellular interactions such as fertilization and immune responses are mediated by protein–protein binding interactions. Depending on the recognition function of these proteins, coevolutionary changes may accumulate as a result of positive selection. Evidence of positive selection has been documented in over a dozen types of recognition proteins (e.g., Hughes and Nei 1988; Clark and Kao 1991; Lee and Vacquier 1992; Swanson and Vacquier 1995). Selection can act on the recognition properties of new variant alleles that arise by a variety of sequence changes, including amino acid replacement mutations and recombinations among existing alleles. Here we present sequence analysis of a gamete recognition locus, at which highly polymorphic alleles showing evidence of positive selection have a common ancestor near the time of speciation.

Genes controlling mate recognition may differentiate between species, leading to sexual reproductive isolation (Coyne 1992; Palumbi 1994). Theoretical studies have shown that when mate recognition loci are highly polymorphic, assortative mating may evolve rapidly (Nei, Maruyama, and Wu 1983; Wu 1985). Heritable variation in male traits and female preferences (reviewed by Pomiankowski and Sheridan 1994) implies that polymorphisms exist at the corresponding genetic loci. Such intraspecific alleleic variation would be required for differentiation of novel assortative mating types. However, because of the complexity of both signaling and receiving processes during courtship, there is little direct evidence of sequence variation at loci responsible for reproductive isolation (Coyne 1992; Wu et al. 1995).

Key words: bindin, *Echinometra,* polymorphism, recombination, positive selection, mate recognition, fertilization, sea urchin.

Address for correspondence and reprints: Edward C. Metz, Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093.

Mol. Biol. Evol. 13(2):397–406. 1996
© 1996 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Species-specific fertilization is maintained by a relatively small number of recognition proteins expressed on gamete surfaces. This provides opportunities for study of the molecular evolution and population genetics of reproductive isolation (Palumbi 1994). In free-spawning marine invertebrates, gamete recognition proteins that have been characterized include sperm bindin and its egg receptor from sea urchins (Vacquier and Moy 1977; Minor, Gao, and Davidson 1989; Foltz and Lennarz 1993; Vacquier, Swanson, and Hellberg 1995) and sperm lysin from abalone (Lee and Vacquier 1992; Lee, Ota, and Vacquier 1995). In sea urchins, species-specificity of fertilization generally occurs during sperm–egg attachment (Summers and Hylander 1975; Metz et al. 1994), which is mediated by the interaction of sperm acrosomal protein bindin and its egg-surface receptor (Minor, Gao, and Davidson 1989; Foltz and Lennarz 1993; Vacquier, Swanson, and Hellberg 1995).

In order to examine patterns of molecular evolution at a mate recognition locus, we sequenced the mature bindin coding region from three Indo–Pacific sea urchin species, *Echinometra mathaei*, *Echinometra oblonga*, and *Echinometra* species 'type A' (Metz et al. 1994). Reproductive isolation between these sympatric species is maintained by a pronounced species-specificity of fertilization. Gamete incompatibilities between the species occur at steps of fertilization known to be mediated by the bindin–receptor interaction (Metz et al. 1994). Based on mitochondrial and nuclear DNA comparisons, these urchins diverged in the Pleistocene and are the most closely related urchins for which genetic relationships are known (Palumbi and Metz 1991).

Materials and Methods

Echinometra mathaei and *E. oblonga* were collected from mixed populations at Kapapa Island in Kaneohe Bay, Oahu, Hawaii. Samples from other locations (Bali,

