High amino acid diversity and positive selection at a putative coral immunity gene (tachylectin-2)

Marshall L. Hayes  
*Cornell University*

Ron I. Eytan  
*Louisiana State University*

Michael E. Hellberg  
*Louisiana State University*

Follow this and additional works at: https://digitalcommons.lsu.edu/biosci_pubs

**Recommended Citation**  
Hayes, M., Eytan, R., & Hellberg, M. (2010). High amino acid diversity and positive selection at a putative coral immunity gene (tachylectin-2). *BMC Evolutionary Biology, 10* (1) https://doi.org/10.1186/1471-2148-10-150

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.
High amino acid diversity and positive selection at a putative coral immunity gene (*tachylectin-2*)

Marshall L Hayes¹, Ron I Eytan² and Michael E Hellberg*²

**Abstract**

**Background:** Genes involved in immune functions, including pathogen recognition and the activation of innate defense pathways, are among the most genetically variable known, and the proteins that they encode are often characterized by high rates of amino acid substitutions, a hallmark of positive selection. The high levels of variation characteristic of immunity genes make them useful tools for conservation genetics. To date, highly variable immunity genes have yet to be found in corals, keystone organisms of the world’s most diverse marine ecosystem, the coral reef. Here, we examine variation in and selection on a putative innate immunity gene from *Oculina*, a coral genus previously used as a model for studies of coral disease and bleaching.

**Results:** In a survey of 244 *Oculina* alleles, we find high nonsynonymous variation and a signature of positive selection, consistent with a putative role in immunity. Using computational protein structure prediction, we generate a structural model of the *Oculina* protein that closely matches the known structure of tachylectin-2 from the Japanese horseshoe crab (*Tachypleus tridentatus*), a protein with demonstrated function in microbial recognition and agglutination. We also demonstrate that at least three other genera of anthozoan cnidarians (*Acropora, Montastrea* and *Nematostella*) possess proteins structurally similar to tachylectin-2.

**Conclusions:** Taken together, the evidence of high amino acid diversity, positive selection and structural correspondence to the horseshoe crab tachylectin-2 suggests that this protein is 1) part of *Oculina*’s innate immunity repertoire, and 2) evolving adaptively, possibly under selective pressure from coral-associated microorganisms. *Tachylectin-2* may serve as a candidate locus to screen coral populations for their capacity to respond adaptively to future environmental change.

**Background**

Host immune systems must be able to recognize a wide range of rapidly evolving microbes; thus, functional variation is a hallmark of responses to potential pathogens and other non-self molecules. Such variation can arise via complex interactions leading to somatic recombination, as in vertebrate immune systems and molluscan fibrinogen-related proteins [1], or more simply via genetic diversity in the host immune system, either at the level of families of genes or alleles at a single locus. Genes used by potential hosts to distinguish self from non-self and to recognize and defend against pathogenic microbes include some of the most genetically variable known, among them those encoding the Major Histocompatibility Complex (MHC) proteins in vertebrates [2] and a histocompatibility protein in tunicates [3], as well as disease resistance (*R*) proteins [4] and ribonucleases of gametophytic self-incompatibility (GSI) [5] in plants.

High levels of variation are generated by diversifying selection that results from either selection favoring heterozygotes or from frequency-dependent selection favoring rare alleles. A molecular signature of positive selection (that is, an excess in the nonsynonymous nucleotide substitution rate dN relative to the synonymous rate dS when compared to neutral expectations) is therefore another distinctive feature of host immunity genes. Indeed, one of the first uses of the now-standard dN/dS ratio approach for detecting positive selection was on sequence data from the MHC binding cleft of mice and...
humans [6]. Subsequent work on natural populations of vertebrates has detected positive selection at MHC many times [7], as well as on other immunity genes from vertebrates [8] and plants [9,10].

Because high variation at immunity genes appears to be sustained by selection, surveys of intraspecific variation at such loci have been employed in ways that neutral markers cannot be. For example, the signature of positive selection at MHC may be difficult to maintain if historical population sizes were small enough for drift to dominate, a situation that can be evaluated by comparing patterns at MHC loci and highly variable, but presumably neutral, nuclear markers such as microsatellites [11]. More generally, surveys of variation at MHC loci have taken on a special role in a conservation context, where they have been seen as proxies for levels of standing adaptive variation in wild populations [12,13]. Such assessments of a population’s ability to withstand challenges from pathogens should hold true to the extent that variation at immunity genes relates to an organism’s fitness and ability to fend off pathogens and parasites, as has been seen for MHC in some vertebrates [14-16].

In contrast to the many works on MHC variation in wild vertebrate populations, findings of immunity gene variation in invertebrate animals are only beginning to emerge. In part, such surveys have been stymied by the heterogeneous ways in which the immune systems of different phyla work [17]. Schulenburg et al. [18] also suggested that invertebrate immunity genes may be limited in their genetic variability, but recent studies suggest otherwise. Exceptionally high diversity has been recorded recently in invertebrate immunity genes, both among paralogous members of the same gene family (e.g. in sea urchins [19]; in nematodes: [20]) and among alleles at a single locus (e.g. mosquitoes: [21]; in mussels: [22]). High polymorphism has not been accompanied universally by positive selection however [23], although most studies have not tested for it.

