

# Characterization of the Sperm Molecule Bindin in the Sea Urchin Genus *Paracentrotus*

Isabel Calderón · Xavier Turon · H. A. Lessios

Received: 29 October 2008 / Accepted: 19 February 2009 / Published online: 28 March 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** Bindin is a sea urchin gamete-recognition protein that plays an essential role in the specificity of egg–sperm interactions and thus may be evolving under sexual selection and be related to speciation. Bindin has been found to evolve under strong selection in some sea urchin genera and neutrally in others. In this study, we characterized bindin in the two extant species of the genus *Paracentrotus*: *P. lividus* from the Atlanto-Mediterranean region and *P. gaimardi* from Brazil. The structure of the bindin molecule in *Paracentrotus* is similar to that of other genera studied thus far, consisting of a conserved core flanked by two variable regions and an intron of variable length located at the same conserved position as in other genera. Polymorphism in *P. lividus* is caused mainly by point substitutions and insertions/deletions, and length variations are caused mainly by the number of repeated motifs in the flanking regions. There is no evidence of recombination. Positive selection is acting on amino acid sites located in two regions flanking the conserved core.

**Keywords** Bindin · PAML · *Paracentrotus* · Positive selection · Sea urchins · Speciation

## Introduction

Proteins involved in reproduction and, more specifically, in gamete recognition—which control the binding and fusion of gametes—may play an essential role in reproductive isolation and speciation in marine organisms (reviewed in Palumbi 1994, 1998; Vacquier et al. 1995; Swanson and Vacquier 2002; Lessios 2007). Reproductive isolation in broadcast spawners may be caused by a number of factors, such as nonoverlapping spawning times, chemical communication, or species-specific gamete interactions. Although evolutionary theory suggests that traits closely related to fitness are under purifying selection (Nei 1987), reproductive proteins often show high rates of diversification (Civetta and Singh 1998; Swanson and Vacquier 2002). Evolutionary patterns of proteins directly implicated in reproductive isolation can provide important information on speciation processes (Coyne 1992; Palumbi 1994; Coyne and Orr 2004).

Marine invertebrates often possess gamete-recognition proteins that are subject to positive selection. In abalones and teguline gastropods, several studies have identified the role of positively selected amino acid sites scattered over the entire length of the protein lysin (Lee and Vacquier 1992; Lee et al. 1995; Swanson and Vacquier 1997) that affect secondary structure of the protein (Hellberg and Vacquier 1999). The abalone lysin changes under positive selection to adapt to variations of the receptor molecule, VERL, located on the egg's surface (Swanson and Vacquier 2002). Interestingly, another abalone sperm protein, sp18, a paralog of lysin (Vacquier et al. 1997), is among the most rapidly evolving metazoan proteins known (Swanson and Vacquier 1995).

In sea urchins, the protein bindin constitutes the major insoluble component of the sperm acrosomal vesicle

---

I. Calderón (✉)  
Department of Animal Biology, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain  
e-mail: calderon@ub.edu

X. Turon  
Center for Advanced Studies of Blanes (CEAB, CSIC), Blanes, Girona, Spain

H. A. Lessios  
Smithsonian Tropical Research Institute, Balboa, Republic of Panama

(Vacquier and Moy 1977; Ulrich et al. 1998). When sperm undergo the acrosomal reaction, bindin is deposited on the tip of the sperm acrosomal process after exocytosis from the vesicle. This protein mediates adhesion of sperm to eggs and fusion between their membranes in a species-specific manner, this being the key stage of the recognition process (Glabe and Vacquier 1977; Glabe and Lennarz 1979; Metz et al. 1994; Vacquier et al. 1995; Vacquier 1998).

The last few years have seen an increase in the number of studies aimed at understanding the evolution of bindin in various sea urchin genera. Zigler and Lessios (2003a) found that bindin was present in a common ancestor of all extant sea urchins >250 million years ago. Different evolutionary patterns have been observed in the 10 genera of echinoids for which bindin has been studied thus far (Lessios 2007). Positive selection has been detected in particular domains of bindin in the genera *Echinometra* (Metz and Palumbi 1996; Geyer and Palumbi 2003; McCartney and Lessios 2004), *Strongylocentrotus* (Biermann 1998), and *Heliocidaris* (Zigler et al. 2003), whereas purifying selection or neutral evolution has been found in *Arbacia* (Metz et al. 1998), *Tripneustes* (Zigler and Lessios 2003b), and *Lytechinus* (Zigler and Lessios 2004). Interestingly, these results suggest that positive selection operates in genera in which species are sympatric; however, it does not operate in genera in which all (or nearly all) species are allopatric. This pattern is not necessarily related to selection to avoid hybridization in sympatric species; however, it may simply reflect the fact that species with incompatible gametes can persist in sympatry (Zigler and Lessios 2003b; McCartney and Lessios 2004).

High levels of polymorphism in bindin have been found in several species (Metz and Palumbi 1996; Biermann 1998), and indications of differential fertilization efficiency within species related to bindin types have been reported (Palumbi 1999; Levitan and Farrell 2006). In addition to reinforcement, sperm competition, interlocus sexual conflict, and sexual selection can accelerate divergence in bindin, leading to species differentiation as a result of intraspecific processes (Vacquier et al. 1997; Howard 1999; McCartney and Lessios 2004; Levitan and Farrell 2006). Comparisons between patterns of bindin evolution of different species of sea urchins can provide information on the processes that have shaped them. It is thus useful to characterize this molecule in as many genera as possible.

The genus *Paracentrotus* is a member of the order Echinoidea, the only order in which bindin has been found to evolve under positive selection. This genus contains only two species: *P. lividus* (Lamarck 1816) is abundant in the Mediterranean Sea and is also found in the eastern Atlantic from the British islands to South Morocco, including Macaronesia. *P. gaimardi* (Blainville 1825) is distributed

from the Gulf of Guinea to Angola on the African coast and from Rio de Janeiro to Florianopolis along the coast of Brazil (Mortensen 1943). Recent studies suggest that *P. lividus* can maintain gene flow over thousands of kilometres (Duran et al. 2004; Calderón et al. 2008). Similarly, the presence of *P. gaimardi* on both sides of the Atlantic Ocean suggests that its larvae can also disperse over long distances. The genetic distance between the two species based on the mitochondrial ATPase 8 and 6 regions is approximately 1.6% (Lessios, unpublished data), which, when calibrated by divergence in the same DNA region of other sea urchins by the rise of the Isthmus of Panama (Lessios 2008), suggests that *P. lividus* and *P. gaimardi* have been isolated during the last half million years. Thus, these two allopatric species provide an interesting model for the study of the evolution of bindin. In this study, we characterized the gene coding for bindin in the genus *Paracentrotus*, and we studied the intraspecific polymorphism and mode of evolution in *P. lividus*.

