EVOLUTION & DEVELOPMENT 17:1, 92–108 (2015)
DOI: 10.1111/ede.12108

Evolution of gamete attraction molecules: evidence for purifying selection in speract and its receptor, in the pantropical sea urchin *Diadema*

Santosh Jagadeeshan,*[1] Simon E. Coppard, and Harilaos. A. Lessios

Smithsonian Tropical Research Institute, Box 0843–03092, Balboa, Panama

*Author for correspondence (e-mail: sanjagster@gmail.com)

[1]Present address: Department of Biology, McMaster University, Hamilton, Ontario, Canada

SUMMARY Many free-spawning marine invertebrates, such as sea urchins, lack any courtship or assortative mating behavior. Mate recognition in such cases occur at the gametic level, and molecules present on the sperm and egg are major determinants of species-specific fertilization. These molecules must also coevolve in relation to each other in order to preserve functional integrity. When sea urchins release their gametes in seawater, diffusible molecules from the egg, termed sperm-activating peptides, activate and attract the sperm to swim toward the egg, initiating a series of interactions between the gametes. Although the compositions and diversity of such sperm-activating peptides have been characterized in a variety of sea urchins, little is known about the evolution of their genes. Here we characterize the genes encoding the sperm-activating peptide of the egg (speract) and its receptor on the sperm, and examine their evolutionary dynamics in the sea urchin genus *Diadema,* in the interest of determining whether they are involved in reproductive isolation between the species. We found evidence of purifying selection on several codon sites in both molecules and of selectively neutral evolution in others. The diffusible speract peptide that activates sperm is invariant across species, indicating that *Diadema* egg peptides do not discriminate between con- and hetero-specific sperm at this stage of the process. Speract and its receptor do not contribute to reproductive isolation in *Diadema*.

INTRODUCTION

Sea urchins are broadcast spawners; males and females release their gametes into the sea, where molecules present on the gametes mediate species-specific sperm and egg interactions, and fertilization (Vacquier 1998). Upon release, sperm acquire motility and are attracted by diffusible molecules released from the egg (Miller 1985). Once they enter the egg jelly, sperm undergo changes in motility and respiration (Ohtake 1976; Garbers 1989), and as they pass through the egg jelly, sulfate polysaccharides bind to receptors in the sperm, inducing the acrosomal reaction (Dan 1967; Vacquier and Moy 1997). Finally, sperm binds and fuses with the Vitelline Envelope (VE), and releases its DNA into the cytoplasm of the egg (Vacquier 1998). Several molecules in the sperm and the egg are known to play crucial roles in mediating each of these steps (Vacquier 1998; Neill and Vacquier 2004; Vieira and Miller 2006; Hirohashi et al. 2008). There is considerable interest in the evolution of these molecules, because their divergence can potentially cause reproductive isolation between species (Swanson and Vacquier 2002; Vacquier and Swanson 2007). Although identifying such divergent molecules is important for understanding how species-specific fertilization occurs and speciation occurs, characterizing gametic molecules that are conserved is equally important to shed light on the ways in which functional integrity is maintained (Lessios 2007).

Gametic proteins that have been identified in free spawning organisms as evolving rapidly mediate processes that occur once the sperm has entered the egg jelly, and when sperm binds to the vitelline envelope (Swanson and Vacquier 2002; Vacquier and Swanson 2007; Lessios 2011). In the egg jelly, fucose sulfate polymers (FSPs) that interact with sperm receptors for egg jelly (*SuRej*) to induce acrosome reaction (Neill and Vacquier 2004) have species-specific structural differences (Biermann et al. 2004). Genes encoding *SuRej* evolve under positive selection in the sea urchin genus *Strongylocentrotus* (Mah et al. 2005). In at least two genera of abalones, the sperm proteins *Lysin* and its receptor on the VE (*VERL*), mediating sperm-VE binding, coevolve rapidly under positive selection (Swanson and Vacquier 1998; Galindo et al. 2003; Clark et al. 2009; Hellberg et al. 2012). In sea urchins, the sperm protein *bindin*, which binds to receptors on the VE, has been shown to be under positive selection in some (but not all) sea urchin genera (Zigler and Lessios 2003; Lessios 2007, 2011) and to evolve rather rapidly in relation to other molecules (Lessios and Zigler 2012). The *egg receptor for bindin* (*EBR1*) has been characterized in

only two species of sea urchins and one starfish (Kamei and Glabe 2003; Hart 2013). The sea star *bindin* and its receptor (*Obi1*) have been shown to experience diversifying selection in some populations (Sunday and Hart 2013; Hart et al. 2014). Although these studies provide valuable insights into the evolutionary dynamics of sperm-egg binding molecules, relatively little is known about the molecular evolution of sperm and egg molecules involved in stages prior to sperm contact with the egg jelly. As a result, whether or not reproductive barriers can be established at the sperm-egg attraction stage in free spawning animals remains unclear.

Sperm-activating peptides and their receptors

Eggs of marine invertebrates produce diffusible molecules with chemoattractant properties to guide sperm (Lillie 1912; Miller 1985; Kaupp et al. 2006). Since the first characterization of *resact* from *Arbacia punctulata* (Hansborough and Garbers 1981), sperm-activating peptides (SAPs), or speracts, and their receptors on the sperm, have been intensively studied to understand the basis of chemotaxis (Kaupp 2012). SAPs are small diffusible egg jelly peptides (approximately 9–15amino acids long) that can activate the sperm at a distance from the egg (Miller 1985). When a speract molecule binds to the speract-receptor, which is localized on the flagellum (Cardullo et al. 1994), it induces cellular activation, increasing respiration and flagellar motility of the sperm (Kopf et al. 1979; Trimmer and Vacquier 1986). In seawater, the diffusible speract molecules provide a "speract-gradient", or a chemoattractant "pathway" to which sperm respond by altering their usual circular swimming trajectories (Kaupp et al. 2008). Essentially, as sperm sample the speract gradient, their flagellar modulation and turning motions are periodically stimulated as they contact speract molecules, causing sperm to "turn" and "run" toward the source of the gradient (Kaupp et al. 2008; Guerrero et al. 2010). Much of this model of chemotactic behavior comes from the studies of *resact*, a protein that is structurally dissimilar, but functionally analogous to speract, and is unique to the sea urchin genus *Arbacia* (Hansborough and Garbers 1981). *Resact* was initially the only speract-like molecule demonstrated to induce a chemotactic response in sperm, causing sperm to accumulate at its source (Ward et al. 1985). Recently, Guerrero et al. (2010) have demonstrated that speract of *Lytechinus pictus* induces similar chemotactic response in conspecific sperm.

Over 75 SAPs have been isolated from the egg jelly of approximately 15 sea urchin species. There is considerable variation in the number and amino acid composition of speract peptides across sea urchin genera (Suzuki 1995). SAPs had been reported to function in a "species-specific" manner — *A. punctulata* resact has no effect on *S. purpuratus* or *L. pictus* sperm, and speracts of the latter species have no effect on *A. punctulata* sperm (Hathaway 1963; Suzuki et al. 1984; Ward et al. 1985). However, a comprehensive review of the nature of "species-specificity" of SAPs (Suzuki and Yoshino 1992; Suzuki 1995) suggests that discriminatory abilities of speract appear to be "genus-specific or order-specific". Relatively little is known about the receptor for speract, primarily because the genes that code for it have only been sequenced in two sea urchin species: *S. purpuratus* (Dangott et al. 1989) and *Hemicentrotus pulcherrimus* (Shimizu et al. 1994). To date, only one study addresses the evolutionary dynamics of genes-encoding speract- and speract receptor-like molecules. Nakachi et al. (2008) have shown that DNA sequences of sea star a*sterosaps* and their receptors on the sperm are conserved across species of the subfamily *Asteriinae*. Here, we present evidence regarding the molecular evolution of speract and its receptor in species of the pantropical sea urchin genus *Diadema*. The *Diadema* speract peptide, first characterized in *Diadema setosum* (Yoshino et al. 1990), is a short, nine amino acid long peptide and has been termed SAP IV as it is structurally unique to the order Diadematoida (Suzuki and Yoshino 1992).

MATERIALS AND METHODS

Specimen collections

Diadema antillarum was collected off the Smithsonian Tropical Research Institute's Punta Galeta and Bocas del Toro marine stations in the Caribbean. *Diadema mexicanum* was collected at Taboguilla Island and Saboga Island, off the Pacific coast of Panama. *Diadema paucispinum* was collected at Moku Ola (Coconut Island), Hawaii. *Diadema savignyi* and *Diadema setosum* were both collected at Namatakula, Coral Coast, Fiji Islands and off Olango Island, Philippines.

Identification of speract and its receptor from *D. antillarum* and *D. mexicanum* cDNA libraries

Speract and its receptor were initially identified by screening ovary-specific and testis-specific cDNA libraries of *D. antillarum* and *D. mexicanum*. Full-length cDNA sequences from other *Diadema* species were later acquired through polymerase chain reaction (PCR) amplifications, as well as next-generation sequencing (NGS) (see below). To construct the cDNA libraries, mRNA extracted from freshly dissected testis and ovary (identified by observing the presence of eggs or sperm under a light microscope) was used as a template to construct normalized cDNA libraries using the CloneMiner cDNA construction kit (Life Technologies, Carlsbad, CA, USA). Clones were partially sequenced using the M13 forward primer (5′-GTAAAACGACGGCCAG-3′) within the pDONR221 vector (Life Technologies) to obtain Expressed Sequence Tags (ESTs). Speract and its receptor were identified through tBLASTx searches of ESTs against the *S. purpuratus* genome (Sodergren et al. 2006). Complete cDNA sequences from *D. antillarum* and *D. mexicanum* were then obtained through

PCR amplification using M13 forward and reverse primers (M13 Rev 5′-CAGGAAACAGCTATGAC-3′).

PCR amplification of speract and its receptor from *D. paucispinum*, *D. savignyi* and *D. setosum*

Gonads collected from *D. paucispinum*, *D. savignyi* and *D. setosum* and stored in RNALater were used for RNA extractions using the PURELINK RNA extraction kit (Life Technologies). Approximately 4 ug of total RNA was used as a template to construct cDNA libraries using a Mint Kit (Evrogen, Moscow, Russia). These libraries were normalized using the Trimmer-Direct normalization kits (Evrogen) to increase the likelihood of amplifying transcripts of low abundance. Five to ten nanograms of cDNA were used in PCR amplifications of speract using primers designed from sequences of *D. antillarum* and *D. mexicanum*, 5′ GAAGGTCATCGCTGCAGTTCTTCT '3, forward and 5′ TCTCCTCGAGGGATCAGCAGAC 3′, reverse. To amplify the speract receptor, the following primers were used—forward 5′ GGCAAAAGACATGATGGCAG 3′, and, reverse 5′ TTGTCAGGGGCTTAGGCAGCAG 3′. Purified PCR products were sequenced using the BigDye Terminator v3.1 system (Applied Biosciences) on an ABI 3130 sequencer (Applied Biosystems, Life Technologies, Carlsbad, CA, USA).