Corals (Phylum Cnidaria, Class Anthozoa, Order Scleractinia) are among the many taxa for which we have no information on the population-level variability of immunity genes. Reef-building (hermatypic) corals are especially suitable targets for studies of immunity gene variation for several reasons. Reef corals create habitat that sustains large numbers of other species, yet these corals have been in decline globally in the recent past [24]. Coincident with this decline has been a rise in reported coral diseases [25,26]. These have often been associated with high water temperature anomalies (e.g. [27]), although there are claims for mechanistic ties between coral bleaching and bacteria [28,29]. The long generation times of many large hermatypic corals would seem to make them especially vulnerable to rapidly evolving microbes with generation times perhaps $10^7$-$10^8$ shorter. The search for polymorphic recognition genes from corals that could effectively match the diversity of potential pathogens is just beginning. One carbohydrate recognition protein (Millectin) from a coral binds both bacterial pathogens and algal symbionts [30], but of the several similar isoforms reported, all came from a single individual, suggesting that they constituted a recently radiating gene family and were non-allelic. A polymorphic kin-recognition locus has been characterized in a hydroid [31], but no such highly variable region has been reported from corals, nor has any coral immunity gene been characterized for intraspecific variation to date.

Here, we characterize sequence variation at a putative coral immunity gene, tachylectin-2, fortuitously identified while generating markers from an EST library [32] from Oculina, a genus that has served previously as a model for studies of coral disease and bleaching [28]. Tachylectin-2 was originally isolated from the Japanese horseshoe crab (Tachypleus tridentatus) and has since been demonstrated experimentally to possess anti-microbial activity [33]. Its crystal structure is unique and composed of a five-bladed β-propeller [34], each blade of which potentially binds N-acetyl sugars such as those associated with lipopolysaccharide (LPS) and peptidoglycan found in the bacterial cell wall.

Homologs of tachylectin-2 have been reported from two cnidarians, although in both cases, functional roles other than host immunity have been proposed. Mali et al. [35] isolated a gene (CTRN) from the hydrozoan Hydractinia echinata that had a primary sequence with a similarity of over 30% to horseshoe crab tachylectin-1 and a typical tachylectin repeat structure. However CTRN was expressed solely in circumoral neurons and, most importantly, expression levels were not induced when challenged with LPS. This suggests that the primary function of CTRN is neither host immunity nor some other form of microbial recognition. Schwarz et al. [36] identified a homolog of horseshoe crab tachylectin-2 via BLASTX screening of an EST library generated from the coral Montastrea faveolata. They speculated that such a protein could mediate interactions between coral hosts and their algal symbionts and suggested further study to investigate this possibility.

To confirm the identity of the Oculina gene, we show that the inferred structure of the Oculina tachylectin-2 corresponds closely to the solved crystal structure of the horseshoe crab tachylectin-2. We also report high nonsynonymous variation for the Oculina tachylectin-2 and demonstrate that positive selection has promoted allelic diversity at this locus. Our results are consistent with adaptive diversification at a host immunity gene, and thus open the door to molecular studies of host susceptibility and population vulnerability in corals.
Results
Cnidarian Homologs of Tachylectin-2 from Tachypleus tridentatus

Pairwise alignments of putative tachylectin-2 amino acid sequences from four cnidarians (three corals and an anemone) and the full-length tachylectin-2 from the Japanese horseshoe crab reveal a high degree of similarity among the proteins, beginning at position 54 of the horseshoe crab tachylectin-2 (Figure 1, Table 1). The secondary structure prediction corroborates the localization of our alignment and indicates that the *Oculina* sequence corresponds to a region spanning nearly two complete tandem repeats in the horseshoe crab tachylectin-2, including two β-strands within β-sheet II, all four strands of β-sheet III and the first two strands of β-sheet IV with each β-sheet being separated by a single α-helix (Figure 1, [34]). Furthermore, consistent with the characteristics of a 5-bladed β-propellor, the *Oculina* primary structure is comprised of highly similar (70% identity) tandem repeats of 40 amino acids each (*LYGXXDKFXYRX-PPTHXSDNLGSAXXIGGXGWXXFXXL*, Figure 1).

Within each of its five equivalent β-sheets, the horseshoe crab tachylectin-2 harbors an individual N-acetyl sugar-binding site comprised of 8 functional residues capable of direct interaction with target ligands. The partial *Oculina* sequence may thus contain 16 binding residues, including a complete 8-residue binding pocket. Moreover, these 16 binding residues are highly conserved between the two proteins. Ten residues are identical to the horseshoe crab tachylectin-2 sequence, while 3 are conserved substitutions (M23L, I70L and V76I), and 3 are semi-conserved (D18A, N31T and D65G).

As an additional test of structural correspondence, we generated three-dimensional (3-D) structural models for the *Oculina*, *Acropora*, *Montastrea* and *Nematostella* homologs and used a threading approach to align the models with the known X-ray crystal structure for tachylectin-2 [PDB:1tl2] from horseshoe crab. For each of the four models, the estimated precision, or percent likelihood of a correct threading match based on e-values, is 100%. Furthermore, respective RMS deviations resulting from the superposition of the *Oculina*, *Acropora*, *Montastrea* and *Nematostella* models onto 1tl2 are 0.42 Å (on 372 atoms), 0.67 Å (on 352 atoms), 0.65 Å (on 324 atoms) and 0.41 Å (on 896 atoms) (Figure 2). Very similar, well-aligned structures generally have RMS deviations <1.0 Å, indicating that all four of our alignments are precise. Divergence among these models and 1tl2 is minimal and limited to the C-terminal extensions in the particular cases of *Oculina* and *Acropora*. Thus, a phylogenetically wide range of corals possesses tachylectin-2-like proteins that are very similar to and align well with the horseshoe crab tachylectin-2.