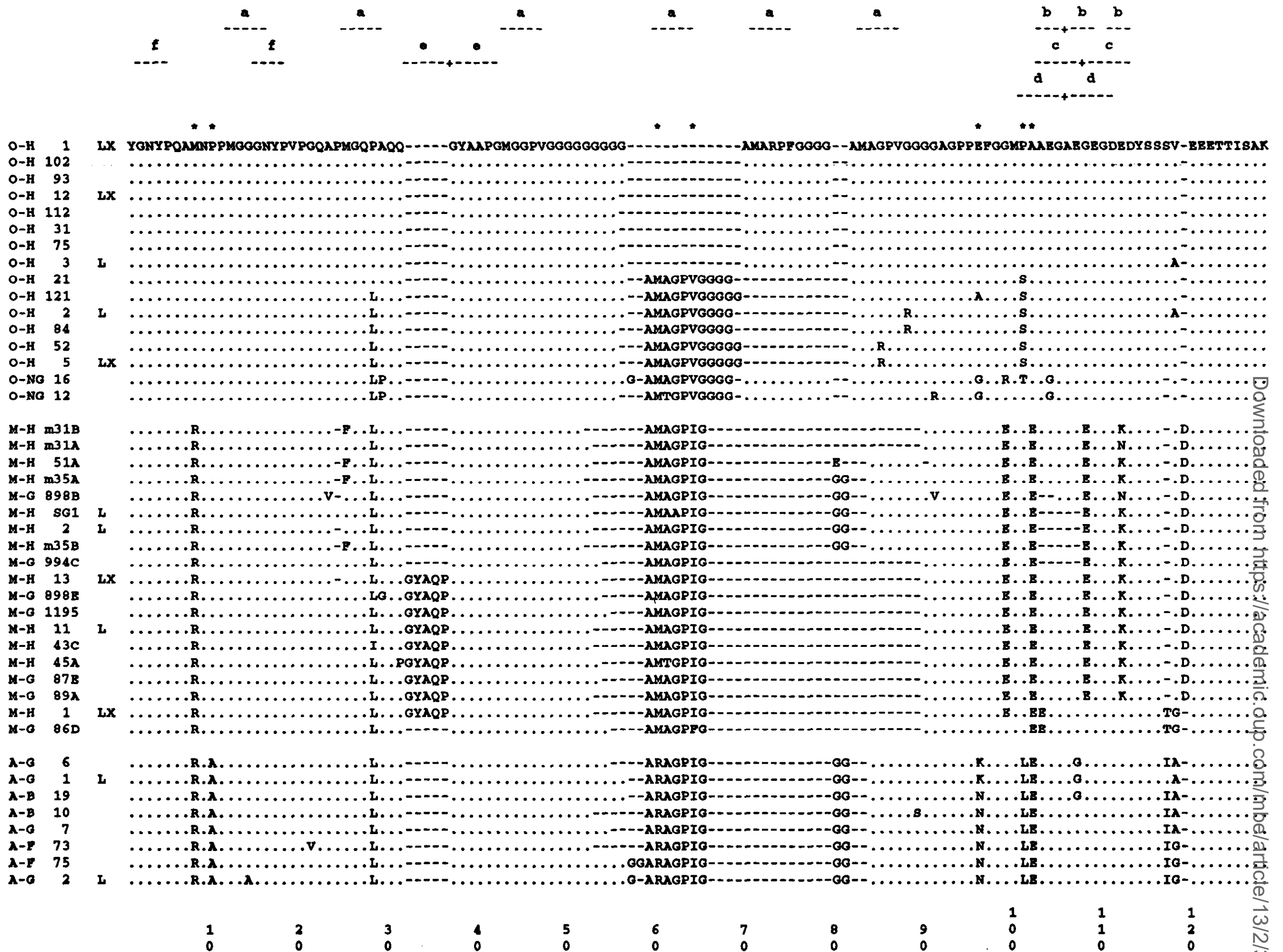

FIG. 1*A*.—Deduced amino acid sequences of mature *Echinometra* bindin showing species-specific and polymorphic variation. Alignment is based on nucleotide sequences, including information at silent sites. Labels denote species (M: *Echinometra mathaei*; O: *E. oblonga*; A: *Echinometra* species type 'A') obtained from Hawaii (H), Guam (G), Fiji (F), Bali (B), and New Guinea (NG). Label numbers are the original sample identification. Sequences derived from mRNA are indicated by 'X'. Dashes indicate alignment gaps. Asterisks indicate alignment positions at which all alleles of one species differ from all alleles of the other two species. Positions of repeated element sets a–f are indicated by dashed overlines. Aligned codon numbering is shown below the sequences. *A*: 5′ bindin sequences showing aligned codons 1–128 up to the position of the intron. *B*: 3′ bindin sequences showing aligned codons 129–277 downstream of the intron. The 5′ portions of these full-length sequences are indicated by 'L' in figure 1*A*. The overline indicates the region of highest sequence conservation among different urchin genera (see Vacquier, Swanson, and Hellberg 1995, fig. 5, alignment positions 365–428). Note: Neighboring polyglycine regions that are not aligned (e.g., positions 76–81 and 250–255) reflect silent differences.

Downloaded from https://academic.oup.com/mbe/article/13/2/397/1083306 by guest on 11 August 2022

Fiji, Guam, and New Guinea) of these two species and *Echinometra* species 'type A' (Metz et al. 1994) were preserved in 70% ethanol prior to shipping.

PCR Templates

Three different sources of templates were used in standard 50-μl PCR amplifications of *Echinometra* bindin sequences: (1) Testis cDNA: Male urchins were injected with 0.5 M KCl in order to induce spawning and clear sperm from the testis. Poly-A RNA (mRNA) was isolated from total testis RNA on oligo-dT cellulose (Sambrook, Fritsch, and Maniatis 1989). AMV reverse transcriptase was used to make cDNA from 1–2 μg of mRNA by random hexamer priming. One microliter of the cDNA reaction was then used directly in PCR reactions with the 3′ "RACE" primer system for poly-A RNA (Frohman, Dush, and Martin 1988). (2) Genomic DNA isolated from sperm. (3) Whole sperm: One microliter of a 1:2,000 dilution of sperm in distilled water was used directly in PCR reactions.

PCR Primers

PCR primers were designed to bracket and amplify the entire region encoding the mature functional portion of *Echinometra* bindin. Conserved regions in the bindin sequences of two *Strongylocentrotus* species (Gao et al.

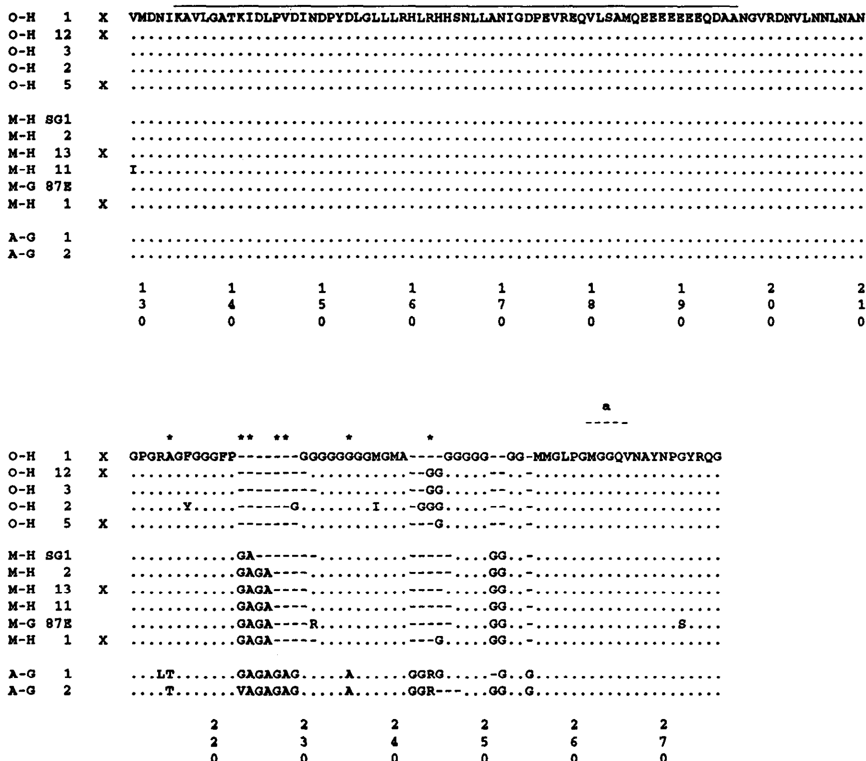

FIG. 1B

1986; Minor et al. 1991) allowed primers to be made that amplify 5′ and 3′ regions flanking the mature *Echinometra* bindin coding sequence. Sequences of these flanking regions were used to identify a pair of *Echinometra*-specific primers (PBE: 5′ TCGGAAAGAGGC-GGACGTCAGAAACG 3′ and E6: 5′ GTWTACCTA-GTTGGTATCTCTTTA 3′) bracketing the mature protein coding sequence. The upstream primer (PBE) is located in the region of the proteolytic cleavage site that releases the mature bindin from the "preprobindin" (Gao et al. 1986; Minor et al. 1991). The downstream primer (E6) is located outside of the bindin stop codon in the 3′ untranslated region. An alternate downstream primer (EJ: 5′ CATTAGGCCTACCTTTGGCAC 3′) was used with PBE to amplify only the 5′ half of the bindin coding sequence.