## Materials and Methods

Samples of *P. lividus* used for initial isolation of bindin mRNA were collected from Rosas (northwestern Mediterranean; Fig. 1). Gonads were removed and immediately preserved in RNAlater (Ambion, Austin, TX). Total RNA was extracted from testis tissue using the Totally RNA Kit (Ambion). Bindin was isolated using a 5' and 3' RACE kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions (Frohman et al. 1988; Park et al. 2003). For the 5' RACE, first-strand cDNA was synthesized with the primer CORE200-: 5'-TCY.TCY.TCY.TCY.TGC.ATI.GC-3' (Zigler and Lessios 2003b). Initial sequences were obtained by combining degenerate primers used by McCartney and Lessios (2004) and by Zigler and Lessios (2003b): MB1130+ (5'-TGC.TSG.GTG.CSA.CSA.AGA.TTG.A-3'), EL1207+ (5'-AGG.CAT.CAC.TCC.AAT.CTC.CTG.GC-3'), CORE 200- and CORE 157- (5'-CIG.GRT.CIC.CHA.TRT.TIG.C-3'). Specific primers for *P. lividus* were designed for the 5' end of the molecule (see later text). For 3' RACE, synthesis of first-strand cDNA was conducted with primer MOBY (5'-AAG.GAT.CCG.TCG.ACA.TCG.ATA.ATA.CGA.CTC.ACT.ATA.GGG.A(T)<sub>18</sub>-3') using nested primers out2 (5'-GAT.CCG.TCG.ACA.TCG.ATA.ATA.CG-3') and in2 (5'-CGA.TAA.TAC.GAC.TCA.CTA.TAG.G-3') (Zigler and Lessios 2003b). Specific primers for *P. lividus* were subsequently designed from the 3' end of the molecule.

The following primers were used for amplification in *Paracentrotus*: PreCleF (located upstream of the preprobindin cleavage site): 5'-CAG.GTG.ATA.CAG.AAA.GAA.GCG.GT-3' and Mid-3'UTR (located downstream of the bindin stop codon): 5'-ATG.TCG.TTG.CAA.TCA.TGA.

**Fig. 1** Localities at which *P. lividus* was sampled



AGG-3'. These primers amplified the region coding for the entire length of the protein from genomic DNA, including the bindin intron (see later text). The same primers were used to amplify one sample of *P. gaimardi* collected in Brazil (Baia de Sepetiba, 23°03' S, 43°45' W) and preserved in high-salt dimethylsulfoxide solution. Amplifications were performed with 1 U DyNAzyme EXT DNA proofreading polymerase (Finnzymes, Espoo, Finland) with 0.4  $\mu$ M each primer. The following polymerase chain reaction (PCR) conditions were used: 94°C for 2 min, 30 cycles of 94°C for 45 s, 47°C for 30 s, and 72°C for 3 min, followed by a final extension of 72°C for 5 min. PCR products were separated on a 1.2% low-melting agarose gel, and DNA was extracted with 0.5 U Gelase (Epicentre Biotechnologies, Madison, WI).

To analyze intraspecific variability in *P. lividus*, we studied individuals from 15 geographically distant localities covering the Atlanto-Mediterranean distribution range of the species (Fig. 1 and Table 1). Gel-purified PCR amplicons were cloned with pGEM-Easy Vector cloning kit (Promega, Madison, WI) to separate individual alleles. Two to five colonies from one individual/population were sequenced using bacterial colonies directly as template for PCR amplification with vector primers M13 F and M13 R. The same PCR conditions as detailed previously were used, with two exceptions: annealing temperature was 50°C, and 1 U *Taq* polymerase (Promega) was used to amplify these clones. Sequencing of both strands was performed using M13 F and M13 R on an ABI 3100 automated sequencer (Applied Biosystems, Carlsbad, CA), and sequences were aligned using Sequencher 4.2 (Gene Codes Corp, Ann Arbor, MI). The single individual of *P. gaimardi* was cloned according to the same protocol as detailed

**Table 1** Populations of *Paracentrotus* studied, number of clones sequenced, and number of unique haplotypes encountered<sup>a</sup>

Population	No. of clones (no. of haplotypes)
<i>P. lividus</i> (total)	63 (23)
Cyprus	5 (1)
Peloponnese	5 (2)
Naples	5 (2)
Corsica	3 (1)
Rosas	5 (2)
Nao	5 (1)
Gata	4 (2)
Ceuta	2 (1)
Tarifa	5 (2)
Cádiz	5 (2)
Azores	6 (2)
Madeira	5 (2)
Tenerife	2 (1)
Roscoff	3 (1)
Ireland	3 (2)
<i>P. gaimardi</i>	5 (1)

<sup>a</sup> One individual per population was sampled

previously. All sequences have been deposited in GenBank (accession nos. FJ713291 to FJ713314 for *P. lividus* and FJ713315 for *P. gaimardi*).

A total of 51 singleton mutations were observed on various sites among the 63 sequences analyzed of *P. lividus* (9 singletons in 5 sequences of *P. gaimardi*). These singleton mutations may represent true mutations or may have arisen from polymerase errors during amplification or cloning. Despite the relatively low rate of error of *Taq*

polymerase, it is likely that some of the singletons observed are in fact PCR and cloning artefacts. Therefore, polymorphisms that appeared only once (i.e., in a single clone) in the whole data set were replaced so as to match the consensus sequence (Villablanca et al. 1998). This may have lead to slightly conservative results; however, it is unlikely that it modified general conclusions.