Positive Selection

Diversity at *tachylectin-2* as indicated by \( \pi \) (the number of nucleotide differences per site between two randomly chosen sequences) was high overall (\( \pi = 0.0179 \)), and was evident at both silent (\( \pi_{\text{syn}} = 0.0299 \)) and replacement sites (\( \pi_{\text{rep}} = 0.0145 \)). Waterson's theta (\( \theta_W \)) was 0.0215. The maximum likelihood analysis recovered a single best tree (Figure 3), although support for nodes was low (as expected for intraspecific data from species lacking strong phylogeographic structure). The neighbor joining tree (Figure 3) supported just three nodes with bootstrap values >50%, and thus was essentially a star phylogeny.

Likelihood ratio tests suggest that positive selection promotes nonsynonymous substitutions among our sample of *tachylectin-2* alleles from *Oculina* (Table 2). The strength of this conclusion varies depending on which allele tree the tests are based upon. Both tests (M7 vs. M8 and M1a vs. M2a) based on the collapsed neighbor joining (cNJ) tree were highly significant. These tests test indicated that while only a small proportion of residues were under selection (about 5.5%), those residues were under very strong selection, with the ratio of nonsynonymous to synonymous substitution rates (\( \omega \)) > 9. Based on the single best maximum likelihood tree, the M7 (where \( \omega \) can take on a range of values between 0 and 1) versus M8 (as for M7, but with an additional positive selection

| Host species               | Accession Number | sequence length (aa) | % identity/similarity to *Oculina* sequence | % identity/similarity to 1tl2 aa 54-140 |
|----------------------------|------------------|----------------------|---------------------------------------------|--------------------------------------|
| *Oculina varicosa*         | FJ966784         | 92                   | 100/100                                     | 60/68                                |
| *Montastrea faveolata*     | FE038913         | 111                  | 54/60                                       | 44/62                                |
| *Acropora millepora*       | EZ038328         | 277                  | 76/85                                       | 52/64                                |
| *Nematostella vectensis*   | EDO38290         | 242                  | 52/67                                       | 55/71                                |

\(^a\) Local alignment of *M. faveolata* is with 1tl2 aa 54-135.

\(^b\) Local alignment of *A. millepora* with 1tl2 aa 54-140 is presented here, although the strongest alignment is with 1tl2 aa 7-95.

\(^c\) Local alignment of *N. vectensis* is with 1tl2 aa 54-145.
category where ω can exceed one) test was significant. The test comparing M1a (with just two ω categories: neutrality (ω = 1) and another between 0 and 1) to M2a (as for M1a, but with an additional positive selection category where ω > 1) was not significant (p = 0.081). For these tests, a higher proportion (about 24%) of codons were inferred to fall in the class experiencing positive selection, but the ω for this class was not as high as for the collapsed NJ analysis (ω = 2.37).

Consistent with the estimated proportions of residues under selection (Table 2), the BEB analysis for sites under positive selection (Table 3) based on the cNJ tree indicated fewer sites (just 5) under stronger selection than the analysis based on the maximum likelihood tree (which indicated 11 residues). One of the sites that was significant for the analysis based on the cNJ tree was not significant for the analysis based on the ML tree. Thus, a total of 12 residues were flagged as potentially evolving under positive selection (Figure 4).

An in-frame indel occurs in the Oculina alignment. The alternate resolution of the alignment without the indel results in five consecutive nucleotide substitutions and two non-synonymous replacements (data not shown). The in-frame indel used in our alignment, by eliminating non-synonymous substitutions, results in a more conservative test for positive selection by removing two sites from consideration in the PAML analyses.

Tajima’s D (-0.680) and Fu and Li’s D (-2.17) were both negative, consistent with an excess of low frequency polymorphisms, but these values were not statistically significant. Fay and Wu’s H (0.631) was also not significant.

Tests for recombination generally did not reveal significant results. The DSS test implemented in TOPALi did not detect recombination in the tachylectin-2 alignment.

---

**Figure 1** Amino acid sequence alignment of conserved regions of the horseshoe crab tachylectin-2 (1tl2), and tachylectin-2s from the anthozoans *Oculina*, *Montastrea*, *Acropora* and *Nematostella*. Black boxing indicates conserved sites, relative to the *Oculina* sequence. Red text highlights those positions in 1tl2 that are functional N-acetyl-sugar-binding sites [34]. The locations of the 40-amino acid tandem repeats characteristic of tachylectin-2 are indicated by the blue lines. Below the sequences, the predicted secondary structural elements of the *Oculina* tachylectin are shown relative to elements of the experimentally derived secondary structure of 1tl2.
Two tests for recombination were implemented using the Datamonkey server, first looking for a single recombination event (SBP test) and second for multiple recombination events (GARD test). Neither test detected any recombination. The four-gamete test detected six recombination events; the estimate of \( R \) per gene was 128.0.

Spatial and Functional Correspondence of Positively Selected Sites in Oculina Tachylectin-2

The results of the Bayes Empirical Bayes (BEB) analysis (Table 3, Figure 4) indicated that 12 codons in *Oculina* tachylectin-2 may have been subject to positive selection. To determine whether the spatial organization of these sites corresponded to regions of functional importance, positively selected sites were first mapped onto the threaded structural model. In the space-filling representation of the modeled molecular surface (Figure 5), all 12 positively selected sites are localized on the protein's exterior and are well exposed. This observation is independently confirmed by calculations of solvent-accessible surface area using the GETAREA algorithm and 10% solvent accessibility as a cut-off value.

Next, we tested whether the spatial distribution of positive selection was clustered, or otherwise nonrandom, in 3-D space by comparing observed codon positions to random permutations of an equal number of surfaces sites along the same length of protein. No significant trend in selected site clustering was observed. Indeed, selected sites appear to occur along the entire length of \( \beta \)-strands, \( \alpha \)-helices and coiled coils present in the *Oculina* tachylectin-2 (cf. Figure 2, Figure 4).