Cloning and Sequencing

PCR products were cloned in T-vectors as described (Palumbi and Baker 1994) in order to separate individual alleles. Bacterial colonies were used directly as template in PCR with vector primers to screen for the presence of plasmids with inserts of the correct size. Single-stranded plasmid DNA was obtained by helper phage infection (VCSM13, Stratagene) and sequenced using a series of primers. To assess polymerase error, multiple clones were sequenced from two or three independent replicate PCR reactions using sperm DNA from each of four individual urchins as templates. Sequence comparisons of clones from different PCR reactions using the same template showed an average of 0.85 difference per full-length bindin coding sequence due to polymerase error. No evidence of insertion/deletion or recombination was found within any of the four sets of replicate sequences.

Sequence Analysis

Phylogenetic analysis was performed using PAUP (Swofford 1993). Alignment gaps (excluding polyglycines) were coded as presence/absence characters. Average proportions of replacement substitutions per replacement site (P_n) and silent substitutions per silent site (P_s) (Nei and Gojobori 1986), with standard errors (Nei and Jin 1989), were calculated using the MEGA program (Kumar, Tamura, and Nei 1993). Corrections for multiple hits were not used because of the close relationship of the taxa. MEGA was used to obtain nucleotide frequency and codon usage in the sequences. From these, measures of nucleotide bias (Irwin et al. 1991) and codon usage bias (Shields et al. 1988) were calculated. MEGA was also used to obtain neighbor-joining phylogenetic reconstructions. The GCG program (Genetics Computer Group 1991) was used to identify re-

Downloaded from https://academic.oup.com/mbe/article/14/4/398/... by guest on 21 August 2022

peat and inverted repeat regions of nucleotide sequences.

Non-conservative amino acid substitutions were identified as changes between amino acid functional classes based on net charge and polarity as described (Hughes, Ota, and Nei 1990); likewise, conservative substitutions were changes within classes. To generate expected values for non-conservative changes under neutral evolution, we randomly substituted bindin sequences at 6% of nucleotide positions. (Six percent reflects the average difference between alleles from different species between aligned codons 80 and 120.) For each randomly substituted sequence, we counted the number of conservative and non-conservative differences from the original sequence. This procedure was repeated 100 times for an allele from each urchin species. The proportion of non-conservative amino acid differences in these 300 sequences provided the neutral expectation. Using the chi-square test, we compared this neutral expectation with the proportion of non-conservative amino acid changes indicated by our phylogenetic reconstruction to have occurred in the bindin alleles.

Results

Bindin Sequences

Full-length mature bindin sequences were obtained for five alleles from cDNA and eight alleles from genomic DNA (fig. 1). Like bindin sequences from other sea urchin genera (Gao et al. 1986; Glabe and Clark 1991; Minor et al. 1991), *Echinometra* cDNA sequences have an inferred polypeptide length of about 240 amino acids, a central highly conserved region of about 80 amino acids, and short repeated elements. An intron is located at aligned codon position 128 (fig. 1). Introns range in length from about 450 to 750 bp, depending on the species. Most variation was found upstream of the intron, so we sequenced 30 additional alleles in this 5′ region (fig. 1*A*).

Bindin sequences from *Echinometra* as well as other genera of urchins (Glabe and Clark 1991; Minor et al. 1991; Lopez, Miraglia, and Glabe 1993; Vacquier, Swanson, and Hellberg 1995) show numerous short repeated sequence motifs and chains of identical codons for glycine. We identified six sequence motifs (excluding polyglycines) that have two to seven repeats at least 40% similar in nucleotide sequence (fig. 1). One of these (set 'a', fig. 1) resembles the "MGGXX" repeat unit found in bindins from other urchin genera (Minor et al. 1991; Lopez, Miraglia, and Glabe 1993). Nucleotide sequences of these units (set 'a') could form 9–12-bp hairpin stems with neighboring repeats of the same type (not shown).

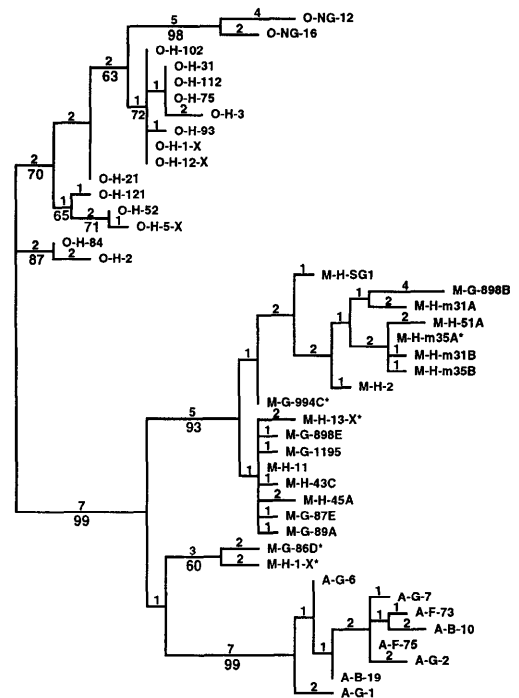

FIG. 2.—Phylogenetic reconstruction of 5′ bindin coding sequence relationships. In addition to nucleotide characters, alignment gaps (excluding polyglycines) were coded as presence/absence characters. Labels are as in figure 1. The consensus tree (PAUP 3.1.1 [Swofford 1993], equal weighting, 1,000 random order additions) has length (102) close to the minimum length trees (101) and a consistency index of 0.794. Numbers above branches indicate branchlengths. Percentage of occurrence in 500 bootstrap searches is indicated below branches (for values above 65%). Asterisks denote unique alleles that may have arisen by recombination among the other alleles (see fig. 4); hence the history of alleles within species is not likely to be described accurately by this reconstruction.