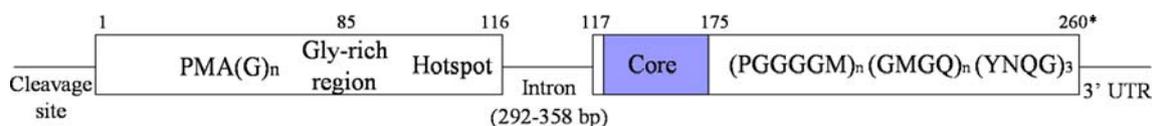
### Sequence Analysis and Tests for Selection

We tested for recombination in our sequences using several algorithms implemented in the Recombination Detection Program (Martin et al. 2005). Bindin haplotype and nucleotide diversity were estimated using DnaSP v. 4.10 (Rozas et al. 2003). For subsequent analyses, bindin was divided in three regions based on patterns of bindin variation detected in other genera: a hotspot region 5' from the conserved core, corresponding to amino acids 85 to 116; a conserved core of 55 amino acids; and a third region corresponding to the rest of the mature molecule (the mature protein includes the whole coding region, 3' from the cleavage site to the stop codon, excluding the intron; Fig. 2). We also identified the single bindin intron between the two exons 5' of the core. For each of these regions, we used MEGA 4.0 (Tamura et al. 2007) to calculate differences in the proportion of synonymous ( $d_S$ ) and nonsynonymous sites ( $d_N$ ) by the method of Pamilo and Bianchi (1993) and Li (1993). Evidence for positive selection was sought for each region of the molecule using Fisher's exact tests on all pairwise comparisons (Zhang et al. 1997) under Nei and Gojobori's (1986) model of evolution available in MEGA 4. The statistical analysis of protein sequences (SAPS; [http://www.isrec.isb-sib.ch/software/SAPS\\_form.html](http://www.isrec.isb-sib.ch/software/SAPS_form.html)) was used to identify separated repeats, simple tandem repeats, and period repeats in each species (Brendel et al. 1992). DnaSP was used to calculate the effective number of codons (ENC) as a measure of codon use bias (Wright 1990).

We computed the McDonald and Kreitman (1991) (MK) test to detect selection on bindin by comparing the ratio of nonsynonymous to synonymous nucleotide changes fixed between species with the same ratio of changes polymorphic within species using DnaSP. However, the criterion that average  $\omega = d_N/d_S$  over the entire molecular sequence should be  $>1$  is too stringent for detecting positive selection that may affect only a few amino acid sites (Crandall

et al. 1999). Thus, we used module Codeml in PAML (Yang 2007) to estimate  $d_N/d_S$  in lineages and sites based on a maximum-likelihood (ML) tree. This tree was constructed using TreeFinder (10000 replicates; Jobb 2007) based on the best-fit model of molecular evolution as determined by ModelTest v.3.8 (Posada and Crandall 1998; Posada 2006). To test for positive selection between the two species, a branch-specific model was used with one ratio  $d_N/d_S$  for *P. gaimardi* and another for all samples of *P. lividus*. This model was compared with a null model with a single ratio for all branches using a likelihood ratio test between the null and the alternative models ( $LRT; 2(\ln L_2 - \ln L_1)$ ). This was compared with the  $\chi^2$  distribution with the degrees of freedom being equal to the difference in the number of parameters between the two models (Nielsen and Yang 1998; Yang et al. 2000). We also tested if any other internal branch of the tree was subject to positive selection to determine whether constraints were acting at a local level, assigning a ratio  $\omega_1$  to the tested branch and a background ratio  $\omega_0$  to all other branches.

To test whether specific sites in the bindin protein have evolved under positive selection, site-specific models were tested. These models were the nearly neutral model (M1a, according to the terminology used in Yang et al. 2005, which specifies two  $d_N/d_S$  ratios:  $0 < \omega_0 < 1$  and  $\omega_1 = 1$ ), positive selection model (M2a, which includes an additional site class with a  $d_N/d_S$  ratio  $\omega_2 > 1$  estimated from the data), beta model [M7, with a flexible beta distribution of  $d_N/d_S$  among sites but limited to the interval [0,1], where 0 represents complete constraint and 1 the expectation under no selective constraint), and beta& $\omega$  model (M8, with an extraclass of sites with  $\omega_2 > 1$  estimated from the data; Yang et al. 2005). Neutral models (M1a and M7), which do not allow for positively selected sites, were compared with alternative models (M2a and M8), which do allow for positively selected sites (see Swanson et al. 2001). According to Yang et al. (2000), comparisons between M1a and M2a as well as between M7 (beta) and M8 (beta& $\omega$ ) provide a powerful test for positive selection. Bayes Empiric Bayes (BEB) calculation of posterior probabilities for site classes was implemented for models M2a and M8 (Yang et al. 2005). We also applied branch-sites models to simultaneously survey variation in selection across sites and lineages, considering *P. gaimardi* the background branch and the ancestral branch of *P. lividus* as a foreground branch. Branch-site models attempt to detect



**Fig. 2** Schematic representation of the gene coding for bindin molecule in *P. lividus*. Numbers represent the amino acidic position in the mature bindin molecule. \*Stop codon. 3'UTR—Untranslated region. See text for details

positive selection that affects only a few sites along a few lineages. We computed the new test 2 (i.e., branch-site test of positive selection), wherein model A with  $\omega_2$  estimated from the data are compared with the null model A with fixed  $\omega_2 = 1$  (Yang et al. 2005; Zhang et al. 2005). Test 2 is a robust test to detect positive selection on foreground branches (Zhang et al. 2005).

## Results

We isolated the region coding for whole mature bindin protein in both *P. lividus* and *P. gaimardi*. The coding portion of this DNA region in *P. lividus* ranged from 708 (Roscoff) to 759 (Cyprus) bp from the preprobindin cleavage site to the stop codon, thus coding for a protein from 236 to 253 amino acids. A conserved multibasic motif RMKR or RTKR (Gao et al. 1986) after the 5' preprobindin region indicated the proteolytic cleavage site, which has been present in all echinoids studied thus far (Zigler and Lessios 2003a). As observed in other species, this molecule contained a first exon, including a hypervariable hotspot, and a second exon, including the conserved 55-amino acid core with 18 invariable amino acids involved in membrane fusion (Ulrich et al. 1998, 1999) and a fragment from 70 to 82 amino acids long containing various repeated polyglycine motifs 3' of the core. These two exons were separated by an intron ranging between 292 and 358 bp located after a conserved valine, 5' from the conserved core region (Fig. 2).

At the nucleotide level, point mutations and indels were responsible for variation in our sequences. No events of recombination were detected in our data set. Indeed, bindin in the genus *Paracentrotus* presented variations in length mainly because of changes in the number of repeats in the glycine-rich motifs. The first exon presented a repeated PMA(G<sub>n</sub>) motif followed by a glycine-rich region that was highly variable in length. Similarly, 3' from the core region, two different glycine-rich motifs (PGGGGM and GMGQ) were repeated, accounting for noticeable

variations in length ( $\leq 24$  bp) in the second coding region. One of the alleles found in the Peloponnese (Greece) contained an insertion FQ not observed in any other population. Finally, a motif YNQG just before the stop codon was repeated three times in every individual analyzed.