Finally, we examined whether the 12 positively selected sites correspond to active N-acetyl sugar-binding residues or are otherwise directly associated with the known functional regions of the horseshoe crab tachylectin-2. Based on our sequence alignment and the degree of conservation at known functional positions in 1tl2, no positively selected site corresponded to any active site within an N-acetyl sugar-binding pocket.

Discussion

**A Recognition Function for Cnidarian Tachylectin-2 Homologs**

Since the initial discovery of tachylectins in the Japanese horseshoe crab *Tachypleus tridentatus*, tachylectin-like proteins have been identified in a wide range of organisms including slime molds [37], sponges [38], hydrozoan cnidarians [35], lancelets [39] and fish [40,41]. This coverage has been recently extended to anthozoan cnidarians on the basis of genome sequence from a sea anemone [42] and EST sequences from three corals [32,36,37]. Thus, tachylectin-2-like proteins are present in at least two cnidarian classes (Hydrozoa and Anthozoa) as well as in both major clades of scleractinians, the complex (e.g. *Acropora*) and the robust (e.g. *Montastrea, Oculina*) [43].

Proteins of the tachylectin family have broad-spectrum anti-microbial activity [44] and play a role in non-self recognition [45], two functions central to cnidarian innate immunity [46,47]. Alignments and structural modeling of translated amino-acid sequences show that our *Oculina* gene encodes a protein homologous to tachylectin-2. Although our characterization of the *Oculina tachylectin-2* is limited because we have yet to identify the start site for the gene’s coding region, several lines of circumstantial evidence suggest that the partial *Oculina* gene encodes an ortholog of *tachylectin-2*. First, based on pairwise sequence comparisons (Figure 1) and protein threading results (Figure 2), the sequences from *Oculina* and *Tachypleus* are highly similar. Second, partial amino-acid sequences from *Oculina* reveal an internal homology that consists of two highly similar tandem repeats, a feature characteristic of protein regions that arrange pseudo-symmetrically to form larger structures comprised of multiple \( \beta \)-propeller folds. Lectins, and tachylectins among them, characteristically derive conformational rigidity, multivalency, and avidity from these \( \beta \)-propeller structures [48,49]. Finally, the predicted secondary structure (i.e. \( \alpha \)-helices, \( \beta \)-strands and coiled coils) of the *Oculina* protein corresponds closely to the experimentally derived X-ray structure of 1tl2 (Figure 1).

Modeling the structure of *Oculina* tachylectin-2 enables us to map the location of residues found to be under positive selection on the 3-D structure of the pro-
tein. Without exception, replacements have occurred at well-exposed sites on the protein's surface. This external localization most likely stems from surface residues having lower structural constraints than their buried counterparts (e.g. [50,51]). Thus, positive selection does not appear to have altered the structure and stability of the core of Oculina tachylectin-2, nor has selection targeted residues known to be involved in carbohydrate recognition and binding. These have been conserved, suggesting that the molecular affinity of the Oculina tachylectin-2 for N-acetyl sugars remains unchanged. However, given that positive selection acts on residues on the surface of the Oculina protein, especially those at the fringes of the N-acetyl sugar-binding pockets (cf. Figure 5 here and Figure 5 from [34]), these selected changes may modulate the architecture, flexibility and specificity of the binding pocket to accommodate a broader range of ligand sizes and configurations from a more diverse pool of non-self sources. Such correlations between structure and activity have been previously described for other classes of lectin (e.g. for galectins, [52]).

Positive Selection on Oculina Tachylectin-2

In our survey of Oculina tachylectin-2 alleles, we found high nonsynonymous variation (25 alleles differing by amino replacements in a survey of 244 alleles) promoted by positive selection (Table 3, Figure 4). Because high variation and positive selection are disproportionately observed in immunity genes relative to other functional gene classes [53], the presence of these signatures is consistent with Oculina tachylectin-2 performing some aspect of coral immunity (even though some immunity genes do not reveal such a signal, e.g. [54]). Experimental assessments of proposed immune functions are critical both to confirming the involvement of this protein in non-self recognition or some other form of coral-microbe
interaction based on N-acetyl sugar specificity, and to identifying the selective forces driving the evolution of tachylectin-2.

An alternate explanation to a purely immune explanation for positive selection on tachylectin-2 is that the protein may mediate a more general process of non-self recognition, limited not only to pathogen detection but also including symbiont identification. Indeed, in the best-characterized host-symbiont interactions (that between legumes and rhizobia), De Mita et al. [55,56] found positive selection in the legume Medicago truncatula at the NODULATION RECEPTOR KINASE (NORK) gene, which functions during the early stages of root infection by symbiotic nitrogen-fixing bacteria and endomycorrhizal fungi. While the authors propose that Medicago-rhizobium interactions may co-evolve in a manner reminiscent of host-pathogen co-evolution, this interpretation is a matter of current debate [57]. The signature of positive selection has also appeared in a study trying to elucidate the molecular underpinnings of specificity among corals and their dinoflagellate symbionts in terms of lectin-mediated interactions. Voolstra et al. [58] presented preliminary evidence of positive selection at a dinoflagellate gene encoding a novel protein of unknown function although, unlike our study, this conclusion of positive selection was not based on allelic variation. Until more studies conclude that adaptive evolution acts to diversify alleles at host symbiont recognition genes, the positive selection seen among alleles at Oculina tachylectin-2 should be considered more consistent with a role in immunity than symbiosis.