In the full-length *Echinometra* bindin sequences, alignment gaps indicate length variation of polyglycine codons typical of microsatellites, and 11 sites at which 1–10 codons are inserted or deleted (fig. 1). Insertion/deletion of one or a few codons probably results from slipped-strand mispairing (Levinson and Gutman 1987). Several repeated units are inserted/deleted in different alleles, and some of these insertion/deletions are bracketed by polyglycines. Alleles that are otherwise identical can differ in the number of polyglycine codons (e.g., alleles M-H 11 and M-G 1195, fig. 1*A*). There is no evidence of insertion/deletion in the highly conserved central region of bindin.

PAUP reconstructions based on coding sequences of the 13 full-length alleles showed bindin alleles to be monophyletic within species (100% bootstrap support in 200 searches). The 43 5′ sequences also grouped according to species with high bootstrap values (fig. 2) with the exception of two apparently paraphyletic *E. mathaei* alleles (see below).

Downloaded from https://academic.oup.com/mbe/article/13/…

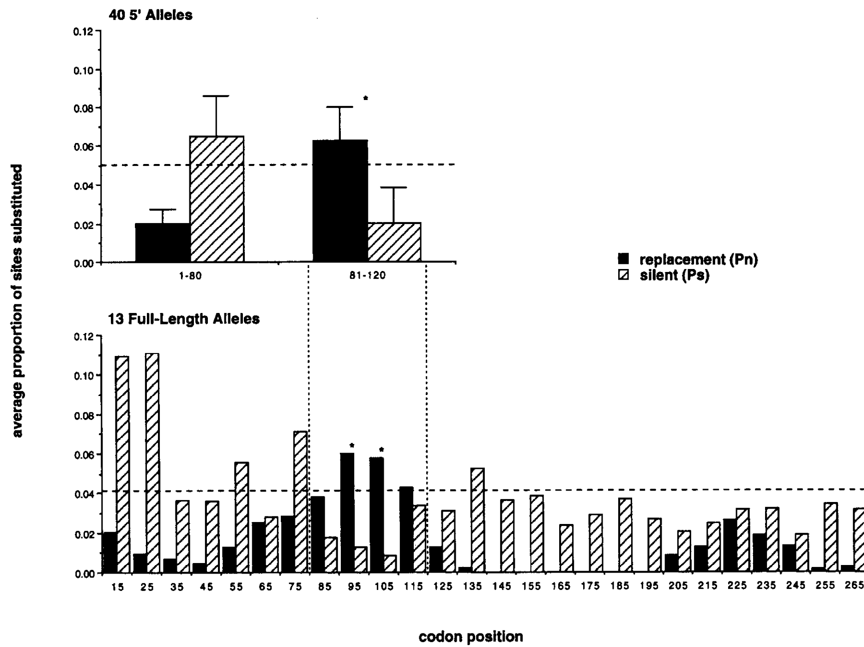

FIG. 3.—Comparison of proportions of replacement substitutions per replacement site (P_n) and silent substitutions per silent site (P_s) along the *Echinometra* bindin coding sequences. Average P_n and P_s with standard errors were calculated using the MEGA program (Kumar, Tamura, and Nei 1993). For all comparisons of 13 full-length bindin alleles, average P_n and P_s are shown in overlapping 30-codon windows, each 10 codons apart (*below*). The profile reveals a region of high replacement between codons 80 and 120. For all comparisons of 40 unique 5′ alleles, average P_n and P_s are shown for the high replacement region between codon positions 80 and 120 and for the remainder of the sequence (*above*). Horizontal dashed lines indicate the overall average P_s for each sequence set. Asterisks indicate comparisons in which P_n is significantly greater than P_s ($P < 0.05$, Z-test; Kumar, Tamura, and Nei 1993).

Amino Acid Replacements

A total of 14 amino acid alignment positions show species-specific differences in our sample of full-length and 5′ alleles (fig. 1*A* and *B*). Of these, 10 are amino acid replacements and seven of the replacements change charge or polarity. Among 10 differences fixed between species in the 5′ nucleotide sequences, seven are replacements and three are silent. Likewise, for changes that are polymorphic within species in this region, 32 are replacements and 19 are silent. Because the polymorphic replacement/silent ratio does not differ from the fixed replacement/silent ratio, *Echinometra* 5′ bindin sequence data do not reveal adaptive evolution in the McDonald-Kreitman test ($P > 0.5$, G-test; McDonald and Kreitman 1991). However, the usual alternative of neutral evolution is probably not indicated because different regions of bindin show evidence of both purifying selection and positive selection.

We compared the proportion of replacement substitutions per replacement site (P_n) to the proportion of silent substitutions per silent site (P_s) in 30-codon overlapping windows along the 13 full-length bindin sequences (fig. 3). Between aligned codon positions 130 and 213, all alleles have the same inferred amino acid sequence (fig. 1*B*), so in this region $P_n = 0$, indicating the action of purifying selection (Hughes and Nei 1988). In windows between alignment codons 80 and 120, average P_n exceeds average P_s (and is greater than average P_s over the entire sequence). Comparing all 40 unique 5′ bindin sequences between codons 80 and 120 showed that $P_n/P_s = 3.0$ (fig. 3 and table 1). P_n is also greater than P_s in comparisons of alleles *within* species (table 1). When $P_n/P_s > 1$, the action of positive Darwinian selection is indicated (Hughes and Nei 1988; Hughes, Ota, and Nei 1990; Clark and Kao 1991; Lee and Vac-

Table 1
Mean P_n ± SE and mean P_s ± SE in Comparisons among Bindin Alleles

	P_n	P_s	P_n/P_s
Codons 1–80			
40 unique 5′ alleles	0.020 ± 0.007	0.065 ± 0.021	0.31
13 *E. oblonga* alleles	0.008 ± 0.005	0.049 ± 0.019	0.16
19 *E. mathaei* alleles	0.008 ± 0.003	0.030 ± 0.013	0.27
8 *E.* type 'A' alleles	0.004 ± 0.003	0.011 ± 0.008	0.28
Codons 81–120			
40 unique 5′ alleles	0.063 ± 0.017	0.021 ± 0.018	3.00*
13 *E. oblonga* alleles	0.028 ± 0.010	0.017 ± 0.012	1.65
19 *E. mathaei* alleles	0.018 ± 0.008	0	>1.0
8 *E.* type 'A' alleles	0.018 ± 0.009	0.016 ± 0.015	1.13

NOTE.—Asterisk indicates comparison in which mean P_n is significantly greater than mean P_s (Z-test, $P < 0.05$; Kumar, Tamura, and Nei 1993).