In the case of the single individual of *P. gaimardi* analyzed, the cleavage site RMKR preceded a 115 amino acid—long first coding region and a 130 amino acid—long second coding region, including the conserved core of 55 amino acids. These two regions were separated by an intron of 519 bp. The general structure of *P. gaimardi*'s bindin gene was similar to that of *P. lividus*, and the main variation in length between these two species lay in the intron, which includes a large 159-bp insertion in *P. gaimardi*.

The 63 complete bindin sequences obtained from *P. lividus* yielded 23 unique haplotypes, of which only 1 was shared by 2 individuals from different populations (Madeira and Tarifa). Haplotype diversity was therefore high ( $H = 0.964$  over all samples). Of the 15 individuals of *P. lividus* analyzed, 6 (40%) were homozygotes, whereas the other 9 (60%) were heterozygotes (Table 1). The single *P. gaimardi* studied contained the same haplotype in 5 sequenced clones, except for singleton mutations, which we assume to be cloning or PCR artefacts.

Nucleotide diversity for the entire coding region of *P. lividus* was low (0.01489), whereas that of the intron was only slightly higher (0.01761; Table 2). In the case of *P. lividus*, most of the variation was observed at the first coding region upstream of the intron. Indeed, this first coding region exhibited average intraspecific divergence of 7.49% and also contained 75% of the sites detected as evolving under positive selection (see later text). On the contrary, the second coding region had average intraspecific divergence of 3.9% and 25% of the positively selected sites. Average intraspecific divergence in the conserved core was only 2.4%. The average number of mutations between two *P. lividus* was 8.57, lower than the average observed when comparing sequences of *P. lividus* with those of *P. gaimardi* (14.35). Seven sites were fixed between species. In the intron, 32 positions (excluding

**Table 2** Amino acid sequence length of bindin, excluding indels, nucleotide diversity, amino acid replacements ( $d_N$ ), and silent ( $d_S$ ) substitutions per site in three regions of the mature bindin of *P. lividus* and their ratio ( $\omega = d_N/d_S$ )<sup>a</sup>

Region	Length in <i>P. lividus</i>	Nucleotide diversity ( $\pi$ )	$d_N$	$d_S$	$\omega = d_N/d_S$	Length in <i>p. gaimardi</i>
Hotspot	31–33 aa	0.01690	0.010	0.036	0.2778	33 aa
Core	55 aa	0.00360	0	0.007	0	55 aa
Rest of the molecule	149–165 aa	0.01520	0.012	0.035	0.3429	158 aa
Total (mature)	236–253 aa	0.01489	0.014	0.022	0.6363	246 aa
Intron	292–358 bp	0.01761				519 bp

aa amino acid; bp base pair

<sup>a</sup> Mean of pairwise comparisons calculated by the method of Pamilo and Bianchi (1993) and Li (1993)

indels) were variable in *P. lividus*, of which 20 were parsimony informative. This intron also contained repeated motifs, although they were not the main contributors to variation in length in this noncoding region. Excluding differences in length, the average number of intron nucleotide substitutions in interspecific comparisons was 8.857 (6 of which were fixed between species), which was higher than the average number of substitutions found within *P. lividus* (4.748).

At the amino acid level, 20 (7.90%) of a maximum of 253 amino acid positions (excluding the indels) were variable within *P. lividus*, and only three fixed differences were found between *P. lividus* and *P. gaimardi*. Bindins of both species were identical to those of all other species studied to date (Zigler and Lessios 2003a) in the 18 amino acids involved in sperm–egg membrane fusion (Ulrich et al. 1999). At the amino acid level, 16 unique sequences were found in *P. lividus*, and an additional 1 was found in *P. gaimardi*. Like other mature bindins, the 2 species of *Paracentrotus* contained no cysteine or tryptophan residues. Glycine was by far the most common amino acid present in the mature bindin, accounting on average for >24% of the amino acid residues in both species. If the core region were excluded, glycine would account for >30% of the amino acid content of mature bindin. As in other species, leucine was the most frequent amino acid in the core, accounting for 17.2% of the 55 amino acids. There was no evidence of codon use bias in any of the observed sequences, with an average ENC of 59.68 in *P. lividus* and 56.51 in *P. gaimardi*.

#### Tests for Selection

As listed in Table 2, average  $d_N$  was lower than average  $d_S$  over the entire length of the mature protein. In the hotspot region and in the rest of the molecule,  $d_S$  was 3 times higher than  $d_N$ . The 55 amino acids within the core region showed no amino acid substitutions, and all changes were synonymous. Neither positive nor negative selection were evident in any of the 3 regions of the molecule ( $P > 0.05$ , Fisher's exact test for each region independently).

MK test for the entire mature protein showed no significant departure from neutrality in the genus *Paracentrotus*. Over the entire mature bindin, 3 nonsynonymous changes and 4 synonymous changes were fixed between species, and 17 replacement and 19 silent polymorphisms were found within *P. lividus* ( $P = 1.0$ , Fisher's exact test).

The ML tree for the observed bindin alleles of *P. lividus*, rooted on *P. gaimardi*, is shown in Fig. 3. Although some groups with several well-supported subgroups of haplotypes appeared, no geographic structure was apparent in the tree. A separate ML tree, based exclusively on the intron, confirmed the existence of some groups (Fig. 3), but the

distance between any two sequences was unrelated to geographic origin.