In our earlier work on several nominal Oculina species [32], we employed three nuclear loci (tachylectin-2, as well as two others tentatively identified using BLASTP as fatty acid elongase and elongation factor 1α) as markers to reveal subdivision among 10 different sampling populations spanning a 2370 km range and also to establish the genetic isolation of a threatened deep-water population (Jeff’s Reef on the Oculina Banks off central Florida). In the latter case, the tachylectin-2 locus was unusual both in being fixed for a single allele and in being putatively associated with host immunity. This observation adds to the ongoing debate over the adaptive potential of reef building corals in the face of global climatic change [59,60]. On the one hand, most Oculina populations segregate ample genetic variation and show the signature of positive, diversifying selection at the tachylectin-2 gene, demonstrating that genotypic diversity and adaptive variation are present. On the other, the deep-water population, presently threatened by illegal trawling [61], is fixed for a single tachylectin-2 allele, which may indicate its future prospects are even bleaker than demography suggests.

That functional genetic variation for host resistance exists within coral populations has been shown recently for another Atlantic species, Acropora cervicornis. Vollmer and Kline [62] showed that there are differences in resistance to white-band disease resulting from in-situ transmission assays among different A. cervicornis genotypes. Genotypes immune to white-band disease, presently threatened by illegal trawling [61], is fixed for a single tachylectin-2 allele, which may indicate its future prospects are even bleaker than demography suggests.

The geographic variation in allelic diversity and positive selection in our data are also intriguing in light of

### Table 2: Tests for positive selection on putative tachylectin-2 from Oculina based on maximum likelihood (ML) and collapsed Neighbor Joining (cNJ) allele trees

| Models compared (allele tree) | 2ΔL     | P1    | ω     |
|-------------------------------|---------|-------|-------|
| M1a vs. M2a (ML)              | 5.02    | 23.9  | 2.37  |
| M7 vs. M8 (ML)                | 6.18*   | 23.9  | 2.37  |
| M1a vs. M2a (cNJ)             | 95.2**  | 5.6   | 9.38  |
| M7 vs. M8 (cNJ)               | 96.8**  | 5.7   | 9.36  |

* - significant at p = 0.05
** - significant at p < 0.001

p1 - percentage of sites under positive selection (under model M8)
ω - ratio of dN/dS for those sites under positive selection (model M8)
recent studies demonstrating that allelic fixation may provide a signature of geographically variable selection in isolated populations [63]. For example, in a geographical survey of variation at 16 immunity genes in six natural populations of teosinte, Moeller and Tiffin [64] found a marked difference in allelic variation at a single locus, the wound-inducible serine protease inhibitor (wip1), in a single population. Taken along with other observations (for example, this population was fixed for a replacement at an active site), wip1 appeared to have undergone a population-specific selective sweep. The results of Moeller and Tiffin also suggest that such signatures are relatively rare, even among immunity genes. Considering the special importance in variation at immunity genes in evaluating population viability for conservation purposes [13], it would be interesting to explore patterns of interpopulation variation in Oculina or other scleractinians using a more comprehensive set of candidate loci including both immunity and non-immunity genes.

Given that our inference of positive selection was based on allelic sequences, the possible role of recombination in creating this signal deserves consideration. Tests for recombination were equivocal. On one hand, the results of the four-gamete test [32] and values for $R$ indicate multiple recombination events and a high per-locus recombination rate. This rate (128) is high enough to cause false positives in site-based tests of selection, especially the M7 vs. M8 test [65]. On the other hand, values for $R$ have been shown to be biased upwards [65,66] and are sensitive to infinite-site violations, which are present in our dataset. Further, neither the DSS test nor the GARD test detected multiple recombination events, even though both are designed to do so. Even if we allow for the presence of one recombination event, this does not appear to be a high enough rate of recombination to cause false positives in the site-based tests of selection [67,68]. The four-gamete test determines whether all four combinations of a pair of variable sites are present in a sample [69]. If so, then recombination is inferred. However, incompatible sites can arise through recurrent mutations as well as recombination. For data sets where the mutational process is best described by any time-reversible mutation model (as ours is), incompatible sites will arise, causing the four-gamete test to be overly conservative (in that it may flag recombination even when it is not present). Indeed, one tachylectin-2 site was dropped from the four-gamete test because it segregates for more than two nucleotides, reinforcing the argument that an infinite sites model is not appropriate for our data.

**Conclusion**

The combination of two lines of evidence suggests the gene region examined here is involved in coral innate immunity. First, both primary amino acid sequence and inferred protein structure are similar to those of tachylectin-2, a protein from horseshoe crabs associated with an innate immune response to bacterial pathogens. Second, the divergence of alleles at tachylectin-2 has been promoted by positive selection, a hallmark of immunity genes. This proposed immunity function should be confirmed experimentally. Nonetheless, given the conservation threats faced by many corals and the special role of genetic variation at immunity genes such as the vertebrate MHC in assessing population viability, the gene encoding tachylectin-2 in corals may serve as a candidate locus to screen anthozoan populations for their potential to respond adaptively to future challenges. Neither widespread assessments of immunity loci nor comparisons of immunity to non-immunity genes have been conducted for corals as they have for other hosts (e.g. [63,70-72]). Combining the identification of additional coral immunity genes with geographic surveys and detailed studies of the functional consequences of naturally occurring variation should provide insights into both how corals defend...
themselves against natural enemies and how better we can preserve these key components of marine biodiversity.