Downloaded from https://academic.oup.com/mbe/article/13/2/396/1033262 by guest on 24 August 2022

quier 1992; Clark 1993; Hughes 1993; Lee, Ota, and Vacquier 1995).

Positive selection may favor an excess of nonconservative substitutions that change the charge or polarity of amino acid residues (Hughes, Ota, and Nei 1990). To test for this in bindin, we compared the proportion of nonconservative changes observed in the high replacement region (codons 80–120) with the proportion of nonconservative changes that would occur by random substitution in this region. The average proportions of nonconservative changes expected to occur randomly are 43% for charge changes and 49% for polarity changes (see Methods). Phylogenetic reconstruction (fig. 2) indicates that 27 amino acid replacements occurred between aligned codons 80 and 120. (In four cases, the same change shows on the tree more than once within species; these changes may result from recombination [see below] and were counted as a single change at each site.) Of the 27 observed replacements, 17 (63%) change charge, compared with 11.6 (43%) expected at random ($P = 0.036$, Chi square). Likewise, 19 (70%) of the observed replacements change polarity, compared with 13.2 (49%) expected at random ($P = 0.026$, Chi square). This excess of non-conservative substitution further suggests that positive selection favors the diversification of bindin.

Explanations for High Replacement/Silent Ratios

In general, a high P_n/P_s ratio indicates positive selection, but it might also result from constraints on silent substitution. We tested for such constraints in two ways. First we examined nucleotide and codon usage bias. Average G+C content among bindin sequences was 37.6% for introns, 51.6% for third codon positions, and 59.2% for all codon positions. Nucleotide frequency bias (Irwin, Kocher, and Wilson 1991), measured on a scale of 0 (no bias) to 1 (maximum bias), was 0.165 for introns, 0.064 for third codon positions, and 0.147 for all codon positions. The scaled Chi square (Shields et al. 1988), indicating codon usage bias on a scale of 0 (no bias) to 1 (maximum bias), was 0.238 for all bindin codons and 0.247 for all fourfold degenerate codons. These measures indicate that nucleotide frequency bias and codon usage bias are both low to slight (Shields et al. 1988; Irwin, Kocher, and Wilson 1991; Lee, Ota, and Vacquier 1995) in the bindin sequences and cannot account for the strong differences seen between replacement and silent substitution.

Second, we compared changes in introns to silent changes in coding regions. Among two *E. mathaei* and two *E. oblonga* alleles for which complete introns were sequenced, the average proportion of changes occurring in the intron (0.052 ± 0.010) is similar to the average proportion of silent substitutions per silent site for the complete coding region (0.044 ± 0.010) and for codons 80–120 (0.039 ± 0.038). Silent sites in the coding sequence appear to be as free to vary as introns.

Table 2
Numbers of Homologous and Orthologous Comparisons of Repeated Elements in Which $P_n > P_s$

Repeat Set	$P_N > P_S$/Total	
	Homologous	Orthologous
a*	9/12	17/93
b*	7/9	7/32
c*	7/9	0/11
d*	9/9	0/12
e	0/2	0/4
f	0/2	0/4

Note.—For each repeat type, all possible pairwise comparisons were made between different sequences that are homologous as defined in our alignment. Likewise, all possible comparisons were made between the different sequences of a repeat type that are orthologous (i.e., not aligned). Location of repeat types a–f is indicated in figure 1. Asterisks indicate repeat types that overlap the region under positive selection (codons 80–120).

In addition, we considered the possibility that high P_n/P_s ratios were an artefact of aligning shuffled repeated sequences instead of truly homologous sequences. To do this, we compared the P_n/P_s ratios in homologous regions with those between orthologous repeated regions. For each of the short repeat types indicated in figure 1, P_n/P_s ratios from all possible homologous (aligned) pairs of sequences were compared to those from orthologous (non-aligned) pairs of sequences. In most of the homologous comparisons $P_n/P_s > 1$, but the reverse was found for orthologous comparisons (table 2). These results are consistent with the action of positive selection and accumulation of silent differences between orthologous repeats that presumably arose by duplication.

Recombination

We used the conservative 'four gamete test' (Hudson and Kaplan 1985; Kliman and Hey 1993) to obtain a minimum estimate of recombination. Under the assumption of no parallel changes for two characters, the only way all four possible allelic types can be found in a sample of alleles is for at least one recombination event to have occurred between the two sites (Hudson and Kaplan 1985). Positive selection may favor convergent changes at a particular nucleotide site, so we restricted our analysis to variable characters defined either as (1) a unique insertion/deletion (excluding polyglycines) or (2) two or more unique nucleotide differences that always group together. Despite the conservative nature of this test, we identified three recombination events in *E. mathaei* (fig. 4). Two of these alleles related by recombination are the apparently paraphyletic *E. mathaei* alleles (M-H 1 and M-G 86D, see figs. 1 and 2).