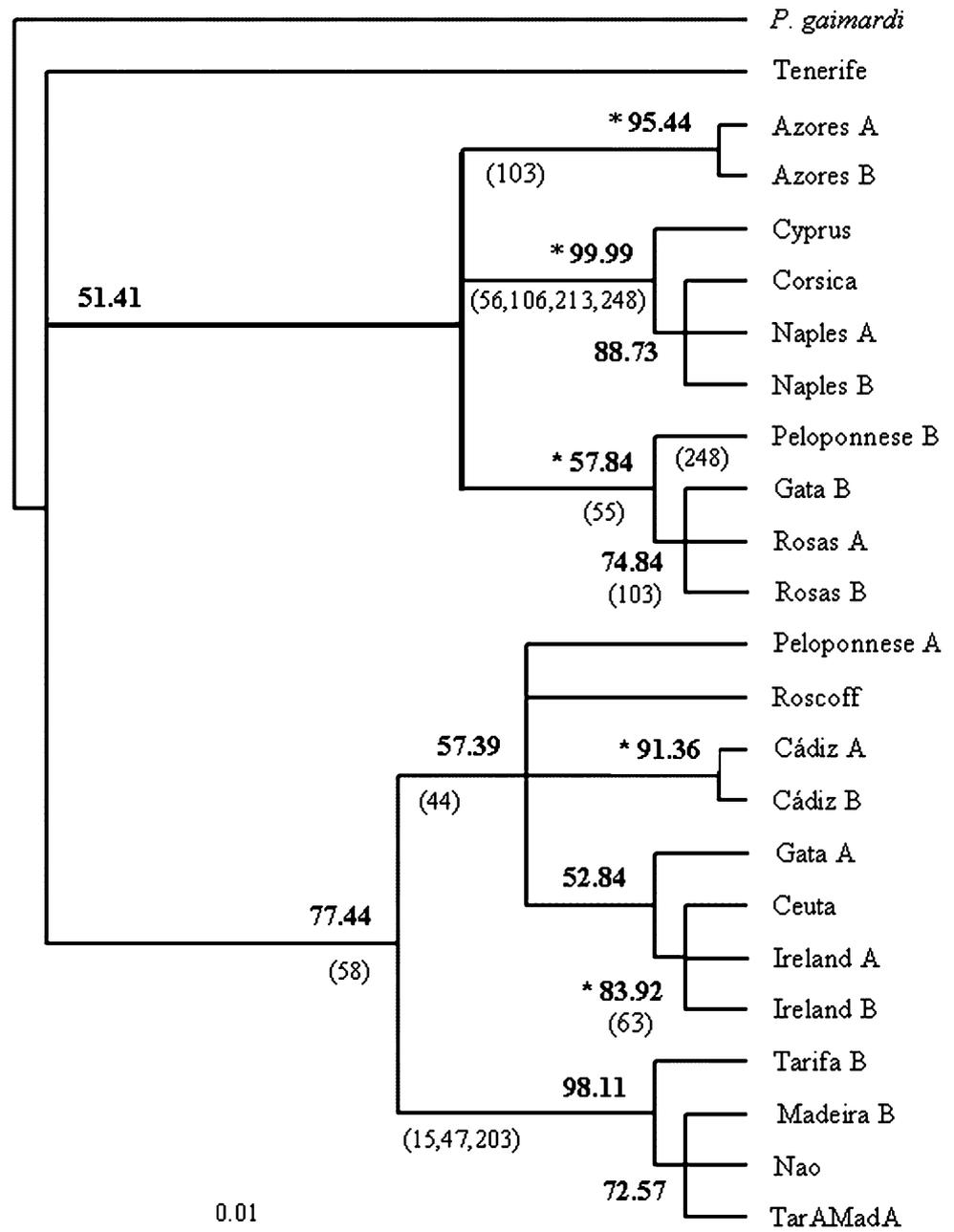
The branch model of PAML showed that there was no positive selection between *P. lividus* and *P. gaimardi*, and alternative models were not significantly better than the one-ratio model (Table 3). Similarly, the tests for particular branches of *P. lividus* did not show evidence for selection (data not shown). However, when applying site models, PAML detected 12 positions potentially subject to positive selection in both regions flanking the core (Table 3). Alternative site models tested (M2a and M8) fit the data significantly better than null models (M1a and M7). Both M2a and M8 identified the same sites as being under selection (BEB: 15Q/R, 44Q/P, 47A/V, 55M/R, 56A/G, 58G/R, 63G/R, 103D/V, 106A/V, 203G/R, 213G/R, 248R/G; Fig. 4 and Table 3), suggesting that positive selection was acting at sites 5' and 3' of the core of bindin in *P. lividus*. Of these 12 sites, 9 (75%) were located within the coding region upstream from the conserved core, whereas only 3 (25%) were located downstream from the core (Fig. 4). These positively selected sites were positioned on the ML tree (Fig. 3) by treating them as shared derived characters of the terminal branch sequences. The branch-site test 2 did not fit the data significantly better than its corresponding neutral model (data not shown). Indeed, estimated  $\omega_2$  was very close to 1 ( $\omega_2 = 1.05811$ ), possibly as a consequence of the high similarity of our sequences to one another. Test 2 was then recomputed considering as the foreground branch each of the branches that accumulate changes in positively selected sites. Again, alternative models did not fit the data better than neutral models for any of the branches tested.

#### Discussion

We isolated bindin in two extant species of the genus *Paracentrotus*. This molecule shows a similar structure to that of other genera studied thus far, including a conserved core flanked by two variable regions and an intron of variable length located at the same conserved position as in other genera. Our study results suggest that despite the high polymorphism detected, no positive selection is acting on particular intraspecific lineages. However, our data also show that some sites of the bindin sequence of *P. lividus* flanking the core are under strong positive selection.

Sequence length of bindin in the genus *Paracentrotus* is approximately the same as in other sea urchins studied to date except *Diadema*, which has a much longer sequence (Zigler and Lessios 2003a). The observed variations in length among the bindin sequences of *Paracentrotus* are mainly caused by differences in the number of repeats of glycine-rich motifs located upstream and downstream from

**Fig. 3** ML tree of mature bindin of *P. lividus* rooted on *P. gaimardi*. Only nodes with >50% bootstrap support are shown in bold. \*Clades supported by >50% bootstrap on the ML tree based exclusively on the intron. A and B represent the two alleles of heterozygotes. TarAMadA represents the only haplotype shared by individuals from two different localities (Tarifa and Madeira). Numbers in brackets indicate branches along which positively selected sites occur according to reconstruction based on parsimony. Scale represents number of substitutions per site



the core region. These regions flanking the conserved core are expected to be important in species-specificity of gamete interactions (Lopez et al. 1993; Biermann 1998). More specifically, a motif GMMXX is thought to play an important role in selectivity in other sea urchins (Minor et al. 1991; Metz and Palumbi 1996). This motif, however, remains invariable between the two species of *Paracentrotus*. The motif YNQG, which is repeated before the stop codon in *Paracentrotus*, has also been detected in *Triploneustes* (Zigler and Lessios 2003b).