Methods

Protein Alignments, Threading, and Structural Analysis of Positively Selected Sites

A tachylectin-2-like sequence was first identified during random sequencing of an *Oculina varicosa* cDNA library [32]. A BLASTX survey of GenBank, using the *Oculina* sequence as a query, identified a putative tachylectin from the larval transcriptome of the reef-building coral, *Acropora millepora* [GenBank:EZ038328] [73]. A *Montastrea faveolata* homolog [EF038913], previously described by Schwarz et al. [36] in an EST screen, was retrieved from GenBank. Finally, a *Nematostella vectensis* homolog [EFO38290] was located in the JGI genome database http://genome.jgi-psf.org/Nemve1/Nemve1.home.html using conserved motifs from the *Oculina* sequence as a BLASTX query. Protein sequences were aligned using MegAlign (DNASTAR) for multiple alignments and EMBoss http://www.ebi.ac.uk/Tools/emboss/align/index.html for pairwise alignments.

Three-dimensional (3-D) structures of tachylectin-2 based on amino acid sequences from *Oculina, Acropora, Montastrea* and *Nematostella* were modeled using computational tools for structure prediction available in the PHYRE Protein Homology/analogY Recognition Engine version 0.2 ([74]; http://www.sbg.bio.ic.ac.uk/phyre/). The PHYRE protocol incorporates four discrete processing steps into one program interface: 1) generation of a protein profile from the user-provided sequence using iterative PSI-Blast to identify both close and remote sequence homologs, 2) prediction of secondary structure by three independent programs (Psi-Pred, SSPro and JNet), 3) prediction of ordered and disordered regions using Disopred, and 4) fold recognition via the application of a profile-profile alignment algorithm to a reference fold library [75]. The resulting product is a downloadable model with associated confidence estimates relating the model to existing structures in the Structural Classification of Proteins (SCOP) and Protein Data Bank (PDB) databases.

Models for tachylectin-2s from *Oculina, Acropora, Montastrea* and *Nematostella* were visualized and further manipulated using DEEPVIEW version 4.0.1 ([76]; http://spdbv.vital-it.ch/). The PDB file of the *T. tridentatus* tachylectin-2 [PDB:1tl2] was downloaded from the RCSB Protein Data Bank and similarly viewed in DEEPVIEW. Tachylectin-2 sequences from *Oculina, Acropora, Montastrea* and *Nematostella* were structurally superposed on 1tl2 using DEEPVIEW’s Iterative Magic Fit function, and the resulting quality of fit was computed as a Root Mean Squared (RMS) deviation in angstroms. To determine the degree to which positively selection sites were either buried or on the protein’s surface, solvent accessible surface area for individual residues of *Oculina* tachylectin-2 mapped onto 1tl2 was calculated using GETAREA version 1.0 beta ([77]; http://curie.utmb.edu/getarea.html). Finally, statistical significance of the spatial clustering of positively selected amino acids and correspondence with functional binding sites was assessed using programs from Clark and Swanson [78].

Tests for Positive Selection

To test for positive selection acting on *Oculina tachylectin-2*, we assembled a dataset consisting of a subset of our population genetic data (Table 4) from Eytan et al. [32]. The 3’ UTR was cut from the original alignment, yielding

| Residue # | ML | cNJ |
|-----------|----|-----|
| 8         | 0.91| 1.0 |
| 10        | 0.92|     |
| 11        | 0.54|     |
| 13        | 0.84|     |
| 14        | 0.55|     |
| 21        | 0.69|     |
| 27        | 1.0 |     |
| 48        | 0.52|     |
| 51        | 0.90| 0.97|
| 53        | 0.90| 1.0 |
| 54        | 0.98| 1.0 |
| 71        | 0.53|     |

The *p*-values for the M7 vs. M8 likelihood ratio tests are shown, as are the sites determined to be under positive selection by Bayes Empirical Bayes analysis based on two allele trees (maximum likelihood, ML, and collapsed neighbor joining, cNJ) and their probabilities of belonging to the group of residues undergoing positive selection.
40 non-identical alleles (each 276-bp in length), 25 of which differed by non-synonymous substitutions.

CODEML, implemented in PAML v4.2 [79] was used to conduct tests for selection using the site models, which allow the omega ratio to vary among codon sites. Two pairs of null and alternate models were used that provide two likelihood ratio tests (LRTs) for positive selection. The first compares the null model of nearly neutral molecular evolution (the M1a model) to the alternate model of positive selection (M2a) [80]. The second compares a model of beta-distributed variable selection pressure (M7) to an alternate model of beta-distributed variable selection pressure plus positive selection (M8) [81]. The LRT employing two degrees of freedom was used to determine if positive selection was present. Codon sites under positive selection were determined using the Bayes Empirical Bayes analysis [82], which has been shown to be both powerful and not excessively prone to false positives [80,82].

A phylogenetic tree is required for the site-based tests of selection implemented in PAML. Before constructing the tree, a model of sequence evolution was determined with jModelTest v0.1.1 [83] using the AIC to choose between models. The TPM2+I model was selected as the appropriate model. A heuristic tree search using maximum likelihood was implemented in GARLI [84]. Starting trees were obtained via stepwise addition. 100 random-addition sequence replications were performed with TBR branch-swapping. Optimal trees from each repetition were saved. The best tree from all the replicates was then used for tests of selection in PAML. Although the ML tests for positive selection implemented by PAML are generally robust to tree topology [85], we also ran the analysis for an alternative Neighbor Joining allele tree (implemented in PAUP* v4b10 [86]) to insure any inference of positive selection were not overly dependent on a single topology.