Downloaded from https://academic.oup.com/mbe/article/13/2/397/983302 by guest on 21 August 2022

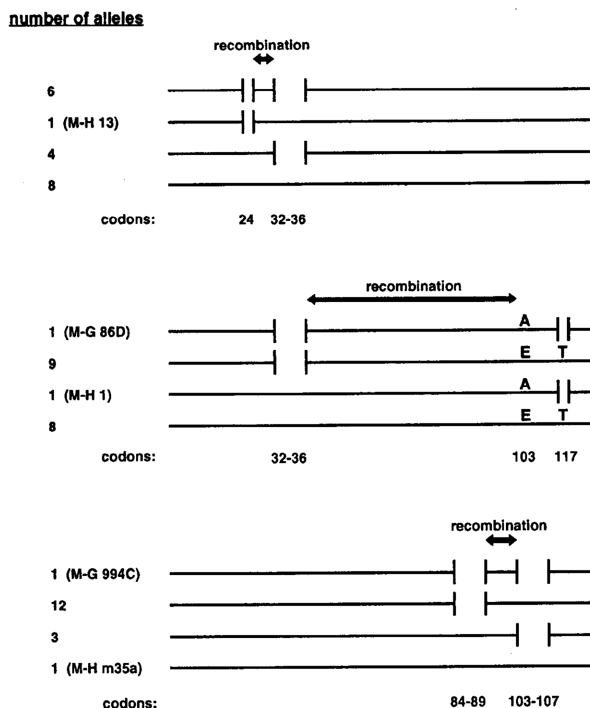

FIG. 4.—Evidence of recombination in *E. mathaei* bindin provided by the "four-gamete test" (Hudson and Kaplan 1985, see text). Schematic diagram shows aligned 5′ bindin alleles. In each of the three cases, all four possible combinations of two unique blocks of sequence in two regions are found in our sample of alleles. To account for this, recombination events probably occurred within the ranges shown by bold arrows. Breaks indicate alignment gaps; letters refer to amino acid substitutions that always occur together. Aligned codon positions of gaps and substitutions (see fig. 1*A*) are given below each set of alleles. The number of alleles in our sample with each profile is shown; when only one allele has a particular profile, the allele designation is given.

In the high replacement region (codons 80–120), these two sequences are approximately equidistant from the remaining alleles from all three species but differ from all alleles at two amino acid positions. However, the full-length sequence available for one of these alleles (M-H 1) groups with other full-length *E. mathaei* alleles in PAUP reconstructions (not shown).

Counting four insertion/deletions and three recombinations (but excluding polyglycine slippage), a minimum of seven unique events that rearranged blocks of sequence are identified in the 19 5′ bindin sequences from *E. mathaei.* As a result of combined effects of replacement, insertion/deletion, and recombination, bindins within species differ at up to 10% of the polypeptide. Most of this variation is found in the first 120 codons, where bindin alleles differ at up to 26 codon positions in our alignment (alleles M-H 1 and M-H m35B, fig. 1*A*) and differ in length by up to 10 codons (e.g., alleles O-H 1 and O-NG 16). In the 40-codon region under positive selection, up to five amino acid positions in our alignment differ in charge between alleles from the same species (e.g., alleles M-H 13 and M-G 86D). Because the sequences are monophyletic within morphological species, this polymorphism must have arisen since the recent divergence of these urchin species.

Discussion

Bindin sequences from *Echinometra* show highly polymorphic alleles that group monophyletically within three closely related morphological species sampled from widely distributed geographical areas. Strong incompatibilities of sperm–egg attachment, which involves the interaction of bindin and its receptor, exist between the *Echinometra* species in this study (Palumbi and Metz 1991; Metz et al. 1994). These incompatibilities presumably result from functional differences in recognition that depend on bindin amino acid sequence differences that distinguish species. "MGGXX" repeat units, which have been suggested to have a role in species-specificity of bindins from other genera of urchins (Minor et al. 1991; Lopez, Miraglia, and Glabe 1993), are variable among *Echinometra* alleles, but most of these differences are not fixed between species. Instead, the seven amino acid alignment positions that differ in charge or polarity between *Echinometra* species may be important for species-specific gamete recognition (fig. 1).

Fixed differences, and the absence of ancestral polymorphism shared between species, show that bindin alleles within each of these three species have had a common ancestor since speciation 1–2 Mya (Palumbi and Metz 1991). A similar pattern of species-specific allele clades was seen for the *period* locus (Kliman and Hey 1993), which may have a role in mate recognition of *Drosophila.* Interestingly, although mitochondrial DNA haplotypes are expected to drift to loss or fixation more rapidly than neutral nuclear loci, some mitochondrial haplotypes are shared among these three *Echinometra* species, suggesting that selection has acted to fix species-specific changes at the bindin locus (S.R.P., unpublished results).

Adaptive evolution between species was not supported by the MacDonald-Kreitman test, but bindin alleles have an unusually high level of amino acid-replacing polymorphism. Previous applications of this test (McDonald and Kreitman 1991; Eanes, Kirchner, and Yoon 1993; Kliman and Hey 1993) worked with sequences showing ratios of replacement to silent polymorphism within species of 0.15 or less. For the 5′ bindin sequences, this intraspecific ratio is 1.7, suggesting that replacement substitutions are favored by selection *within* species. In this case, selection could produce similar high replacement to silent ratios both within and between species, which would lead to a failure of the

MacDonald-Kreitman test for selection. A high replacement/silent ratio for intraspecific polymorphism has also been found at the *Drosophila melanogaster ref(2)P* locus (Dru et al. 1993), in which diversification within species is thought to result from the role of ref(2)P in controlling slightly deleterious viral infections.

Recombination events cannot be resolved to a short 'hotspot' region in bindin, but G+C-rich regions (including polyglycine codons and "MGGXX" units) with repeat and inverted repeat structure are similar to hotspots of recombination known for other loci (Lebo et al. 1983; Kobori et al. 1986). Even in highly variable regions, short repeats and polyglycines have been conserved in bindin across genera (Minor et al. 1991; Lopez, Miraglia, and Glabe 1993). Similarly, a threonine-glycine repeat region of the *Drosophila* period gene is highly variable, yet it is required for function of the protein (Peixoto et al. 1993). Peixoto et al. (1993) suggested that a predisposition to sequence rearrangement could facilitate evolution of functional diversity at this locus. Bindin alleles appear to have much the same predisposition. In addition to acting on amino acid replacements, positive selection is also likely to act on variants generated by insertion/deletion and recombination among alleles.