Speract and speract receptor sequences from Illumina sequencing

We also obtained additional sequences of speract and its receptor from *D. setosum*, *D. paucispinum* and *D. savignyi* through NGS. About 7 ug of total RNA from gonads of *D. setosum*, *D. paucispinum* and *D. savignyi* (a single specimen from each species) were used as templates for library preparation and Illumina Hi-Seq sequencing. RNA quality prior to sequencing was analyzed on a BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) to ensure RIN $>$8.0. Library preparation, Illumina Hiseq2000 transcriptome sequencing to generate 100bp paired-end reads, de novo assembly using Trinity (Grabherr et al. 2011) and reference mapping to the *S. purpuratus* genome were done at sequencing facilities at Genome Quebec, McGill University. After filtering adaptors and low-quality reads, transcriptome sequences with coverages of 76 million reads for *D. savignyi* and *D. paucispinum* and 72 million reads for *D. setosum* were used to search for sequences of speract and its receptor. Full-length coding sequences of speract and its receptor were identified through tBLASTx searches (E value = 1e-5) of assembled contigs to the *S. purpuratus* genome, as well as through BLASTn searches of PCR sequences from *Diadema* species to their respective transcriptome databases, using Galaxy (Goeck et al. 2010). NGS sequences, thus served to confirm sequences obtained from Sanger sequencing methods. Species sequences obtained from NGS are labeled *D. 'species'ngs* in this article. All sequences have been deposited in GenBank (accession numbers for speract: KJ882342 - KJ882358 and for speract receptor: KJ882359 - KJ882375).

Sequence editing and alignments

Speract and its receptor sequenced by ABI 3130 were manually edited in Sequencher 4.6 (Gene Codes Corporation). Sequences alignments were done using MacClade 4.08a (Maddion and Maddison 2005), MUSCLE (Multiple Sequence Comparison by Log-Expectation) (Edgar 2004), RevTrans (Wernersson and Pedersen 2003) and PRANK (Loytynoja and Goldman 2010).

Phylogenetic analyses

Nucleotide alignments were used to construct phylogenetic trees using Bayesian (MrBayes) (Ronquist and Huelsenbeck 2003) and Maximum Likelihood (RAxML) methods (Stamatakis et al. 2005). The best model for phylogenetic analyses was determined using jmodeltest2 (Darriba et al. 2012) based on the AIC criterion (Akaike 1974). The Hasegawa, Kishino and Yano model (Hasegawa et al. 1985) was selected as the best-fit evolutionary model for the speract receptor (HKY + I; $I = 0.8332$), and a transition/transversion model TIM + I was selected as best fit for speract, where $I = 0.7062$.

Sequence analyses

Intraspecific measures of nucleotide diversity, Tajima's D (Tajima 1989), Fu's F as well as Fu and Li's D* and F* (Fu and Li 1993; Fu 1997) tests of neutrality, were carried out in DNAsp5.10 (Librado and Rozas 2009). Intraspecific estimates of the proportions of synonymous (d_S) and nonsynonymous (d_N) substitutions per site, of speract and its receptor in *Diadema* were computed using the methods of Pamilo and Bianchi (1993) and Li (1993), as implemented in MEGA6 (Molecular Evolutionary Genetic Analysis) (Tamura et al. 2013). Interspecific pairwise estimates of synonymous (d_S) and nonsynonymous (d_N) substitutions per site, as well as ω (d_N/d_S), were estimated following the maximum likelihood model of Yang and Nielsen (2000), using the program *Yn.00* contained in the software package PAML4.7a (Phylogenetic Analyses by Maximum Likelihood) (Yang 2007).

Tests of selection

The Single Break Point (SBP) and Genetic Algorithm Recombination Detection (GARD) tests (Kosakovsky Pond et al. 2006) implemented in HyPhy (Hypothesis testing using Phylogenies, Kosakovsky Pond et al. 2005) were used to check for recombination in the molecules. MEGA 6 was used for Fisher's exact tests of neutrality, and to construct Neighbor Joining (NJ) trees that formed the basis of subsequent Maximum

Likelihood (ML) tests of selection using PAML. The program *codeml* implemented in PAML was used to test whether speract and speract receptor contained molecular signatures of having evolved under positive selection. Variation in ω across codon sites was tested using the site-specific models implemented in PAML (Yang and Nielsen 2000; Yang et al. 2000). Likelihood ratio tests (LRT) were used to compare whether amino acid site substitutions in speract and speract receptor conformed to the nearly neutral null models M1a and M7 (β) (as inferred by codon sites with $0 < \omega < 1$), versus models of selection, M2a and M8 (β & ω) (as inferred by codon sites with $\omega > 1.0$). All LRTs in the site-specific models have two degrees of freedom (Yang 2007). We also performed branch site tests of selection (Yang and Nielsen 2002; Yang et al. 2005) to determine whether any of the sampled sequences of speract and its receptor showed evidence of positive selection in specific lineages. In the branch site model, the selection model (Model A, selection) assumes two ω ratios ($0 < \omega_0 < 1$, $\omega_1 = 1$) for all background branches (ω_{back}) but allows user-specified foreground branches of interest (ω_{fore}), to have an additional $\omega_2 > 1.0$. The selection model is compared with a null model (Model A, null) that fixes the foreground ω_2 at 1.0. LRT tests for this comparison have one degree of freedom. We also used the HyPhy software package, which implements a variety of methods to test for evidence of selection (Kosakovsky Pond et al. 2005). Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Random Effects Likelihood (REL) (Kosakovsky Pond and Frost 2005) and Fast Unconstrained Bayesian Approximate method (FUBAR) (Murrell et al. 2013) methods were used to estimate rates of synonymous substitutions per site (d_S, or α) and nonsynonymous substitutions per site (d_N, or ß), and to find codons with signatures of positive or negative selection. The Branch-site REL test (Kosakovsky et al. 2011) was used to identify sites with evidence of pervasive positive selection in specific lineages, and MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) was used to identify codon sites that may have experienced episodic events of positive selection (temporal variation in selection across the tree) in specific lineages (Kosakovsky et al. 2011).

Protein domain predictions

Signal, transmembrane and extra cellular domains were identified using the Simple Modular Architecture Research Tool (SMART), a web-based tool that implements hidden Markov-Models for the identification and annotation of protein architecture (Ponting et al. 1999; Letunic et al. 2012), as well as SignalP4.0 (Petersen et al. 2011), which uses neural networking to specifically identify signal peptides. Complete protein sequences of speract and its receptor were used as inputs on the SMART and SignalP web-servers. Only domains and motifs annotated with high confidence (E-values < 0.001) were chosen.

Comparing nucleotide and protein divergence across sea urchin genera

Sequences of speract (GenBank Accession number NM_214606) (Ramarao et al. 1990) and its receptor (NM_214607) (Dangott et al. 1989) from *S. purpuratus* and *Pseudocentrotus depressus* (Yamano et al. unpublished, AB594707), as well as speract of *Hemicentrotus pulcherrimus* (D38490) (Kinoh et al. 1994) and its receptor (D21101) (Shimizu et al. 1994) were retrieved from the National Center for Biotechnological Information (NCBI). We also retrieved a partial coding sequence of speract of *Eucidaris tribuloides* (Order Cidaroida) using *Diadema* sequences to search against *E. tribuloides* RNA sequence database created by the Center for Computational Regulatory Genomics at Caltech (user interface at www.SpBase.org/ET/). Note that the *E. tribuloides* speract partial sequence is missing the carboxy terminal exon(s), which code(s) for the speract peptide(s). In order to compare the extent of speract gene and protein divergence across genera, amino acid sequences of *S. purpuratus, H. pulcherrimus, E. tribuloides* and *D. setosum* were first aligned using MUSCLE (Edgar 2004). The amino acid alignment was then used as a template to obtain a reverse translated alignment of nucleotide sequences using the software RevTrans (Wernersson and Pedersen 2003). We also used the PRANK alignment software (Loytynoja and Goldman 2010) on webPRANK (http://www.ebi.ac.uk/goldman-srv/webPRANK/) specifically to verify the alignment of the speract peptide domains between genera. Prank makes use of phylogenetic information to align sequences and provides posterior probability support of alignment of each residue. ML trees of speract nucleotide sequences from members of different genera, constructed using RAxML were used as a guide tree for PRANK. Pairwise estimates of (d_S), (d_N) and ω were obtained by the method of Yang and Nielsen (2000), using the program *Yn.00* contained in PAML4.7a.

RESULTS

The structure of speract and its receptor in *Diadema*

We obtained the complete coding sequence (cds) of speract and its receptor from the Caribbean *D. antillarum*, the Tropical Eastern Pacific *D. mexicanum*, and the Indo-Pacific *D. savignyi, D. setosum* and *D. paucispinum*. The *Diadema* speract cds is 603 bp long with no length variation between species. This cds is 288bp shorter than the *S. purpuratus* speract cds, and aligns to 9 of the 24 exons annotated in the complete *S. purpuratus* speract gene sequence (Tu et al. 2012). The differences in length primarily arise from the additional exons that code for multiple speract peptides in *S. purpuratus* (Ramarao et al. 1990). The receptor for speract in *Diadema* is a much larger molecule with a cds length of approximately 1.6 kb, with slight length variations between species. *Diadema setosum* has the longest sequence of

1629 bp (542aa) with two additional codons in the 3′ region of the gene, missing in all other *Diadema* species. *Diadema savignyi, D. mexicanum* and *D. antillarum* all share a cds of 1623 bp (540aa). *Diadema paucispinum* has the shortest cds of 1617 bp (538aa), missing two codons at the 5′ end present in the former three *Diadema* species. The cds of *Diadema*'s receptor for speract is 18–30 bp longer than that of *S. purpuratus*, but aligns to all 11 exons annotated in the complete gene sequence of the *S. purpuratus* receptor for speract (Dangott et al. 1989; Tu et al. 2012).

Gene genealogies of speract and its receptor in *Diadema*

We reconstructed gene genealogies based on nucleotide sequences of speract and its receptor in *Diadema* using Maximum Likelihood (RAxML, Stamatakis et al. 2005) and Bayesian (2003) methods. The two approaches produced similar tree topologies for both speract and its receptor (Fig. 1). Tree topologies of both speract and its receptor agree with the mitochondrial gene tree topology for *Diadema* species (Lessios et al. 2001). The Indo-Pacific species *D. setosum* (diverged approximately 10–7 Ma from the rest of species used in this study, Lessios et al. 2001), is placed as a sister lineage to the more recently diverged species pairs of *D. mexicanum* (Tropical Eastern Pacific) and *D. antillarum* (Caribbean), diverged approximately 3 Ma (Lessios et al.,2001), and the Indo-Pacific *D. paucispinum* and *D. savignyi,* which have diverged from each other and from *D. antillarum* <2 Ma (Lessios et al. 2001). We found no evidence of recombination in both molecules using the Single Breakpoint (SBP) and Genetic Algorithm Recombination Detection (GARD) tests (Kosakovsky et al. 2006) implemented in HyPhy (Kosakovsky Pond et al. 2005) that would affect phylogenetic inferences or detection of signatures of selection in the molecules.