We also tested for selection using summary statistic-based methods. Tajima’s D, Fu and Li’s D, and Fay and
### Table 4: Sequences used in this study

| GenBank Acc. # | Frequency | Populations where present |
|---------------|-----------|---------------------------|
| FJ966794      | 3         | NC                        |
| FJ966795      | 99        | NC(21), GA(24), JAX(16), DAY(13), FtP(7), HSH(2), CFL(5), SAR(7), PAN(4) |
| FJ966802      | 16        | NC(2), GA(2), DAY(4), SAR(8) |
| GQ917185      | 2         | GA                        |
| FJ966742      | 2         | GA                        |
| FJ966864      | 17        | GA, FtP, HSH(2), SAR(2), PAN(11) |
| FJ966775      | 3         | JAX                       |
| FJ966781      | 1         | JAX                       |
| FJ966731      | 3         | JAX, DAY, SAR             |
| FJ966715      | 8         | JAX, FtP(2), HSH(3), CFL, PAN |
| GQ917186      | 1         | DAY                       |
| GQ917187      | 3         | DAY(2), PAN               |
| GU827980      | 1         | DAY                       |
| FJ966724      | 1         | DAY                       |
| FJ966729      | 2         | DAY                       |
| FJ966733      | 2         | DAY                       |
| FJ966730      | 7         | DAY(2), FtP, HSH, SAR, PAN(2) |
| FJ966734      | 2         | DAY, HSH                  |
| FJ966812      | 25        | FtP(10), HSH(5), CFL(7), PAN(3) |
| GU827979      | 3         | FtP, CFL, PAN             |
| FJ966816      | 1         | FtP                       |
| FJ966809      | 1         | FtP                       |
| GQ917188      | 3         | FtP, HSH, PAN             |
Table 4: Sequences used in this study (Continued)

| Sequence     | Frequency | Locality |
|--------------|-----------|----------|
| FJ966814     | 1         | FtP      |
| FJ966824     | 1         | FtP      |
| FJ966825     | 1         | FtP      |
| FJ966826     | 1         | FtP      |
| FJ966827     | 1         | FtP      |
| FJ966763     | 1         | HSH      |
| FJ966709     | 1         | CFL      |
| FJ966714     | 2         | CFL      |
| FJ966717     | 2         | CFL      |
| FJ966843     | 1         | SAR      |
| FJ966845     | 1         | SAR      |
| FJ966846     | 1         | SAR      |
| FJ966850     | 1         | SAR      |
| FJ966862     | 1         | SAR      |
| FJ966833     | 1         | PAN      |
| GU827978     | 1         | PAN      |
| *FJ966784    | 20        | JR80     |

Total 244

Frequency indicates the number of allele copies encountered over a geographic survey of 244 alleles from 10 different populations [32]. * indicates the fixed allele from the deep water (80 m) population at Jeff’s Reef (JR80). Locality abbreviations: NC (North Carolina), GA (Georgia), JAX (Jacksonville, Florida), DAY (Daytona Beach, Florida), FtP (Fort Pierce, Florida), HSH (Horseshoe Reef, Florida), CFL (Cape Florida), SAR (Sarasota, Florida), PAN (Panama City, Florida).

Wu’s H were all implemented in DNAsp v5.0 [87]. To conduct Fay and Wu’s H test, a single *tachylectin-2* sequence from the coral *Solenastrea hyades* (GenBank: FJ966866) was used as an outgroup, and significance was tested using coalescent simulations (without recombination) as implemented in DNAsp. Summary statistics for levels of variation (π and θW) were also calculated using DNAsp.

Recombination events can cause the failure of tree-based tests for detecting non-neutral evolution of codons [65], with high rates of false-positives being particularly problematic [67]. We tested for the presence of recombination in the *tachylectin-2* alignment using several different methods. The tests that we used were: the DSS method [88] implemented in TOPALi v2.5 [89], the SBP and GARD methods of Kosakovsky-Pond *et al.* [68] implemented online via the Datamonkey webserver [90],...
and the four-gamete test [91] implemented in DNAsp v5.0. The preferred [68] corrected AIC criterion was used to evaluate the significance of these tests. In addition, we calculated Hudson’s $R$, the per-locus population scaled recombination rate [91], also implemented in DNAsp v5.0.

Authors’ contributions
All authors carried out the molecular genetic studies and drafted the manuscript. RIE performed the phylogenetic analyses and statistical analyses of variation and positive selection. MLH conducted the protein modeling and visualization. MEH conceived of the study and coordinated its design. All authors read, edited and approved the final manuscript.

Authors’ information
MLH applies molecular genetics, functional genomics and biochemical approaches to the study of host-microbe interactions, using plants, invertebrates and bacteria as model organisms. RIE is interested in the origin and maintenance of marine biodiversity, particularly in coral reef taxa. MEH is broadly interested in population isolation, speciation, and molecular evolution in marine animals, especially anthozoans and gastropods.

Acknowledgements
Nathan Clark conducted several statistical analyses and also contributed helpful commentary on the manuscript. Sequencing efforts beyond those referenced in [32] were partially supported by the National Science Foundation (OCE-0550270 to MEH and Iliana Baums).