The variation of glycine-rich regions is probably caused by slipped-strand mispairing. Indeed, *Paracentrotus* alleles

that are otherwise identical differ in the number of these repeats. Repeats have also been found in other genera of the order Echinoidea (*Echinometra*, *Strongylocentrotus*, *Triploneustes*, and *Lytechinus*). However, they were absent in sampled members of the Cidaroida, Diadematoida, Clypeasteroida, and Arbacioida orders (Metz et al. 1998; Zigler and Lessios 2003a). The only example of a genus outside the Echinoidea order that is known to show extensive repeat structure is the spatangoid *Moiria* (Zigler and Lessios 2003a). There seems to be a correlation between extensive variation in the number of repeats in the region 5' of the core and the presence of positive selection in the

**Table 3** Models of variable  $\omega = d_N/d_S$  ratios across branches and sites<sup>a</sup>

Model type	Ln Neutral	Ln Positive	2(lnL <sub>2</sub> – lnL <sub>1</sub> )	df	P	Positively selected sites
<b>Branch</b>						
Two ratio vs one ratio	–1938.931 ( $\omega = 0.314$ )	–1940.201 ( $\omega_0 = 0.318$ ; $\omega_1 = 0.170$ )	–2.540	1	>0.9	
<b>Site</b>						
M1a vs M2a	–1877.345 ( $p_0 = 0.862$ ; $p_1 = 0.138$ )	–1853.536 ( $p_0 = 0.943$ ; $p_1 = 0.001$ ; $p_2 = 0.055$ ; $\omega_2 = 5.322$ )	47.618	2	<0.001	M2a: (BEB) 15*, 44**, 47*, 55, 56*, 58**, 63, 103*, 106*, 203*, 213*, 248*
M7 vs M8	–1880.502 ( $p = 0.0087$ ; $q = 0.0586$ )	–1853.466 ( $p_0 = 0.946$ ; $p = 0.507$ ; $q = 8.397$ ) ( $p_1 = 0.054$ ; $\omega = 5.447$ )	54.073	2	<0.001	M8: (BEB) 15**, 44**, 47**, 55*, 56**, 58**, 63*, 103**, 106**, 203**, 213**, 248**

<sup>a</sup> Likelihoods for each model and likelihood ratios for each comparison are shown. BEB was implemented in comparisons that involve models M2a and M8. Proportions ( $p$  or  $q$ ) and  $\omega$  are presented in parentheses for each model

\*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$

nonrepeat region (Zigler and Lessios 2004). *Paracentrotus* fits this pattern in showing evidence of both; however, the functional link between these two properties remains elusive.

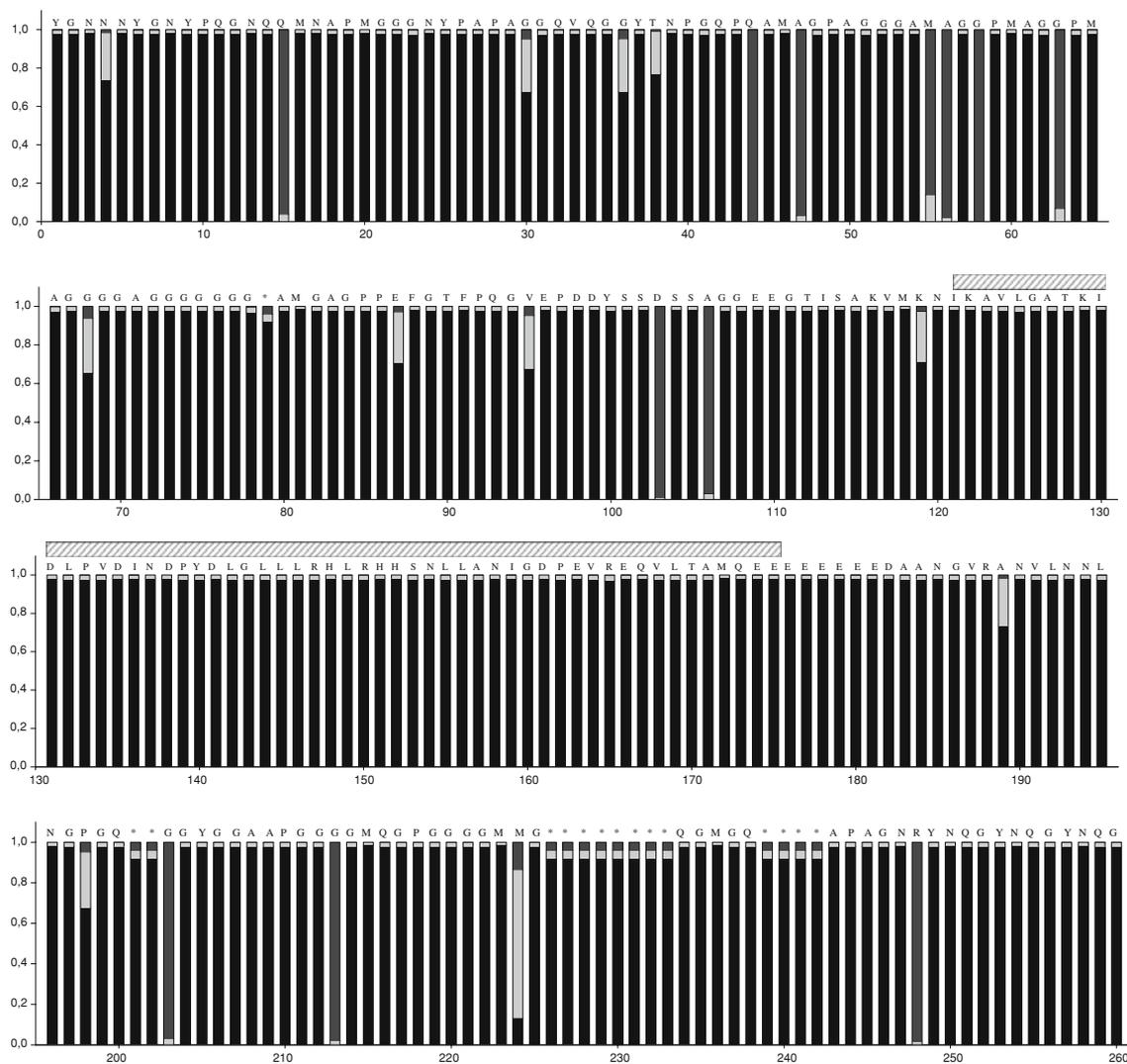
The conserved core, implicated in fusion between bindin and the female egg membrane in vivo, shows no amino acid substitutions, indicating the action of purifying selection in this region (Ulrich et al. 1998, 1999). Higher variation in the sequence, including insertions of glycine-rich repeats, is found outside the core. Such structures have been observed in other proteins with adhesive functions, such as components of spider silk (Guerette et al. 1996) or mussel byssus (Coyne et al. 1997), and in immunorecognition proteins (Hughes 1993). Conversely, these structures are absent in lysin of abalones, which tunnels through the egg membrane in a nonenzymatic and species-selective process (Swanson and Vacquier 1997). In lysin, function is conserved by retaining secondary structure, and species-specificity is maintained by varying selected amino acids scattered through the entire length of the molecule (Lee and Vacquier 1992; Swanson and Vacquier 1997).