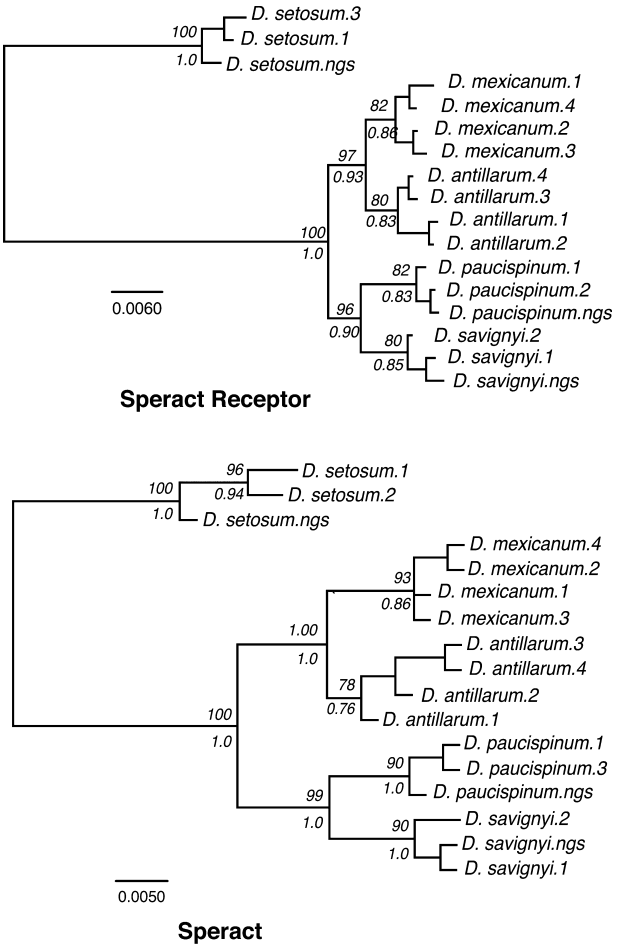

Fig. 1. Phylogenetic trees of speract and its receptor in Diadema, constructed using Bayesian and Maximum Likelihood methods. Trees are unrooted. Posterior probabilities from Bayesian analyses are indicated below nodes and bootstrap support values from Maximum Likelihood analyses are above nodes. Numbers next to species represent specimen identification numbers. ngs - sequences obtained from next generation sequencing..

Molecular evolution of speract and its receptor

We analyzed the rates and patterns of nucleotide substitutions to determine the mode of evolution in speract and its receptor. Intraspecific estimates of nucleotide diversity and the proportions of nonsynonymous substitutions (d_N) to synonymous substitutions (d_S) are low in both speract and its receptor (Table 1). Tajima's D (Tajima 1989), Fu's F as well as Fu and Li's D* and F* tests of neutrality (Fu and Li 1993; Fu 1997) did not reject a model of neutral evolution for either of these molecules (Table 1). Interspecific pairwise estimates of the ratio of replacement and silent substitutions, ω (d_N/d_S), are less than 1.0 in all interspecific pairwise comparisons in both speract and its receptor (Table 2). None of these comparisons (Tables 1 and 2) showed any evidence of departures from neutrality according to Fisher's exact tests ($P > 0.05$ in all cases).

Estimates of ω averaged over the entire gene are insufficient to detect signatures of selection that may in fact be scattered among individual codon sites across these reproductive molecules, or a few codons localized in specific regions of the genes (e.g., see Hughes and Nei 1989). We, therefore, performed maximum likelihood tests of selection to determine if indeed specific codon sites in speract and its receptor show signatures of directional selection. We first implemented the sites model (Nielsen and Yang 1998; Yang et al. 2000) to compare the fit of nearly neutral null models, M0, M1a and M7, and the fit of models M2a and M8, which detect positive selection in the data. None of the log likelihood ratio tests between the selection and null models were significant (Table 3), and none of the few codon sites that were indicated as having elevated ω in speract and its receptor molecule had posterior

Table 1. Intraspecific estimates of nucleotide diversity (π), tests of neutrality, and proportions of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per non-synonymous site (d_N) in speract and its receptor in *Diadema*

Species	N	π	Tajima's[1]D	Fu's Fs[2]	Fu & Li's[2] D^*	Fu & Li's[2] F^*	d_S(± SE)	d_N(± SE)	d_N/d_S
Speract									
D. antillarum	4	0.003	1.917	1.919	1.656	1.791	0.027 (0.010)	0.004 (0.002)	0.022 (0.010)
D. mexicanum	4	0.003	1.116	1.960	1.095	1.301	0.020 (0.011)	0.008 (0.003)	0.027 (0.011)
D. savignyi	3	0.001	–	2.022	1.414	1.412	0.010 (0.006)	0.002 (0.001)	0.008 (0.004)
D. paucispinum	3	0.002	–	2.638	1.421	1.421	0.017 (0.010)	0.003 (0.001)	0.014 (0.010)
D. setosum	3	0.006	–	1.272	1.486	1.492	0.049 (0.018)	0.007 (0.002)	0.042 (0.019)
Speract receptor									
D. antillarum	4	0.003	2.198	−0.439	2.197	2.163	0.006 (0.002)	0.001 (0.001)	0.005 (0.002)
D. mexicanum	4	0.003	1.198	−0.399	1.745	1.824	0.010 (0.006)	0.001 (0.001)	0.009 (0.004)
D. savignyi	3	0.002	–	0.308	0.743	0.760	0.011 (0.006)	0.001 (0.001)	0.010 (0.005)
D. paucispinum	3	0.001	–	−0.341	1.604	1.600	0.006 (0.004)	0.001 (0.001)	0.005 (0.004)
D. setosum	3	0.004	–	0.855	0.087	0.697	0.016 (0.006)	0.003 (0.001)	0.015 (0.006)

SE, standard error. [1]Tajima'D (Tajima 1989) – $P > 0.10$ in all cases. [2]Fu's F, Fu and Li's D^* and F^* (Fu and Li 1993; Fu 1997) – $P > 0.10$ in all cases

probabilities >90% (Table 3). We also implemented the branch-site models tests (Yang et al. 2005; Zhang et al. 2005) to determine if codon sites were evolving under selection in specific lineages, but found no evidence of positively selected sites in any of the sequences of speract and its receptor that were sampled in this study (Table 4).

We also used HyPhy (Kosakovsky Pond et al. 2005), which contains several codon-based likelihood methods to find

Table 2. Proportions of synonymous substitutions per synonymous site (d_S), nonsynonymous substitutions per nonsynonymous site (d_N) and ω (d_N/d_S), in speract and its receptor between *Diadema* species

Species pairs	d_S (± SE)	d_N (± SE)	$\omega(d_N/d_S)$
Speract			
D. antillarum vs. D. mexicanum	0.029 (0.015)	0.003 (0.002)	0.072
D. antillarum vs. D. paucispinum	0.109 (0.031)	0.013 (0.005)	0.106
D. antillarum vs. D. savignyi	0.124 (0.034)	0.015 (0.006)	0.124
D. antillarum vs. D. setosum	0.104 (0.027)	0.009 (0.002)	0.089
D. mexicanum vs. D. paucispinum	0.094 (0.028)	0.009 (0.028)	0.092
D. mexicanum vs. D. savignyi	0.108 (0.031)	0.017 (0.007)	0.157
D. mexicanum vs. D. setosum	0.088 (0.027)	0.007 (0.002)	0.079
D. paucispinum vs. D. savignyi	0.084 (0.027)	0.013 (0.006)	0.145
D. paucispinum vs. D. setosum	0.070 (0.024)	0.006 (0.003)	0.092
D. savignyi vs. D. setosum	0.138 (0.036)	0.007 (0.003)	0.109
Speract Receptor			
D. antillarum vs D. mexicanum	0.019 (0.008)	0.005 (0.002)	0.314
D. antillarum vs. D. paucispinum	0.045 (0.012)	0.004 (0.002)	0.096
D. antillarum vs. D. savignyi	0.046 (0.010)	0.006 (0.002)	0.135
D. antillarum vs. D. setosum	0.118 (0.017)	0.026 (0.005)	0.223
D. mexicanum vs. D. paucispinum	0.055 (0.011)	0.005 (0.002)	0.080
D. mexicanum vs. D. savignyi	0.051 (0.010)	0.006 (0.002)	0.122
D. mexicanum vs. D. setosum	0.123 (0.018)	0.026 (0.005)	0.213
D. paucispinum vs. D. savignyi	0.032 (0.008)	0.005 (0.002)	0.164
D. paucispinum vs. D. setosum	0.127 (0.018)	0.027 (0.005)	0.214
D. savignyi vs. D. setosum	0.120 (0.017)	0.024 (0.004)	0.242

SE, standard error.

Table 3. Maximum likelihood tests of codon substitution models of variation in ω for speract and its receptor in *Diadema* as determined by PAML

Null model[1] -ℓ	*dn/ds*	Parameter estimates under null model	Alternative model[1] -ℓ	*dn/ds*	Parameter estimates under alternative model	$-2\Delta\ell$	P	Sites under positive selection[2]
Speract								
M1a (nearly neutral)			M2a (selection)					
1005.262	0.115	$p_0 = 0.884$, $p_1 = 0.115$ $\omega_0 = 0.000$, $\omega_1 = 1.00$	1003.070	0.693	$p_0 = 0.916$, $p_1 = 0.0$, $p_2 = 0.071$, $\omega_2 = 1.703$	4.384	0.111	3S, 9I
M7 (neutral, ß) 1003.171	0.118	$p = 0.039$, $q = 0.219$	M8 (selection, ß & ω) 1000.987	1.475	$p_0 = 0.918$, ($p_1 = 0.081$), $p = 0.007$, $q = 2.165$, $\omega = 1.790$	4.368	0.112	3S, 9I
Speract receptor								
M1a (nearly neutral)			M2a (selection)					
2826.360	0.187	$p_0 = 0.820$, $p_1 = 0.179$ $\omega_0 = 0.009$, $\omega_1 = 1.00$	2825.269	0.204	$p_0 = 0.865$, $p_1 = 0.119$, $p_2 = 0.015$, $\omega_2 = 1.000$	2.200	0.332	459I, 503A, 507V
M7 (neutral, ß) 2794.892	0.182	$p = 0.028$, $q = 0.164$	M8 (selection, ß & ω) 2798.833	0.187	$p_0 = 0.821$, ($p_1 = 0.178$), $p = 0.0564$, $q = 1.766$, $\omega = 1.00$	0.118	0.943	58I, 495I, 500A, 504V

ω, ratio of nonsynonymous to synonymous substitutions (d_N/d_S). [1]Models as designated in Yang (2007). -ℓ: -log likelihood. -$\Delta\ell$: -log likelihood ratio values. P = probability derived from the χ^2 distribution. [2]Sites indicated by the analysis as evolving under positive selection- the numbers represent amino acid position in the alignment (Figs. 2 and 3) and letters correspond to amino acid in the first sequence of alignment (*D. savignyi* for speract and *D. setosum* for speract receptor).

signatures of selection in speract and its receptor. For speract, all of the codon site models—Single Likelihood Ancestor Counting (SLAC), Random Effects Likelihood (REL), Fixed Effects Likelihood (FEL) (Kosakovsky Pond and Frost 2005) and the Fast Unconstrained Bayesian Approximation (FUBAR, Murrell et al. 2013) detected several codon sites with strong evidence of purifying selection [PP(Posterior Probability) >0.95, Table 5]. The REL and FUBAR models inferred a single codon site, 9L, in speract to be evolving under diversifying selection but the evidence was not strong ($PP_{REL} = 0.913$, $PP_{FUBAR} = 0.909$). In the receptor for speract, all site models detected strong signatures of purifying selection ($PP > 0.95$) on several codon sites (Table 5). A single site, 507V, inferred to be evolving under positive selection by FEL and FUBAR was not statistically well supported (P $value_{FEL} = 0.0495$, $PP_{FUBAR} = 0.903$). The Branch-site REL and MEME models (Kosakovsky et al. 2011; Murrell et al. 2012) failed to detect sites in any of the lineages with evidence of pervasive or episodic positive selection in either molecule.