Accumulation of three times more replacement substitutions than silent substitutions per site provides strong evidence that differences between species reflect positive selection in a 40-codon domain of bindin (fig. 3 and table 1). Although the mechanism of interaction with the egg-surface receptor remains unknown, bindin is not highly modified after translation, so its function is determined primarily by its amino acid sequence (Lopez, Miraglia, and Glabe 1993). As in other proteins (e.g., Hughes and Nei 1988), different amino acid positions in bindin probably experience different degrees of selection. Part of the region of bindin under positive selection is highly charged; up to half of the amino acids between positions 96 and 122 are negatively charged in *Echinometra* alleles (fig. 1). Conservation of this negative charge signature in bindins from other genera (see Vacquier, Swanson, and Hellberg 1995, fig. 5) suggests that this region is functionally important despite its variability.

Why positive selection should favor polymorphism in bindin is not yet clear. It is challenging to suggest a plausible mechanism by which selection could favor diverse alleles within a species at a gamete recognition locus. Instead, such a locus is expected to undergo strong purifying or directional selection for efficient gamete recognition (Carson 1987). Some systems in which positive selection has been documented, including MHC loci in primates and rodents (Hughes and Nei 1988; Hughes, Ota, and Nei 1990; Hughes, Hughes, and Watkins 1993), surface antigens in *Plasmodium* (Hughes 1992, 1993), and self incompatibility loci in plants (Clark and Kao 1991; Clark 1993), are characterized by highly polymorphic alleles maintained by balancing selection for periods of time that transcend speciation. Clustering of bindin alleles within species (fig. 2) indicates that balancing selection probably was not a major force in *Echinometra* bindin evolution. Abalone lysin is a sperm surface recognition protein from a marine free-spawning invertebrate in which positive selection has generated marked interspecific differences (Lee and Vacquier 1992; Lee, Ota, and Vacquier 1995; Swanson and Vacquier 1995). However, lysin polymorphism within abalone species has not yet been observed. Vacquier and Lee (1993) have suggested that egg-surface receptors may be in an evolutionary race with microbial pathogens gaining access to the egg, although there is no direct evidence to support this idea.

Differences between bindin alleles are pronounced and may result in functional differentiation *within* species as well as between species. The concentration of sperm required for 100% fertilization varies from male to male (S.R.P., unpublished results), but very high percentages of eggs become fertilized in all intraspecific *Echinometra* crosses with high sperm concentrations (Palumbi and Metz 1991; Metz et al. 1994). Accordingly, the egg-surface receptor for bindin must be capable of recognizing the variety of polymorphic forms of bindin that we find within species. The binding domain of this receptor is related to the binding domain of hsp70 (Foltz, Partin, and Lennarz 1993). Like MHC, hsp70 molecules are capable of recognizing a suite of related peptides yet exhibit greatest affinities for particular peptides (Rippmann et al. 1991; Gething and Sambrook 1992; Elliot et al. 1993; Gragerov and Gottesman 1994). This suggests that in a similar manner the egg-surface receptor molecule might have greatest affinity for certain forms of bindin, while accepting a range of related bindins.

Given some tolerance by the receptor (female) locus, many slightly different functional variants at the bindin (male) locus could accumulate in natural populations instead of being removed rapidly by purifying selection. Under these conditions, selection on the receptor locus would be relaxed, because all eggs eventually fertilize, regardless of which allelic form of bindin is encountered on sperm. However, selection could act on the bindin locus if a particular receptor had greatest affinity for a particular bindin. Especially under conditions of mass spawning common in sea urchins, receptors on the egg surface may select among competing sperm based on affinities for different bindins.

Key elements of models that lead to the evolution of assortative mating include functional polymorphism

Downloaded from https://academic.oup.com/mbe/article/13/2/397/1033026 by guest on 21 August 2022

of mate recognition loci (Nei, Maruyama, and Wu 1983; Wu 1985) and relaxed selection on female loci (Wu 1985). In such models, sexual isolation arises most readily when multiple alleles at a female choice locus interact with multiple alleles at a male trait locus in partially isolated subpopulations. Sea urchins and other marine taxa with extensive ranges are subject to fluctuation of population sizes and short-term fragmentation of populations by a variety of biogeographic processes (Palumbi 1994). We have found that bindin allele frequencies based on restriction fragment polymorphism differ between *Echinometra* populations across the Indo–Pacific (S.R.P. and E.C.M., unpublished results). Thus the action of selection on polymorphic bindin loci may set the stage for population level differentiation of the specificity of fertilization and might be a prelude to species formation.

Sequence Availability

The nucleotide sequences reported in this paper have been deposited in the GenBank data base: accession numbers U39502–U39544.

Acknowledgments

We thank J. Minor, R. Britten, and E. Davidson for providing *Strongylocentrotus* bindin sequences we used to design PCR primers, W. K. Thomas for providing a preprobindin primer, D. Hansell, L. Tashima, and G. Bryant-Greenwood for help with mRNA isolation, and M. Rivera for help sequencing. We also thank C. S. Baker, C. Biermann, T. Duda, K. Foltz, R. Hickson, A. Hofmann, B. Kessing, Y.-H. Lee, H. Lessios, A. P. Martin, M. McCartney, W. O. McMillan, G. Roderick, S. Romano, and V. Vacquier for technical advice and discussions. Supported by NSF predoctoral fellowship and Edmondson Research Fellowship to E.C.M. and NSF grants to S.R.P.

LITERATURE CITED

CARSON, H. L. 1987. The genetic system, the deme, and the origin of species. Ann. Rev. Genet. **21**:405–423.

CLARK, A. G. 1993. Evolutionary inferences from molecular characterization of self-incompatibility alleles. Pp. 79–108 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Japan Scientific Societies Press, Tokyo, and Sinauer Associates, Inc., Sunderland, Mass.

CLARK, A. G., and T.-H. KAO. 1991. Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proc. Natl. Acad. Sci. USA **88**:9823–9827.

COYNE, J. A. 1992. Genetics and speciation. Nature **355**:511–515.

DRU, P., F. BRAS, S. DEZELEE, P. GAY, A.-M. PETITJEAN, A. PIERRE-DENEUBOURG, D. TENINGES, and D. CONTAMINE. 1993. Unusual variability of the *Drosophila melanogaster ref(2)P* protein which controls the multiplication of sigma rhabdovirus. Genetics **133**:943–954.

EANES, W. F., M. KIRCHNER, and J. YOON. 1993. Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90**:7475–7479.