No codon use bias has been observed in the genus *Paracentrotus*, with ENC values close to the average detected in other genera by Zigler and Lessios (2003a). Although reasons for the lack of such bias are not clear, this seems to be a common feature in sex-related genes (Civetta and Singh 1998; Ferris et al. 2002).

MK (1991)) test failed to detect positive selection for divergence between the two species, and neither the branch model nor the branch-sites model indicated selection in any specific branch of the tree. These results, however, could well be related to the low power of the statistical tests employed, especially considering that a single sequence of *P. gaimardi* was analyzed and that the sequences obtained in *Paracentrotus lividus* were similar to each other. Further

sampling on *P. gaimardi* would provide a clearer view of relations between these congeneric species. In contrast, the site-specific models of Yang et al. (2000) were able to detect sites on which strong positive selection was acting. The fact that models M2a and M8, compared with their neutral alternatives, identified the same sites as being under selection suggests a clear pattern of positive selection for some amino acids, which is obscured by averages over longer sequences in other tests of selection. These selected sites are clearly associated to some branches of the tree (Fig. 3), further suggesting the action of intraspecific processes on these same codon positions. Nevertheless, branch-site models were unable to detect selection associated to any of these branches, possibly as a consequence of their short length. Thus, bindin evolution in this genus, despite its paucity of species and their lack of sympatry, resembles evolution of bindin in *Echinometra* (Metz and Palumbi 1996; McCartney and Lessios 2004), *Strongylocentrotus* (Biermann 1998), and *Heliocidaris* (Zigler et al. 2003).

What is the source of positive selection on the bindin of *Paracentrotus*? Because it is highly unlikely that *P. lividus* and *P. gaimardi* have ever overlapped geographically, the possibility of reinforcement, suggested from comparison between genera (Metz et al. 1998), can be ruled out. Significant departures of Hardy-Weinberg equilibrium detected for different markers in this species (Calderón et al. 2008, 2009) suggest that mating may not be random in this broadcast spawner. Indeed, variation at the intra-specific level suggests that some sort of sexual selection may be acting on *Paracentrotus*. Bindin divergence may be possibly associated to variation in the egg receptor, which may lead to the establishment of groups of preferential mating within populations. McCartney and Lessios (2004) suggested that strong positive selection on the bindin of *Echinometra lucunter*, but not on its neotropical congeners,



**Fig. 4** Posterior probabilities of codon sites assigned to classes with different values of  $\omega$  calculated using the selection model M2a. Horizontal axis represents amino acid positions of the consensus sequence of *P. lividus*. Black bars represent the posterior probability of assignment to  $\omega_1 = 1.000$ . Light grey bars represent probability of

assignment to  $\omega_0 = 0.051$ . The posterior probability of positive selection (dark grey bars [ $\omega_2 5.3216$ ]) is 1 minus the probability of neutral or purifying selection. The hatched line delimits the 55 amino acid-conserved core. \*Positions where indels were found

might be caused by the fertilization environment encountered by this locally abundant shallow-water species. Levitan and Farrell (2006) experimentally demonstrated that when sperm are limited, common bindin alleles of *Stongylocentrotus franciscanus* achieve most of the fertilization; however, when sperm is abundant, rare alleles are favoured by avoiding polyspermy. *P. lividus* is found at high densities in many communities (Palacín et al. 1998; Boudouresque and Verlaque 2001); therefore, there is potential in this species for sperm competition and for the maintenance of high bindin polymorphism. Different modes of evolution of bindin have been detected in different genera, and the source of selection on bindin and

other reproductive proteins is difficult to determine. This work contributes to identifying the evolutionary mode of bindin in the genus *Paracentrotus*, containing two allopatric species, which seems to evolve subject to positive selection. Comparative data gathered from various genera hold promise of clarifying patterns of existing modes of evolution of bindin in sea urchins.

**Acknowledgements** Special thanks to Axel and Ligia Calderón of STRI-Naos for their help in the laboratory. Thanks to Laura Geyer for providing useful comments on the manuscript. This research was funded by Project No. CTM2007-66635 of the Spanish Government and by general research funds from the Smithsonian Tropical Research Institute.