Protein structure and amino acid variation

To analyze protein architecture and annotate signal peptides, motifs and domains present in speract and its receptor, we used SignalP (Petersen et al. 2011) and SMART (Ponting et al. 1999). Although SMART uses several alignment-based methods to search large databases, SignalP uses neural networking and machine learning to detect signal molecules. Both analyses detected an amino terminal signal peptide in speract (Fig. 2). The receptor for speract in *Diadema* has a more complex structure (Fig. 3); it contains an amino terminal signal peptide, a large extracellular domain with four distinct scavenger receptor cysteine rich (SRCR) domains, a transmembrane domain and a short (13aa) region at the carboxyl terminal end that extends into the cytoplasm. These structures are similar to what has been annotated in the speract receptor of *S. purpuratus* (Dangott et al. 1989) and in *H. pulcherrimus* (Shimizu et al. 1994).

Amino acid sequence alignments between species of *Diadema* reveal little variation in the speract precursor molecule. The speract-peptide -Gly-Cys-Pro-Trp-Gly-Gly-Ala-Val-Cys- is invariant across all five *Diadema* species (Fig. 2). Of the 14 amino acid substitutions observed in speract between *Diadema* species, substitutions replacing serine, threonine and arginine may be relevant to posttranslational modifications, as these amino acids are commonly subject to phosphorylation and glycosylation (Mann and Jensen 2003; Bedford and Clarke 2009). One such substitution (*D. setosum* 3Ser substituted to Phe/Val in the other species) has occurred in

Table 4. Maximum likelihood tests of branch specific variation in ω for speract and its receptor in *Diadema* as determined by PAML

Foreground branch	-ℓ	Parameter estimates under ModelA (null)[1]	-ℓ	Parameter estimates under ModelA (selection)[1]	-2$\Delta\ell$	P	Sites under positive selection[2]
Speract							
D. antillarum	1088.385	$p_o = 0.836$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.163$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$, $\omega_{fore} = 1$	1088.385	$p_o = 0.836$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.163$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$ $\omega_{fore} = 1$	0.00	1.00	43D
D. mexicanum	1088.282	$p_o = 0.749$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.138$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.09$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.02$, $\omega_{back} = 1$, $\omega_{fore} = 1$	1088.178	$P_o = 0.713$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.138$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.12$, $\omega_{back} = 0$, $\omega_{fore} = 2.78$ $p_{2b} = 0.02$, $\omega_{back} = 1$, $\omega_{fore} = 2.79$	0.208	0.901	
D. savignyi	1086.338	$p_o = 0.098$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.014$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.769$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.11$, $\omega_{back} = 1$, $\omega_{fore} = 1$	-1086.337	$p_o = 0.098$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.014$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.769$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.11$, $\omega_{back} = 1$, $\omega_{fore} = 1$	0.002	0.998	18S, 21R, 148V
D. paucispinum	1088.385	$p_o = 0.836$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.163$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0$, ω_{fore}=*n*1 $p_{2b} = 0.0$, $\omega_{back} = 1$, $\omega_{fore} = 1$	1088.385	$p_o = 0.836$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.163$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$, $\omega_{fore} = 1$	0.00	1.00	50V
D. setosum	1088.209	$p_o = 0.801$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.138$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.05$, $\omega_{back} = 0$, $\omega_{fore} = 1$ $p_{2b} = 0.005$, $\omega_{back} = 1$, $\omega_{fore} = 1$	1088.208	$p_o = 0.817$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.141$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.035$, $\omega_{back} = 0$, $\omega_{fore} = 1.53$ $p_{2b} = 0.006$, $\omega_{back} = 1$, ω_{fore}= 1.42	0.002	0.998	5A, 96I, 184S
Speract receptor							
D. antillarum	3117.706	$p_o = 0.695$, $\omega_{back} = 0.023$, $\omega_{fore} = 0.023$ $p_1 = 0.132$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.146$, $\omega_{back} = 0.023$, $\omega_{fore} = 1$ $p_{2b} = 0.027$, $\omega_{back} = 1$,	3117.544	$p_o = 0.811$, $\omega_{back} = 0.024$, $\omega_{fore} = 0.024$ $p_1 = 0.152$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.030$, $\omega_{back} = 0.024$, $\omega_{fore} = 6.97$ $p_{2b} = 0.005$, $\omega_{back} = 1$	0.324	0.850	90K, 430G

continued

Table 4. (*Continued*)

Foreground branch	-ℓ	Parameter estimates under ModelA (null)[1]	-ℓ	Parameter estimates under ModelA (selection)[1]	-2$\Delta\ell$	P	Sites under positive selection[2]
		$\omega_{fore} = 1$		$\omega_{fore} = 6.97$			
D. mexicanum	3118.042	$p_o = 0.837$, $\omega_{back} = 0.026$, $\omega_{fore} = 0.026$ $p_1 = 0.162$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0.026$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$, $\omega_{fore} = 1$	3118.042	$p_o = 0.836$, $\omega_{back} = 0.026$, $\omega_{fore} = 0.026$ $p_1 = 0.163$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0.026$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$ ω_{fore} 1	0.00	1.00	269V
D. savigyni	3116.572	$p_o = 0.551$, $\omega_{back} = 0.002$, $\omega_{fore} = 0.002$ $p_1 = 0.117$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.273$, $\omega_{back} = 0.002$, $\omega_{fore} = 1$ $p_{2b} = 0.058$, $\omega_{back} = 1$, $\omega_{fore} = 1$	3116.572	$p_o = 0.551$, $\omega_{back} = 0.002$, $\omega_{fore} = 0.002$ $p_1 = 0.117$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.272$, $\omega_{back} = 0.002$, $\omega_{fore} = 1.0$ $p_{2b} = 0.058$, $\omega_{back} = 1$ $\omega_{fore} = 1.0$	0.00	1.00	34E, 35Y, 226D, 375T.
D. paucispinum	3118.042	$p_o = 0.837$, $\omega_{back} = 0.026$, $\omega_{fore} = 0.026$ $p_1 = 0.163$, $\omega_{back} = 1$ $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0.026$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$, $\omega_{fore} = 1$	3118.042	$p_o = 0.837$, $\omega_{back} = 0.026$, $\omega_{fore} = 0.026$ $p_1 = 0.162$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.0$, $\omega_{back} = 0.026$, $\omega_{fore} = 1$ $p_{2b} = 0.0$, $\omega_{back} = 1$ $\omega_{fore} = 1$	0.00	1.00	
D. setosum	3115.340	$p_o = 0.736$, $\omega_{back} = 0$, $\omega_{fore} = 0$ $p_1 = 0.109$, $\omega_{back} = 1$ $\omega_{fore} = 1$ $p_{2a} = 0.134$, $\omega_{back} = 0.00$, $\omega_{fore} = 1$ $p_{2b} = 0.019$, $\omega_{back} = 1$, $\omega_{fore} = 1$	3115.221	$p_o = 0.788$, $\omega_{back} = 0.028$, $\omega_{fore} = 0.028$ $p_1 = 0.114$, $\omega_{back} = 1$, $\omega_{fore} = 1$ $p_{2a} = 0.085$, $\omega_{back} = 0.028$, $\omega_{fore} = 186.54$ $p_{2b} = 0.012$, $\omega_{back} = 1$ $\omega_{fore} = 186.54$	0.238	0.887	103K, 154T, 209 D, 215L, 318K, 380G, 323E, 450N, 477D

ω-ratio of nonsynonymous to synonymous substitutions (d_N/d_S). ω_{fore}, ω_{back}: background ratio. [1]Models as designated in Yang (2007). -ℓ: -log likelihood. -$\Delta\ell$: -log likelihood ratio values. P = probability derived from the χ^2 distribution. [2]Sites indicated by the analysis to be evolving under positive selection- the numbers represent amino acid position and letters correspond to amino acid in the first sequence of alignment (Figs. 2 and 3).

the signal peptide. Five other substitutions affect hydrophobicity, polarity and charge downstream of the signal peptide (Fig. 2). None of the lysine residues, which are potential docking or cleaving sites (Caron et al. 2005) change in speract, which suggests that functional integrity is maintained between species. Amino acid variation in the much larger receptor molecule does not appear to be concentrated in any specific region of the molecule (Fig. 3). Owing to indels, *D. setosum* has two additional amino acids in the amino terminal signal peptide region, and *D. paucispinum* is missing two amino acids in the carboxyl terminal transmembrane region. Despite various substitutions across species that affect polarity, hydrophobicity and amino acids commonly associated with posttranslational modifications, all cysteine residues, integral to disulphide bond formation, are conserved, suggesting no major structural variations across species (Fig. 3).

Table 5. Sites identified to be under purifying selection by HyPhy

	Codon	SLAC[1]	FEL[1]	REL[1,2]	FUBAR[3]
Speract					
	51	−5.149	−117.015	−4.419 (0.992)	−2.449 (0.989)
	63	−4.140	−532.139	6.995 (0.995)	−8.229 (0.996)
	69	−3.855	−289.331	−5.761 (0.966)	−.5.733 (0.977)
	127	**−4.621**	**−388.389**	−6.206 (0.994)	−5.727 (0.999)
	161	**−7.437**	**−295.657**	−6.016 (0.995)	−5.606 (0.998)
	190	−8.577	**−962.249**	−6.693 (0.996)	−7.616 (0.995)
Speract Receptor					
	21	−13.726	−136.508	−4.991 (0.988)	−4.558 (0.971)
	59	−13.589	−136.981	−3.993 (0.991)	−4.588 (0.976)
	153	**−34.750**	**−586.44**	−26.661 (1.00)	−11.853 (1.00)
	169	−27.174	**−583.542**	−26.595 (1.00)	−11.547 (0.999)
	176	−13.788	−123.733	−3.225 (0.999	−5.657 (0.989)
	242	−20.264	**−380.696**	−14.298 (0.999)	−10.690 (0.993)
	280	−13.678	−140.228	−3.992 (0.990)	−2.999 (0.967)
	332	−20.010	**−140.03**	−3.243 (0.999)	−4.786 (0.987)
	362	−13.842	−148.301	−4.882 (0.991)	−3.658 (0.969)
	365	−22.248	−144.029	−6.201 (0.989)	−4.742 (0.988)
	373	**−44.496**	**−890.56**	−27.354 (1.00)	−13.587 (1.00)
	389	**−20.999**	**−899.999**	−9.899 (0.999)	−5.982 (0.989)
	414	−22.248	**−294.827**	−11.289 (0.999)	−6.537 (0.998)
	423	−13.001	−303.222	−7.772 (0.990)	−5.468 (0.987)
Speract	540	−20.381	**−309.314**	−20.772 (0.999)	−8.178 (0.998)

[1]d_N/d_S values, $P < 0.05$ for values in normal font, $P < 0.005$ for values in bolded font. [2]Values in parentheses are mean posterior probability for $d_N > d_S$ and (ω) > 1.0 at a site. [3]ß/α (d_N/d_S) values in parentheses are mean posterior probability for ω (=β/α) > 1.0 at a site. See text for definition of test acronyms.