ELLIOT, T., M. SMITH, P. DRISCOLL, and A. MCMICHAEL. 1993. Peptide selection by class I molecules of the major histocompatibility complex. Current Biol. **3**:854–866.

FOLTZ, K. R., and W. J. LENNARZ. 1993. The molecular basis of sea urchin gamete interactions at the egg plasma membrane. Dev. Biol. **158**:46–61.

FOLTZ, K. R., J. S. PARTIN, and W. J. LENNARZ. 1993. Sea urchin egg receptor for sperm: sequence similarity of binding domain and hsp70. Science **259**:1421–1425.

FROHMAN, M. A., M. K. DUSH, and G. R. MARTIN. 1988. Rapid amplification of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.

GAO, B., L. E. KLEIN, R. J. BRITTEN, and E. H. DAVIDSON. 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.

GENETICS COMPUTER GROUP. 1991. Genetics computer group package. Genetics Computer Group, Madison, Wis.

GETHING, M.-J., and J. SAMBROOK. 1992. Protein folding in the cell. Nature **355**:33–45.

GLABE, C. G., and D. CLARK. 1991. The sequence of the *Arbacia punctulata* bindin cDNA and implications for the structural basis of species-specific sperm adhesion and fertilization. Dev. Biol. **143**:282–288.

GRAGEROV, A., and M. E. GOTTESMAN. 1994. Different peptide binding specificities of hsp70 family members. J. Mol. Biol. **241**:133–135.

HUDSON, R. R., and N. L. KAPLAN. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–164.

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

———. 1993. Coevolution of immunogenic proteins of *Plasmodium falciparum* and the host's immune system. Pp. 109–127 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Japan Scientific Societies Press, Tokyo, and Sinauer Associates, Inc., Sunderland, Mass.

HUGHES, A. L., M. K. HUGHES, and D. I. WATKINS. 1993. Contrasting roles of interallelic recombination at the *HLA-A* and *HLA-B* loci. Genetics **133**:669–680.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

HUGHES, A. L., T. OTA, and M. NEI. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. **7**:515–524.

Downloaded from https://academic.oup.com/mbe/article/13/2/397/983302 by guest on 21 August 2022

IRWIN, D. M., T. D. KOCHER, and A. C. WILSON. 1991. Evolution of the cytochrome b gene of mammals. J. Mol. Evol. **32**:128–144.

KLIMAN, R. M., and J. HEY. 1993. DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133**:375–387.

KOBORI, J. A., E. STRAUSS, K. MINARD, and L. HOOD. 1986. Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. Science **234**: 173–179.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park.

LEBO, R. V., A. CHAKRAVARTI, K. H. BUETOW, M.-C. CHEUNG, H. CANN, B. CORDELL, and H. GOODMAN. 1983. Recombination within and between the human insulin and beta-globin gene loci. Proc. Natl. Acad. Sci. USA **80**:4808–4812.

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.

LEE, Y.-H., and V. D. VACQUIER. 1992. The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. Biol. Bull. **182**:97–104.

LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

LOPEZ, A., S. J. MIRAGLIA, and C. G. GLABE. 1993. Structure/function analysis of the sea urchin sperm adhesive protein bindin. Dev. Biol. **156**:24–33.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351**:652–654.

METZ, E. C., R. E. KANE, H. YANAGAMACHI, and S. R. PALUMBI. 1994. Fertilization between closely related sea urchins is blocked by incompatibilities during sperm–egg attachment and early stages of fusion. Biol. Bull. **187**:23–34.

MINOR, J. E., D. R. FROMSON, R. J. BRITTEN, and E. H. DAVIDSON. 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.

MINOR, J. E., B. GAO, and E. H. DAVIDSON. 1989. The molecular biology of bindin. Pp. 73–88 *in* H. SCHATTEN and G. SCHATTEN, eds. The molecular biology of fertilization. Academic Press, San Diego.

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

NEI, M., and L. JIN. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6**:290–300.

NEI, M., T. MARUYAMA, and C.-I. WU. 1983. Models of evolution of reproductive isolation. Genetics **103**:557–579.

PALUMBI, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Ann. Rev. Ecol. Syst. **25**:547–572.

PALUMBI, S. R., and C. S. BAKER. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. **11**:426–435.

PALUMBI, S. R., and E. C. METZ. 1991. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). Mol. Biol. Evol. **8**:227–239.

PEIXOTO, A. A., S. CAMPESAN, R. COSTA, and C. P. KYRIACOU. 1993. Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. Mol. Biol. Evol. **10**:127–139.

POMIANKOWSKI, A., and L. SHERIDAN. 1994. Linked sexiness and choosiness. Trends Ecol. Evol. **9**:242–244.

RIPPMANN, F., W. R. TAYLOR, J. B. ROTHBARD, and N. M. GREEN. 1991. A hypothetical model for the peptide binding domain of hsp70 based on the peptide binding domain of HLA. EMBO J. **10**:1053–1059.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5**:704–716.

SUMMERS, R. G., and B. L. HYLANDER. 1975. Species-specificity of acrosome reaction and primary gamete binding in echinoids. Exp. Cell Res. **96**:63–68.

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.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Smithsonian Institution, Washington, D.C.

VACQUIER, V. D., and Y.-H. LEE. 1993. Abalone sperm lysin: unusual mode of evolution of a gamete recognition protein. Zygote **1**:181–196.

VACQUIER, V. D., and G. W. MOY. 1977. Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. Proc. Natl. Acad. Sci. USA **74**:2456–2460.

VACQUIER, V. D., W. J. SWANSON, and M. E. HELLBERG. 1995. What have we learned about sea urchin sperm bindin? Dev. Growth Differ. **37**:1–10.

WU, C.-I. 1985. A stochastic simulation study on speciation by sexual selection. Evolution **39**:66–82.

WU, C.-I., H. HOLLOCHER, D. J. BEGUN, C. F. AQUADRO, Y. XU, and M.-L. WU. 1995. Sexual isolation in *Drosphila melanogaster*: a possible case of incipient speciation. Proc. Natl. Acad. Sci. USA **92**:2519–2523.

SHOZO YOKOYAMA, reviewing editor

Accepted October 27, 1995

Downloaded from https://academic.oup.com/mbe/article/13/2/397/983302 by guest on 21 August 2022