## References

- Biermann CH (1998) The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Mol Biol Evol* 15:1761–1771
- Boudouresque CF, Verlaque M (2001) Ecology of *Paracentrotus lividus*. In: Lawrence JM (ed) Edible sea urchins: biology and ecology. Elsevier, Amsterdam, The Netherlands, pp 177–216
- Brendel V, Bucher P, Nourbakhsh I, Blaisdell BE, Karlin S (1992) Methods and algorithms for statistical analysis of protein sequences. *Proc Natl Acad Sci USA* 89:2002–2006
- Calderón I, Giribet G, Turon X (2008) Two markers and one history: phylogeography of the edible common sea urchin *Paracentrotus lividus* in the Lusitanian region. *Mar Biol* 154:137–151
- Calderón I, Turon X, Pascual M (2009) Isolation of 9 nuclear microsatellites in the common Mediterranean sea urchin *Paracentrotus lividus* (Lamarck). *Mol Ecol Resour*. doi:10.1111/j.1577.0998.2009.02585.x
- Civetta A, Singh RS (1998) Sex-related genes, directional sexual selection, and speciation. *Mol Biol Evol* 15:901–909
- Coyne JA (1992) Genetics and speciation. *Nature* 35:511–515
- Coyne JA, Orr HA (2004) Speciation. Sinauer, Sunderland, MA
- Coyne KJ, Qin XX, Waite JH (1997) Extensible collagen in mussel byssus—a natural block copolymer. *Science* 277:1830–1832
- Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP (1999) Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol Biol Evol* 16:372–382
- Duran S, Palacín C, Becerro MA, Turon X, Giribet G (2004) Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). *Mol Ecol* 13:3317–3328
- Ferris PJ, Armbrust EV, Goodenough EH (2002) Genetic structure of the mating-type locus of *Chlamydomonas reinhardtii*. *Genetics* 160:181–200
- Frohman MA, Dush MK, Martin GR (1988) Rapid production 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, 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
- Geyer LB, Palumbi SR (2003) Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* 57:1049–1060
- Glabe CG, Lennarz WJ (1979) Species-specific sperm adhesion in sea urchins: a quantitative investigation of bindin-mediated egg agglutination. *J Cell Biol* 83:595–604
- Glabe CG, Vacquier VD (1977) Species specific agglutination of eggs of bindin isolated from sea urchin sperm. *Nature* 267:822–824
- Guerette PA, Ginzinger DG, Weber BH, Gosline JM (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* 272:112–115
- Hellberg ME, Vacquier VD (1999) Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Mol Biol Evol* 16:839–848
- Howard DJ (1999) Conspecific sperm and pollen precedence and speciation. *Annu Rev Ecol Syst* 30:109–133
- Hughes AL (1993) Coevolution of immunogenic proteins of *Plasmodium falciparum* and the host's immune system. In: Takahata N, Clark AG (eds) Mechanisms of molecular evolution. Sinauer, Sunderland, MA, pp 109–127
- Jobb G (2007) TREEFINDER February 2007 version. Munich, Germany. Available at: <http://www.treefinder.de>. Accessed June 2008
- Lee Y, Vacquier VD (1992) The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol Bull* 182:97–104
- Lee Y, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol Biol Evol* 12:231–238
- Lessios HA (2007) Reproductive isolation between species of sea urchins. *Bull Mar Sci* 81:191–208
- Lessios HA (2008) The Great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annu Rev Ecol Evol Syst* 39:63–91
- Levitan DR, Farrell DL (2006) Selection on gamete recognition proteins depends on sex, density and genotype frequency. *Science* 312:267–269
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. *J Mol Evol* 36:96–99
- Lopez A, Miraglia SJ, Glabe CG (1993) Structure/function analysis of the sea urchin sperm adhesive protein bindin. *Dev Biol* 156:24–33
- Martin DP, Williamson C, Posada D (2005) RDP2: Recombination detection and analysis from sequence alignments. *Bioinformatics* 21:260–262
- McCartney MA, Lessios HA (2004) Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Mol Biol Evol* 21:732–745
- 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
- Metz EC, Kane RE, Yanagamachi H, Palumbi SR (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
- Metz EC, Gómez-Gutiérrez G, Vacquier VD (1998) Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Mol Biol Evol* 15:185–195
- Minor JE, Fromson DR, Britten RJ, Davidson EH (1991) Comparison of bindin proteins of *Strongylocentrotus franciscanus*, *S. purpuratus* and *Lytechinus variegatus*: Sequences involved in the species-specificity of fertilization. *Mol Biol Evol* 8:681–795
- Mortensen T (1943) A monograph of the Echinoidea. Camarodonta, vol. III. C. A. Reitzel, Copenhagen, Denmark
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Mol Biol Evol* 3:418–426
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936
- Palacín C, Turon X, Ballesteros M, Giribet G, López S (1998) Stock evaluation of three littoral echinoid species on the Catalan Coast (North-Western Mediterranean). *Pubblicazione della Stazione Zoologica di Napoli: Mar Ecol* 19:163–177
- Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annu Rev Ecol Syst* 25:547–572
- Palumbi SR (1998) Species formation and the evolution of gamete recognition loci. In: Howard DJ, Berlocher SH (eds) Endless forms: species and speciation. Oxford University Press, Oxford, UK, pp 271–278

- Palumbi SR (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc Natl Acad Sci USA* 96:12632–12637
- Pamilo P, Bianchi O (1993) Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Mol Biol Evol* 10:271–281
- Park DJ, Pask AJ, Renfree MB, Graves JA (2003) 3' RACE walking along a large cDNA employing tiered suppression PCR. *Biotechniques* 34(750–752):754–756
- Posada D (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucl Ac Res* 34:700–703
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Rozas J, Sánchez del Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Swanson WJ, Vacquier VD (1995) Extraordinary divergence and positive Darwinian selection in a fusogenic protein coating the acrosomal process of abalone spermatozoa. *Proc Natl Acad Sci USA* 92:4957–4961
- Swanson WJ, Vacquier VD (1997) The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proc Natl Acad Sci USA* 94:6724–6729
- Swanson WJ, Vacquier VD (2002) Rapid evolution of reproductive proteins. *Nature Rev Genet* 3:137–144
- Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA* 98:2509–2514
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Ulrich AS, Otter M, Glabe CG, Hoekstra D (1998) Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. *J Biol Chem* 273:16748–16755
- Ulrich AS, Tichelaar W, Forster G, Zschornig O, Weinkauff S, Meyer HW (1999) Ultrastructural characterization of peptide-induced membrane fusion and peptide self-assembly in the lipid bilayer. *Biophys J* 77:829–841
- Vacquier VD (1998) Evolution of gamete recognition proteins. *Science* 281:1995–1998
- Vacquier VD, Moy GW (1977) Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. *Proc Natl Acad Sci USA* 74:2456–2460
- Vacquier VD, Swanson WJ, Hellberg ME (1995) What have we learned about sea urchin sperm bindin? *Dev Growth Differ* 37:1–10
- Vacquier VD, Swanson WJ, Lee Y (1997) Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *J Mol Evol* 44(Suppl 1): S15–S22
- Villablanca FX, Roderick GK, Palumbi SR (1998) Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol Ecol* 7:547–560
- Wright F (1990) The “effective number of codons” used in a gene. *Gene* 87:23–29
- Yang Z (2007) PAML4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Nielsen R, Goldman N, Krabbe Pedersen AM (2000) Codon substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Yang Z, Wong WSW, Nielsen R (2005) Bayes Empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118
- Zhang J, Kumar S, Nei M (1997) Small-sample tests of episodic evolution: a case study of primate lysozymes. *Mol Biol Evol* 14:1335–1338
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479
- Zigler KS, Lessios HA (2003a) 250 million years of bindin evolution. *Biol Bull* 205:8–15
- Zigler KS, Lessios HA (2003b) Evolution of bindin in the pantropical sea urchin *Tripneustes*: Comparisons to bindin of other genera. *Mol Biol Evol* 20:220–231
- Zigler KS, Lessios HA (2004) Speciation on the coasts of the new world: Phylogeography and the evolution of bindin in the sea urchin genus *Lytechinus*. *Evolution* 56:1225–1241
- Zigler KS, Raff EC, Popodi E, Raff RA, Lessios HA (2003) Adaptive evolution of bindin in the genus *Heliocidaris* is correlated with the shift to direct development. *Evolution* 57:2293–2302