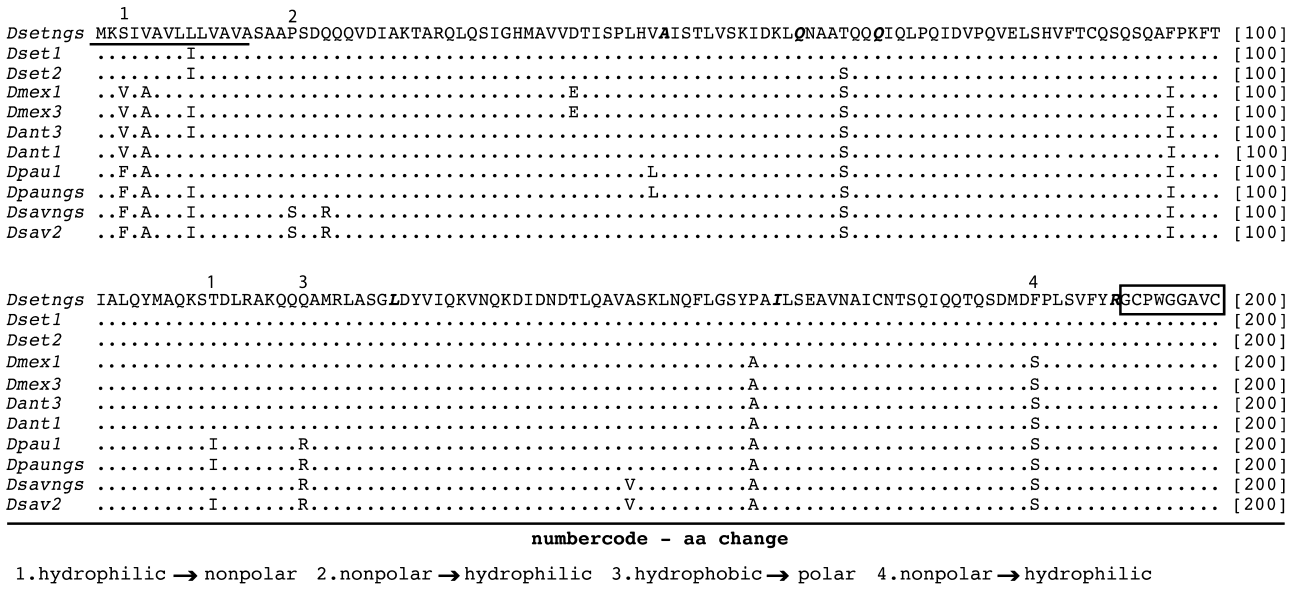

Fig. 2. Amino acid variation in speract between species of *Diadema*. The signal peptide sequence is underlined. The sperm activating peptide (SAP IV) is boxed. Number codes denote amino acid property changes that have occurred (note that the arrows do not imply any direction of change between species). Residues in bolded italics are sites identified to be under purifying selection. Numbers next to species names represent specimen identification numbers. ngs: sequences obtained from next generation sequencing. *Dant - Diadema antillarum*, *Dmex - D. mexicanum*, *Dpau - D. paucispinum*, *Dsav - D. savignyi*, *Dset - D. setosum*.

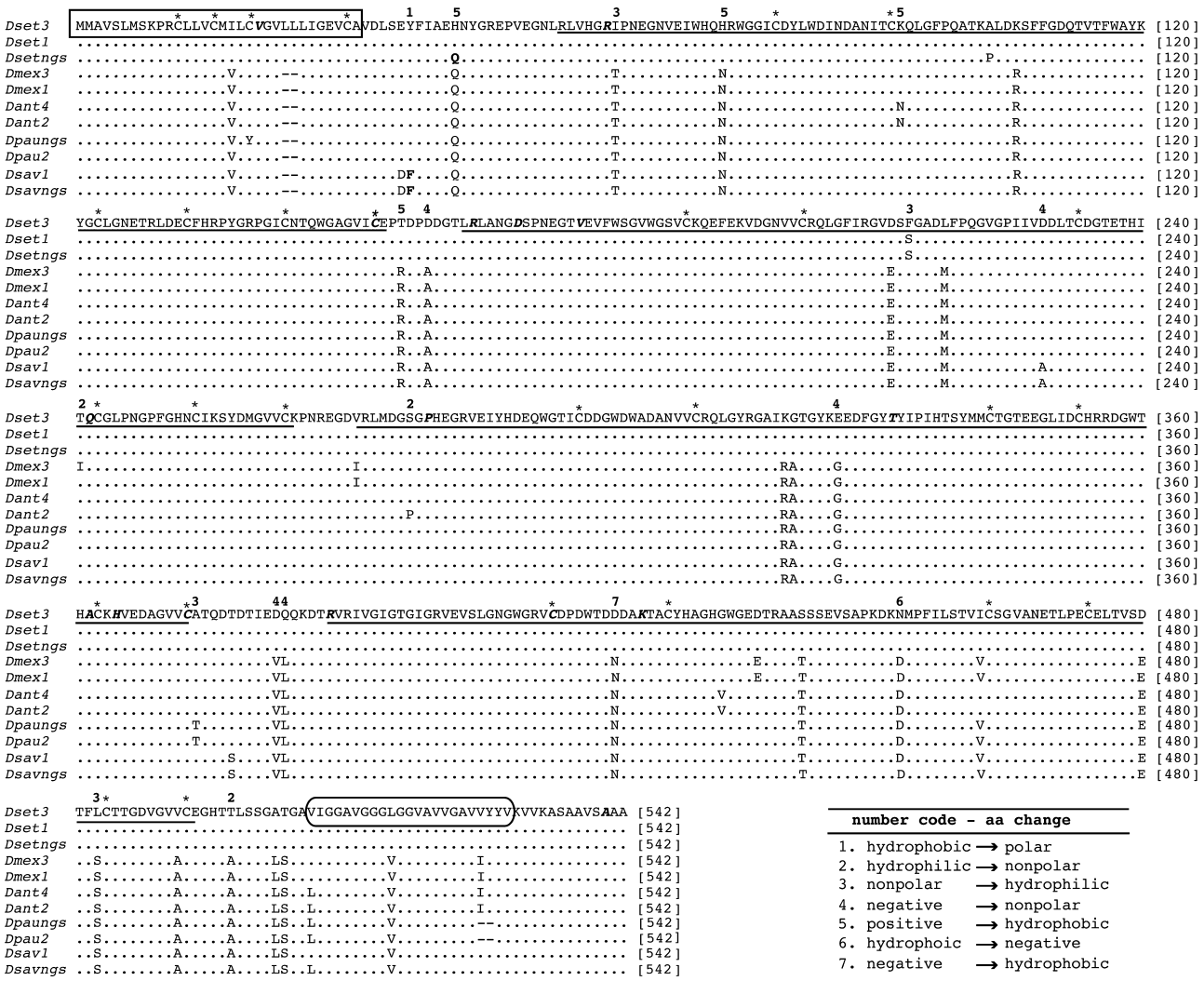

Fig. 3. Amino acid variation in Diadema speract receptor. Signal peptide sequence is boxed. Extracellular (SRCR) domains are underlined. Box with rounded edges - transmembrane domain. Cysteine residues are indicated by asterisks. Species name abbreviations are the same as in Fig 2, and specimen identification numbers are the same as in Fig 1. Numbers denote amino acid property changes that have occurred (note that the arrows do not imply any direction of change between species). Residues in bolded italics are sites identified to be under purifying selection.

Variation across sea urchin genera

Is it possible that speract and its receptor are under functional constraint across sea urchin higher taxa? We compared speract and its receptor's protein sequence of *Diadema* to orthologous sequences of *S. purpuratus* (Dangott et al., 1989; Ramarao et al., 1990), *H. pulcherrimus* (Kinoh et al., 1994; Shimizu et al., 1994), *P. depressus* (Yamano et al. unpublished) and to a partial speract cds from *E. tribuloides* (order Cidaroida) to gain additional insights into the evolutionary dynamics of these reproductive molecules across sea urchin genera. At the amino acid level, speract proteins of *Diadema* and *Eucidaris* are approximately 58% similar, and the speract proteins of *Diadema* and members of the family *Strongylocentrotidae* are approximately 40–50% similar. This is not unexpected given that diadematoids, cidaroids and strongylocentrotids diverged from each other roughly 300–250 Ma (Kroh and Smith 2010). There is considerable variation in the speract molecule; conserved amino acids are more easily observed among members of the Strongylocentrotidae (and to some extent, between *Diadema* and the Strongylocentrotidae), but even in this case, speract of *P. depressus* is only 42% similar to the speract of *S. purpuratus* and 44% similar to the speract of *H. pulcherrimus*. Phylogenetic relationships of speract nucleotide sequences constructed using Maximum Likelihood methods are concordant with tree topologies of species inferred from molecular, as well as fossil data (Smith et al. 2006; Kroh and Smith 2010)(Fig. 4). Pairwise computations of the rates of divergence in speract (ignoring

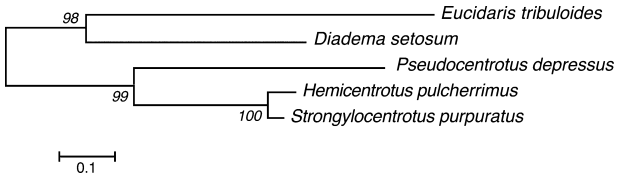

Fig. 4. Maximum Likelihood phylogeny of speract across sea urchin genera. Tree is unrooted. Bootstrap support values are presented at nodes.

gaps) show that the proportions of silent substitutions are consistently higher than replacement substitutions in all comparisons, as would be expected over such large evolutionary distances. The most structurally relevant difference is variation in the number of tandem repeats of speract peptides and their amino acid compositions, which have been shown to be considerably different between genera and orders of sea urchins (Yoshino et al. 1990). All *Diadema* species have a single, nine amino acid, speract peptide. *P. depressus* also has a single speract peptide. However, *H. pulcherrimus* has four additional speract peptides compared to *S. purpuratus*, though some amino acid variation is shared between the two (Fig. 5 also see Kinoh et al. 1994). The *Diadema* and *P. depressus* speract peptides align to the last speract peptide repeat of *S. purpuratus* and *H. pulcherrimus* (Fig. 5) with reasonably good support as inferred by PRANK ($PP_{Diadema} \geq 90\%$, $PP_{P.\ depressus} > 95\%$).

In contrast to speract, the *Diadema* receptor for speract (Fig. 6) shares slightly higher amino acid identities of 69% and 60% with *H. pulcherrimus* and *S. purpuratus,* respectively. Despite over 250 million years of divergence, highly conserved arginine and cysteine residues with very few changes in glycine and glutamine residues suggest that the general structure of the speract receptor protein is conserved across *Hemicentrotus*, *Strongylocentrotus* and *Diadema* (Fig. 6). There are, however, a few amino acid changes within the extracellular SRCR and transmembrane domains that may be of some relevance to how sperm respond to their respective speracts.

DISCUSSION

Chemoattractant properties of eggs are common in metazoans (Eisenbach 1999; Eisenbach and Giojalas 2006) but are of particular importance in free spawning organisms that rely on these molecules to mediate the first stage of fertilization—activating and attracting the sperm to swim toward the egg (Miller 1985). Understanding whether gametic molecules involved in this process are conserved or whether they evolve under different selective pressures should shed light on the relative importance of species-specificity and opportunities for species discrimination during the stage of sperm-egg attraction.

Evidence for purifying selection on speract and its receptor

Our data reveal very little variation in the protein-coding nucleotide and amino acid sequences of speract and its receptor between species of *Diadema*. Of the nucleotide variation that is

```
E.tri MKSVVCLLLFA T-------MVCGAPPSS---------SSSSSSTWQSDT-RIDITSSSKSIQQIAHAAVVDALTPLHDALSSLEYQISNLVDVAQKNRQ [100]
D.set ...I.AV..LV A-------VASA..----------..QQQV..AKTARQL.S.G.M....TIS...V.I.T.VSK.DK.QNA.SQQQI [100]
P.dep ..FILV..CIVIA-------V.Q.S.I..TDDMHPESQ.N---------.Q---T.SD.AMY..GVS.M.I..MFS..MS.....DL.L.K.HNI.VQIGV [100]
H.pul .TFI....VLV A-------VAS.K.VI.---------GQ--------.Q_KYTLAD...DLEG...M.I..S.S...IS.....SAWN...SI.TQEKY [100]
S.pur -------------MPPGPRGVAS.K.VI.---------GQ--------.Q_KYTLAD...DLEG...M.I..S.S...IS.....SAWN..LNI.SQEDY [100]

E.tri DLPQISMRKVDLPACSNAKPSRGPRPASCSRSSTRDQTSRAKSTRPRHSAVGLNLSSKRSTISRLIETLWKRQLTELTSSWNNTQNCQSKSSRKPVLHRS [200]
D.set Q....DVPQ.E.------------------------------------------------------------------------SHVFT...Q.-------QA [200]
P.dep TI.SVVVP.I.M--------------------------------------------------------------------------TVIYE.KA.YE-------- [200]
H.pul II.EL.IP.I.V------------------------------------------------------------------------KSILT.KP.Y.-------- [200]
S.pur SI.EL.IP.I.V------------------------------------------------------------------------KSILS.KP.Y.-------- [200]

E.tri LQTFT-------------------------------------------------------------------------E------------------------ [300]
D.set IPK..IALQYMAQKSTDLRAKQQQ-------AMRLASGLDYVIQKVNQKDIDNDTLQAVASKLNQFLGSYASILS.AVNAIC-----------NTSQIQ [300]
P.dep -PKYSPALTY-------LMSKEDEIESHIDTATLIAHGIEYVNNLLVNDEMDQKTLQAMSSKLNDFLSDYSSLISGLQSSVCVDQNHKIRETMQQ----E [300]
H.pul -PKYPVVLQY-------ISDHQVQVQDHIANANELVEGLKFVSQLIMYKKVDHDTLASVSKMLSDYLTDYASTISSIKSVVC------------Q----D [300]
S.pur -PKYPVVLQY-------ISDHYVQVQDHIANAKELTEGLKFVSQLIMYKKIDHDTLASVSKMLSNYLTDYASTISSLKSVVC-----------------D [300]

E.tri ---------------------------------------------------------------------------------------------------- [400]
D.set QT-------QSDMDSPLSVF---------------------------------------------------------------------------------- [400]
P.dep HT-------EEYMNPALDIF-------------------------------------------------------------------------------- [400]
H.pul PTAPSHPMAESYMDTPLSMFLKGTMPTGGGVEKGFALGGGGVGKGFDLNGGGVGKGFDLNGGGVGKGFDLNGGGVGKGFDLNGGGVSKGFELNGGGVGKG [400]
S.pur QTAPSHPMDESYMDTPLSMILKGTMPTGAGVDKGFALGGGGVGKGFNLNGGGVGKGFDLNGGGVGKGFDLNGGGVGKGFDLNGGG-------------- [400]

E.tri ---------------------------------------------------------- [466]
D.set ------------------------------------------Y--RG-CPWGGAVC-------- [466]
P.dep ------------------------------------------FSERGNCLTGGCIDK------- [466]
H.pul FDLNGGGVGKGFDLTGGGVGKGFDLNGGGVGKGFALGGGGVGKGFS-----LTGGSVDREVETEGW [466]
S.pur ----------VGKGFDLNGGGVGKGFALGGGGVGKGFS-----LTGGGVGREVEIKDW [466]
```

Fig. 5. Speract amino acid variation across sea urchin genera. Signal peptide sequence is underlined. Speract and speract-like peptides are boxed. *Etri - Eucidaris tribuloides*, *Dset - Diadema setosum*, *Pdep - Pseudocentrotus depressus*, *Hpul - Hemicentrotus pulcherrimus*, *Spur - Strongylocentrotus purpuratus*.

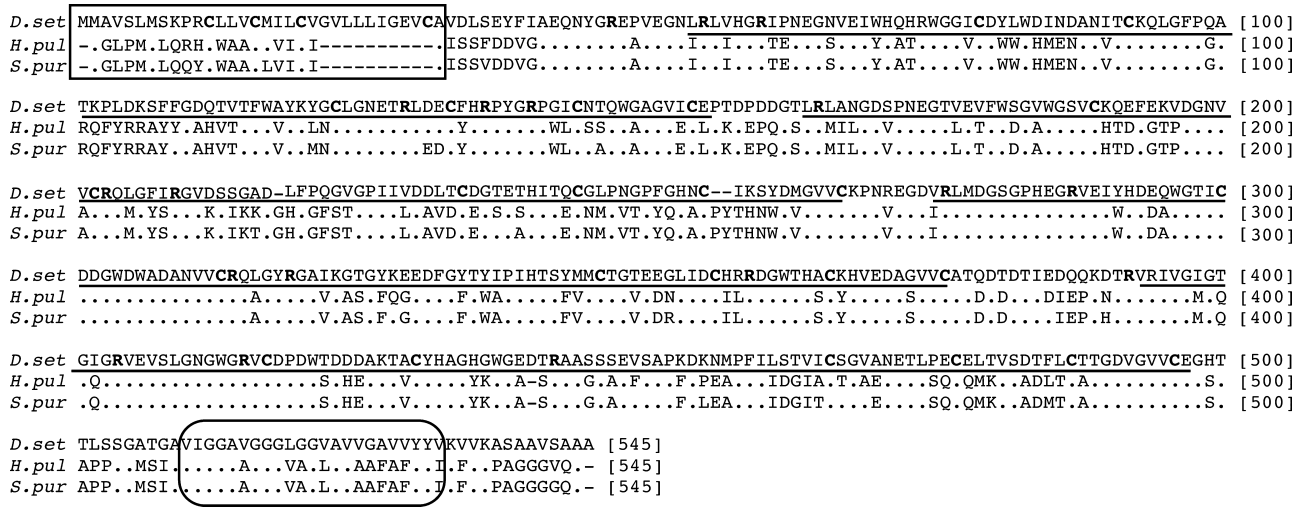

Fig. 6. Amino acid variation in speract receptor across sea urchin genera. *D.set - Diadema setosum*, *H. pul - Hemicentrotus pulcherrimus*, *S. pur - Strongylocentrotus purpuratus*. Signal peptide sequence is boxed. SRCR domains are underlined. Box with rounded edges - transmembrane domain. Conserved cysteine (C) and Argnine (R) are in bold letters.

present, rates of synonymous and nonsynonymous substitutions per site and their ratios show a consistent pattern of $d_S > d_N$ across species (Table 2). Maximum Likelihood tests of selection using PAML and HyPhy failed to find sites with significant evidence (at the 95% level) of diversifying selection in both molecules. However, some site models in HyPhy did find marginal evidence of diversifying selection at the 90% confidence level at site 9 L in speract ($PP_{REL} = 0.913$, $PP_{FUBAR} = 0.909$) and site 507 V in the receptor for speract ($p_{FEL} = 0.0495$, $PP_{FUBAR} = 0.903$,). Site 9 L lies within the signal peptide in speract, which is not directly involved in gamete attraction, and site 507 V is in the transmembrane domain of the receptor. There are no differences in the patterns of substitutions at these sites between allopatric and sympatric species that may be indicative of interspecific challenges in sympatry. Further sampling and detailed population studies may perhaps shed some light on intraspecific processes that can influence the diversification of these gametic molecules (see McCartney and Lessios 2004; Geyer and Lessios 2009; Hart et al. 2014).

On the other hand, we find substantial evidence of purifying selection acting on several sites in both proteins (Table 5, Figs. 2 and 3). Although the rapid and often adaptive evolution of some reproductive proteins may indeed be associated with speciation (Swanson and Vacquier 2002; Vacquier and Swanson 2007; Jagadeeshan et al. 2011), nearly neutral, or even purifying selection in reproductive proteins is not unusual (Kimura 1983; Nei 1987; Rooney et al. 2000; Findlay and Swanson 2009). What is particularly noteworthy in our data is that there is no evidence of diversifying selection on codon sites that encode the diffusible speract peptide, which is responsible for activating sperm. Does this indicate that *Diadema* speract indiscriminately activate and/or attract sperm of congeners? There are no data on the nature of species-level discrimination in sperm egg interactions or gamete incompatibility in the genus *Diadema*. In fact, species-level discriminatory properties of speract have rarely been studied in sea urchins (Suzuki 1995). However, species ecology and biogeography provide some insights into whether speract and its receptor face any selective pressures to diversify in *Diadema*.

Species of the genus *Diadema* are known for their conspicuous lunar spawning rhythms, which are typically temporally out of phase between species (Yoshida 1966; Pearse 1975; Lessios 1984; Coppard and Campbell 2005a). Sperm and egg of some congeners, therefore, rarely come in contact with each other. However, some *Diadema* species do face the risk of hybrid production in nature (Lessios and Pearse 1996), which is expected to affect the evolution of certain reproductive molecules (Coyne and Orr 2004). *Diadema setosum* and *D. savignyi* are sympatric throughout much of the Tropical West Pacific and Indian oceans, and show no spatial isolation on coral reefs (Pearse 1975; Lessios and Pearse 1996; Lessios et al. 2001; Coppard and Campbell 2005b). Although Coppard and Campbell (2005a) reported that *D. savignyi* spawns at full moon and *D. setosum* spawns at new moon in Fiji, geographical variation in spawning times and seasons (Pearse 1975; Coppard and Campbell 2005a) suggests that eggs and sperm of *D. savignyi*, *D. setosum* may at times come in contact with one another, creating the possibility of hybridization between these species. In some regions of the Indo-Pacific ocean, there is also the possibility for hybridization between *D. savignyi*, *D. setosum* and *D. paucispinum* (spawning rhythm unknown) (Lessios and Pearse 1996; Lessios et al. 2001). Indeed, a few hybrids between all three have been reported (Lessios and Pearse 1996) and Uehara et al. (1990) successfully cross fertilized *D. savignyi* and *D. setosum* in the laboratory. None of these Indo-Pacific species show elevated rates of evolution nor

contain any notable signatures of diversifying selection in speract or the receptor for speract; in fact, our analyses reveal significant selective constraint on these molecules.

At the amino acid level, most variation in speract is observed between the Indo-Pacific *D. setosum, D. savignyi* and *D. paucispinum*, whereas only two amino acid sites are different between the allopatric *D. mexicanum* and *D. antillarum* (Fig. 2). Nevertheless, the diffusible speract nonapeptide is invariant across all species, suggesting that *Diadema* speract can indiscriminately activate con- and –heterospecific sperm in seawater. But speract is also required for optimal induction of acrosome reaction in the egg jelly; speract works as a cofactor with fucose sulfate polysaccharides to induce acrosome reaction once the sperm enters the egg jelly (Yamaguchi et al. 1988). This may warrant some speculation regarding whether variations in the precursor molecule are of consequence to differences in signal transduction between species. In line with this speculation, amino acid variation in sites commonly affected by posttranslational modifications may influence how speract molecules are processed between the species. Whether these differences affect speract's function within the egg jelly will need to be studied. In the *Diadema* receptor for speract, the SRCR domains within the extracellular domain as well as transmembrane domains are integral for protein–protein interaction and ligand binding (Pancer et al. 1999). Within this region, all cysteine and arginine residues that are integral for protein conformation and structure (Hohenester et al. 1999) are conserved across *Diadema* species (Fig. 3), providing no evidence for any conformational differences between species. All of the sites with evidence of purifying selection lie within the SRCR domains (Table 5, Fig. 3). However, the indels in *D. setosum* and *D. paucispinum*, and changes in amino acids commonly associated with post-translational modifications can be potential sources of variation to how signal transduction occurs between species. For instance, these differences may be relevant to the manner by which sperm respond to speract's induction of intracellular changes, such as Ca^{+} influx, N^{+}/H^{+} exchange, increase in cyclic AMP and pH (Repaske and Garbers 1983; Schackmann and Chock 1986), all of which are important not only for motility and respiration, but also for undergoing optimal induction of acrosome reaction (Yamaguchi et al. 1988; Neill and Vacquier 2004). Overall, the trend of molecular evolution observed in our study provides little evidence of major functional or structural changes in speract or its receptor between *Diadema* species that would point to their involvement in establishing reproductive barriers at the stage of gamete-attraction.

Variation across sea urchin genera

Our comparison of speract and its receptor's protein sequence of *Diadema* to orthologous sequences of *S. purpuratus* (Dangott et al. 1989; Ramarao et al. 1990), *H. pulcherrimus* (Kinoh et al. 1994; Shimizu et al. 1994), *P. depressus* (Yamano et al. unpublished) and a partial speract cds from *E. tribuloides* produced some tentative insights into the evolutionary dynamics of these reproductive molecules across sea urchin genera. Pairwise computations of silent and amino acid replacements in speract produced no evidence of selection, even on distant comparisons of species that have diverged 300–250 million years ago. Although this would suggest that speract may be under functional constraint or may evolve neutrally across genera, this could be a misleading deduction without further analysis of divergences within each of these genera (as in this study). Heterogeneity in rates of divergence and selective pressures across lineages, and within species on reproductive molecules may not be uncommon (McCartney and Lessios 2004; Geyer and Lessios 2009; Sunday and Hart 2013; Hart et al. 2014). The differences in tandem amino acid repeats in the molecules of different genera may have evolutionary implications. Tandem peptide repeats in a coevolving system may present opportunities for concerted evolution (Swanson and Vacquier 1998). It may also be of interest that the egg molecule is more divergent between genera than its receptor on the sperm. If this difference holds true, pending more thorough investigations in additional sea urchin taxa, it would provide a contrast with what has been proposed previously in internally fertilizing animals, as well as in the case of *VERL* and *Lysin* in abalones, in which male reproductive molecules tend to evolve faster (Swanson and Vacquier 1998; Jagadeeshan and Singh 2005; Haerty et al. 2007).

CONCLUSIONS

In investigating whether reproductive barriers can be established at the sperm-egg attraction stage, our study indicates that the genes encoding the sperm-activating molecule and its receptor in *Diadema* show similar trends of molecular evolution in both allopatric and sympatric species. The majority of sites are evolving neutrally but several sites in both molecules contain strong signatures of purifying selection. Most importantly, lack of variation in the diffusible speract peptide between *Diadema* species implies that these molecules activate/attract sperm of their congeners indiscriminately. As such it is unlikely that speract or its receptor contribute to reproductive isolation in *Diadema*. In all likelihood, molecules involved in subsequent stages, when the sperm passes through the egg jelly, or binds to the vitelline envelope, may be involved. On the other hand, evolution of asynchronous spawning rhythms may be more important than rapid and adaptive sequence evolution in *Diadema*. As Coyne and Orr (2004) have suggested, isolating barriers acting earlier in the process of mating are under stronger selection because divergence at an earlier stage obviates any need for modifications at a later one. The only other analogous study of speract-like molecules by Nakachi et al. (2008) in sea

stars also did not find any evidence for diversifying selection in asterosap, or its receptor. Although comparisons across a few genera point to the possibility that speract and its receptor may not be under directional selection, they do not necessarily reflect factors operating within genera and species (e.g., concerted evolution) that may be responsible for the considerable variation observed in the number of speract repeats across genera. Additional investigations in other sea urchin genera are needed to determine whether the mode of evolution of these molecules is similar to that of *Diadema,* and to infer the importance of sperm-egg attraction in reproductive isolation of sea urchins.

Acknowledgements

We thank Mathew Ross for collecting *D. paucispinum* from Hawaii. We are grateful to McGill University and Genome Quebec Innovation Centre for NGS. Axel Calderon helped with collections of *Diadema* in Panama. We also thank Danny Absalon Gonzalez, Yherson Francisco Molina, Ligia Calderon, Axel Calderon and Laura Geyer for their assistance in the laboratory. Autoridad de Recursos Acuáticos de Panamá (ARAP) kindly permitted collections in Panama. We also thank two anonymous reviewers for valuable comments and suggestions that improved this manuscript. This study was funded by a Smithsonian Molecular Evolution Fellowship to S. J, a Smithsonian Next Generation Small Grant to H.A.L, S. J and S. E. C, and by General Research funds from STRI.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19: 716–723.

Bedford, M. T., and Clarke, S. G. 2009. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell.* 33: 1–13.

Biermann, C. H., Marks, J. A., Vilela-Silva, A. C., Castro, M. O., and Mourao, P. A. 2004. Carbohydrate-based species recognition in sea urchin fertilization: another avenue for speciation. *Evol. Dev.* 6: 353–361.

Cardullo, A. R., Herrick, S. B., Peterson, M. J., and Dangott, L. J. 1994. Speract receptors are localized on sea urchin sperm flagella using a fluorescent peptide analog. *Dev. Biol.* 162: 600–607.

Caron, C. 2005. Regulatory cross-talk between lysine acetylation and ubiquitination: role of the control of protein stability. *Bioessays* 27: 408–415.

Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F., and Swanson, W. J. 2009. Coevolution of interacting fertilization proteins. *PLoS Genet.* 5: e1000570.

Coppard, S. E., and Campbell, A. C. 2005a. Lunar periodicities of diadematoid echinoids breeding in Fiji. *Coral Reefs* 24: 324–332.

Coppard, S. E., and Campbell, A. C. 2005b. Distribution and abundance of reglar sea urchins on two coral feefs in Fiji. *Micronesia* 37: 249–269.

Coyne, J. A., and Orr, H. A. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

Dan, J. C. 1967. Acrosome reaction and lysins. In C. B. Metz and and A. Monroy (eds.). *Fertilization: Comparative Morphology, Biochemistry, and Immunology.* Academic Press, San Diego.

Dangott, L. J., Jordan, J. E., Bellet, R. A., and Garbers, D. L. 1989. Cloning of the mRNA for the protein that crosslinks to the egg peptide speract. *Proc. Natl. Acad. Sci. U S A* 86: 2128–2132.

Darriba, D., Taboada, G. L., Doalla, R., and Posada, D. 2012. JModeltest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9: 772.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.. *Nucleic Acids Res.* 32: 1792–1797.

Eisenbach, M. 1999. Sperm chemotaxis. *Rev. Reprod.* 4: 56–66.

Eisenbach, M., and Giojalas, L. C. 2006. Sperm guidance in mammals - an unpaved road to the egg. *Nat. Rev. Mol. Cell. Biol.* 7: 276–285.

Findlay, G. D., and Swanson, W. J. 2009. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *Bioessays* 32: 26–36.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.

Fu, Y. X., and Li, W. H. 1993. Statistical tests of neutrality for DNA samples from a population. *Genetics* 143: 557–570.

Galindo, B. E., Vacquier, V. D., and Swanson, W. J. 2003. Positive selection in the egg receptor for abalone sperm lysin. *Proc. Natl. Acad. Sci. U S A* 100: 4639–4643.

Garbers, D. L. 1989. Molecular basis of fertilization. *Annu. Rev. Biochem.* 58: 712–742.

Geyer, L. B., and Lessios, H. A. 2009. Lack of character displacement in the male recognition molecule, bindin, in Altantic sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* 26: 2135–2146.

Goeck, J., Nekrutenko, A., Taylor, J., and Team, G. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11: R86.

Grabherr, M. G., et al. 2011. Full length transcriptome assemby from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29: 664–652.

Guerrero, A., Nishigaki, T., Carneiro, J., Yoshiro, T., Wood, C. D.,and Darzon, A. 2010. Tuning sperm chemotaxis by calcium burst timing. *Dev. Biol*. 344: 52–65.

Guerrero, A., Wood, C. D., Nishigaki, T., Carneiro, J., and Darzon, A. 2010. Tuning sperm chemotaxis. *Biochem. Soc. Trans*. 38:1270–1274.

Haerty, W., et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177: 1321–1335.

Hansborough, J. R., and Garbers, D. L. 1981. Speract Purification and characterization of a peptide associated with eggs that activates spermatozoa. *J. Biol. Chem.* 256: 1447–1452.

Hart, M. W. 2013. Structure and evolution of the sea star egg receptor for sperm bindin. *Mol. Ecol.* 22: 2143–2156.

Hart, M. W., Sunday, J. M., Popovic, I., Learning, K. J., and Konrad, C. M. 2014. Incipient speciation of sea star populations by adaptive gamete recognition coevolution. *Evolution* 68: 1294–1305.

Hasegawa ,M., Kishino ,H., and Yano ,T.-A. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22: 160–174.

Hathaway, R. R. 1963. Activation of respiration in sea urchin spermatozoa by egg water. *Biol. Bull.* 125: 486–498.

Hellberg, M. E., Dennis, A. B., Arbour-Reily, P., Aagaard, J. E., and Swanson, W. J. 2012. The *Tegula* tango: a coevolutionary dance of interacting, positively selected sperm and egg proteins. *Evolution* 66: 1681–1694.

Hirohashi, N., Kamei, N., Kubo, H., Sawada, H., Matsumoto, M., and Hoshi, M. 2008. Egg and sperm recognition systems during fertilization. *Dev. Growth Differ.* 50: S221–S238.

Hohenester, E., Sasaki, T., and Timpl, R. 1999. Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. *Nat. Struc. Biol.* 6: 228–232.

Hughes, A. L., and Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U S A* 86: 958–962.

Jagadeeshan, S., and Singh, R. S. 2005. Rapidly evolving genes of *Drosophila*: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. *Mol. Biol. Evol.* 22: 1793–1801.

Jagadeeshan, S., Haerty, W., and Singh, R. S. 2011. Is speciation accompanied by rapid evolution? Insights from comparing reproductive and nonreproductive transcriptomes in *Drosophila*. *Int. J. Evol. Biol.* 2011: 595121.

Kamei, N., and Glabe, C. G. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1,a novel ADAMTS protein. *Genes Dev.* 17: 2502–2507.

Kaupp, U. B. 2012. 100 years of sperm chemotaxis. *J. Gen. Physiol.* 140: 583–586.

Kaupp, U. B., Hildebrand, E., and Weyand, I. 2006. Sperm chemotaxis in marine invertebrates–molecules and mechanisms. *J. Cell Physiol.* 208: 487–494.

Kaupp, U. B., Kashikar, N. D., and Weyand, I. 2008. Mechanisms of sperm chemotaxis. *Annu. Rev. Physiol.* 70: 93–117.

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

Kinoh, H., Shimizu, T., Fujimoto, H., and Suzuki, N. 1994. Expression of a putative precursor mRNA for sperm-activating peptide I in accessory cells of the ovary in the sea urchin *Hemicentrotus pulcherrimus*. *Roux's Arch. Dev. Biol.* 203: 381–388.

Kopf, G. S., Tubb, D. J., and Garbers, D. L. 1979. Activation of sperm respiration by a low molecular weight egg factor and by 8-bromoguanosine 3;5′-monophosphate. *J. Biol. Chem.* 254: 8554–8560.

Kosakovsky Pond, S. L., and Frost, S. D. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22: 1208–1222.

Kosakovsky Pond, S. L., Frost, S. D., and Muse, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D., Delport, W., and Scheffler, K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 28: 3033–3043.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23: 1891–1901.

Kroh, A., and Smith, A. B. 2010. The phylogeny and classification of post-Palaeozoic echinoids. *J. Syst. Palaentology* 8: 141–212.

Lessios, H. A. 1984. Reproductive periodicities of the echinoids *Diadema* and *Echinometra* on the two coasts of Panama. *J. Exp. Mar. Biol. Ecol.* 50: 47–61.

Lessios, H. A. 2007. Reproductive isolation between species of sea urchins. *Bull. Mar. Sci.* 81: 191–208.

Lessios, H. A. 2011. Speciation genes in free-spawning marine invertebrates. *Integr. Comp. Biol.* 51: 456–465.

Lessios, H. A., Kessing, B. D., and Pearse, J. S. 2001. Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* 55: 955–975.

Lessios, H. A., and Pearse, J. S. 1996. Hybridization and introgression between Indo-pacific species of *Diadema*. *Mar. Biol.* 126: 715–723.

Lessios, H. A. and Zigler, K. S. 2012. Rates of sea urchin bindin evolution. In R. S. Singh, J. P. Xu and and R. J. Kulathinal (eds.). *Rapidly evolving genes and genetic systems*. Oxford University Press, Oxford.

Letunic, I., Doerks, T., and Bork, P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40:D302–D305.

Li, W. H. 1993. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 36: 96–99.

Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Lillie, F. R. 1912. The production of sperm Iso-Agglutinins by ova. *Science* 36: 527–530.

Loytynoja, A., and Goldman, N. 2010. WebPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* 11: 579.

Maddion, D. R., and Maddison, W. P. 2005. MacClade 4: analysis of phylogeny and character evolution.

Mah, S. A., Swanson, W. J., and Vacquier, V. D. 2005. Positive selection in the carbohydrate recognition domains of sea urchin sperm receptor for egg jelly (suREJ) proteins. *Mol. Biol. Evol.* 22: 533–541.

Mann, M., and Jensen, O. N. 2003. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21: 255–261.

McCartney, M. A., and Lessios, H. A. 2004. Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* 21: 732–745.

Miller, A. 1985. Sperm chemo-orientation in the metazoa. In A. Monroy and and C. Metz (eds.). *Biology of Fertilization*. Academic press, New York, pp. 275–337.

Murrell, B., et al. 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30: 1196–1205.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764.

Nakachi, M., Hoshi, M., Matsumoto, M., and Moriyama, H. 2008. Conserved sequences of sperm-activating peptide and its receptor throughout evolution, despite speciation in the sea star *Asterias amurensis* and closely related species. *Zygote* 16: 229–237.

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

Neill, A. T., and Vacquier, V. D. 2004. Ligands and receptors mediating signal transduction in sea urchin spermatozoa. *Reproduction* 127: 141–149.

Nielsen, R., and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929–936.

Ohtake, H. 1976. Respiratory behavior of sea-urchin spermatozoa. II. Sperm-activating substance obtained from jelly coat of sea-urchin eggs. *J. Exp. Zool.* 198: 313–322.

Pamilo, P., and Bianchi, N. O. 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* 10: 271–281.

Pancer, Z., Rast, J. P., and Davidson, E. H. 1999. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. *Immunogenetics* 49: 773–786.

Pearse, J. S. 1975. Lunar reproductive rhythms in sea urchins. A review. *J. Interdiscipl. Cycle Res.* 6: 47–52.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8: 785–786.

Ponting, C. P., Schultz, J., Milpetz, F., and Bork, P. 1999. SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic. Acids Res.* 27: 229–232.

Ramarao, C. S., Burks, D. J., and Garbers, D. L. 1990. A single mRNA encodes multiple copies of the egg peptide speract. *Biochemistry* 29: 3383–3388.

Repaske, D. R., and Garbers, D. L. 1983. A hydrogen ion flux mediates stimulation of respiratory activity by speract in sea urchin spermatozoa. *J. Biol. Chem.* 258: 6025–6029.

Ronquist, F., and Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Rooney, A. P., Zhang, J., and Nei, M. 2000. An unusual form of purifying selection in a sperm protein. *Mol. Biol. Evol.* 17: 278–283.

Schackmann, R. W., and Chock, P. B. 1986. Alteration of intracellular [Ca2+] in sea urchin sperm by the egg peptide speract. Evidence that increased intracellular Ca2+ is coupled to Na+ entry and increased intracellular pH. *J. Biol. Chem.* 261: 8719–8728.

Shimizu, T., Yoshino, K. I., and Suzuki, N. 1994. Identification and characterization of putative receptors for sperm activating peptide I (SAP-I) in spermatozoa of the sea urchin *Hemicentrotus pulcherrimus*. *Dev. Growth Differ.* 36: 209–221.

Smith, A. B., Pisani, D., Mackenzie-Dodds, J. A., Stockley, B., Webster, B. L., and Littlewood, D. T. J. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol. Biol. Evol.* 23: 1832–1851.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., and Gibbs, R. A. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314: 941–952.

Stamatakis, A., Ludwig, T., and Meier, H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456–463.

Sunday, J. M., and Hart, M. W. 2013. Sea star populations diverge by positive selection at a sperm-egg compatibility locus. *Ecol. Evol.* 3: 640–654.

Suzuki, N. 1995. Structure, function and biosynthesis of sperm-activating peptides and fucose glycoconjugate in the extracellular coat of sea urchin eggs. *Zool. Sci.* 12: 13–27.

Suzuki, N., et al. 1984. A peptide associated with the egg causes mobility shifts in a major plasma membrane protein of spermatozoa. *J. Biol. Chem.* 14: 14874–14879.

Suzuki, N., and Yoshino, K-I. 1992. The relationship between amino acid sequenes of sperm activating peptides and the taxonomy of echinoids. *Comp. Biochem. Physiol. B* 102: 679–690.

Swanson, W. J., and Vacquier, V. D. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* 281: 710–712.

Swanson, W. J., and Vacquier, V. D. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3: 137–144.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30: 2725–2729.

Trimmer, J. S., and Vacquier, V. D. 1986. Activation of sea urchin gametes. *Ann. Rev. Cell Biol.* 2: 1–26.

Tu, Q. R., Cameron, A., and Worley, K. C. 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res.* 22: 2079–2087.

Uehara, T., Asakura, H., and Arakaki, Y. 1990. Fertilization blockage and hybridization among species of sea urchins. M. Hoshi and and O. Yahashita, (eds.). In *Advances in invertebrate reproduction*. Elsevier, Amsterdam, pp. 305–310.

Vacquier, V. D. 1998. Evolution of gamete recognition proteins. *Science* 281: 1995–1998.

Vacquier, V. D., and Moy, G. W. 1997. The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. *Dev. Biol.* 192: 125–135.

Vacquier, V. D., and Swanson, W. J. 2007. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb. Perspect. Biol.* 3: a002931.

Vieira, A., and Miller, D. J. 2006. Gamete interaction: is it species-speciific. *Mol. Rep. Dev.* 73: 1422–1429.

Ward, G. E., Brokaw, C. J., Garbers, D. L., and Vacquier, V. D. 1985. Chemotaxis of *Arbacia punctulata* spermatozoa to resact, a peptide from the egg jelly layer. *J. Cell Biol.* 101: 2324–2329.

Wernersson, R., and Pedersen, A. G. 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* 31: 3537–3539.

Yamaguchi, M., Niwa, T., Kurita, M., and Suzuki, N. 1988. The participation of speract in the acrosome reaction of *Hemicentrotus pulcherrimus*. *Dev. Growth Differ.* 30: 159–167.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586–1591.

Yang, Z., and Nielsen, R. 2000. Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17: 32–43.

Yang, Z., and Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908–917.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A-M. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.

Yang, Z., Wong, W., S-W., and Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22: 1107–1118.

Yoshino, K. I., et al. 1990. A species-specific sperm-activating peptide from the egg jelly of the sea urchin *Diadema setosum*. *Comp. Biochem. Phsiol.* 95B: 423–429.

Zhang, J., Nielsen, R., and 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, K. S., and Lessios, H. A. 2003. Evolution of bindin in the pantropical sea urchin *Tripneustes*: comparisons to bindin of other genera. *Mol. Biol. Evol.* 20: 220–231.