Author Details
1Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, USA and 1Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

Received: 29 September 2009 Accepted: 19 May 2010
Published: 19 May 2010

References
1. Zhang SM, Adema CM, Kepler TB, Loker ES: Diversification of Ig superfamily genes in an invertebrate. Science 2004, 305:251-254.
2. Peitney SB, Oliver MK: The evolutionary ecology of the major histocompatibility complex. Heredity 2006, 96:7-21.
3. De Toma D, Nyholm SV, Palmer JI, Ishizuka KJ, Ludington WB, Wang JH, Weisshaar J, Willmer CM, Marsh LP, Nystrom M, Palumbi SR, Pandolfi JM, Rosen B, Roughgarden J: Climate change, human impacts, and the resilience of coral reefs. Science 2003, 301:929-933.
4. Parma melis A, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nikoljdo C, Simard F, Caccone A, Powell JR: The molecular evolution of four anti-malarial genes in the Anopheles gambiae species complex. BMC Evol Biol 2008, 8:79.
5. Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, Grosberg R, Hoegh-Guldberg O, Jackson JBC, Kleypas J, Lough JM, Marshall P, Nystrom M, Palumbi SR, Pandolfi JM, Rosen B, Roughgarden J: Climate change, human impacts, and the resilience of coral reefs. Science 2003, 301:929-933.
6. Harvell CD, Kim K, Burkholder JM, Colwell RR, Epstein PR, Grimes DJ, Hofmann EE, Lass EK, Osterhaus ADME, Overstreet RM, Porter JW, Smith GW, Vasta GR: Emerging marine diseases - climatic links and anthropogenic factors. Science 1999, 285:1505-1510.
7. Mykytliodz, Jones LE, Harvell CD: Inmate immunity, environmental drivers, and disease ecology of marine and freshwater invertebrates. Annu Rev Ecol Evol Syst 2006, 37:251-288.
8. Bruna S, Selig ER, Casey KM, Page CA, Willis BL, Harvell CD, Swannam H, Meendy AH: Thermal stress and coral cover as drivers of coral disease outbreaks. PLoS Biol 2007, 5:1229-1227.
9. Rosenberg E, Falkovitz L: The Vibrio shiloi/Oculina patagonica model of coral bleaching. Annu Rev Microbiol 2004, 58:143-159.
10. Rosenberg E, Kushmaro A, Kramarsky-Winter E, Banin E, Loya Y: The role of microorganisms in coral bleaching. ISME J 2009, 3:139-146.
11. Kvennefors ECE, Leggata W, Hoegh-Guldberg O, Degnan BM, Barnes AC: An ancient and variable mannose-binding lectin from the coral Acropora millepora binds both pathogens and symbionts. Dev Comp Immunol 2008, 32:1582-92.
12. Nicotra ML, Powell AE, Rosengarten RD, Moreno M, Grimmer J, Lakkis FG, Dellaporta SL, Buss LW: A hypervariable invertebrate allodatum. Curr Biol 2009, 19:583-589.
13. Eyton RJ, Hayes M, Arbouet-Reilly P, Miller M, Hellberg ME: Nuclear sequences reveal mid-range isolation of an imperiled deep-water coral population. Mol Ecol 2009, 18:2375-2389.
14. Okino N, Kawabata S, Saito T, Hirata M, Takagi T, Iwamango S: Purification, characterization, and cDNA cloning of a 27KDa lectin (L10) from horseshoe crab hemocytes. J Biol Chem 1995, 270:31008-31013.
15. Biesel HG, Kawabata S, Iwamango S, Huber R, Bode W, Tachylectin-2: crystal structure of a specific GlcNAC/GalNAc-binding lectin involved in the
innate immunity host defense of the Japanese horseshoe crab

_Tachypleus tridentatus._ EMBIO 1999; 18:2315-2322.

35. Muli B, Soza-Reed J, Frethme, M Frank U. Structural but not functional conservation of an immune molecule: a tachylectin-like gene in _Hypotrypanum._ Develop Comp Immunal 2006; 30:275-281.

36. Schwa J, Brooksten PB, Voolstra C, Perry AT, Miller DJ, Szmat AM, Coffroth MA, Medina M. Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, _Acropora palmata_ and _Montastrea faveolata._ BMC Genomics 2008; 9:97.

37. Huh GG, Aldrich J, Mottaheddah J, Kwon H, Johnson C, Marsh R. Cloning and characterization of the _Physarum polycephalum_ tectonin: homologues of _Limulus_ lectin L-6. J Biol Chem 1998; 273:6565-6574.

38. Schroeder HC, Ushijima H, Krasco A, Garnulun V, Thakur NK, Diels-Weiher B, Muller IM, Muller WG. Emergence and disappearance of an immune molecule, an antimicrobial lectin, in _Nematoda: a tachylectin-related_ protein in the _spider_ _Subterranea dominulcana._ J Biol Chem 2003; 278:32810-32817.

39. Lu L, Zhang S, Liang Y, Sun X. Identification, expression and antibacterial activity of a tachylectin-related homolog in amphibious _Branchiostoma belcheri_ with implications for involvement of the digestive system in acute phase response. Fish Shellfish Immunol 2009; 26:235-242.

40. Galliano M, Minchioti L, Campagnoli M, Sala A, Visi L, Armesano A, Pucci P, Casbarra A, Causa M, Perduca M, Monaco HL. Structural and biochemical characterization of a new type of lectin isolated from carp eggs. Biochem J 2003; 376:433-40.

41. Tsoi SCM, Ewart KV, Penny S, Liebms C, Darby NJ. Principles of structures of animal and plant lectins. Biochem J 1986; 233:S83-6.

42. Phillips JH. Immune mechanisms from the phylog Ceneterata. In _The Lower Metazoa_ Edited by: Dougherty EC, Brown ZN, Hanson ED, Hartman WD. Berkeley, University of California Press; 1963:425-431.

43. Miller DJ, Hemmrich G, Ball EE, Haywood DC, Khatrullin K, Runyanama N, Agata K, Bosch TCG. The innate immune repertoire in _Cnidaria_ - ancestrality, complexity and stochastic gene loss. _BMC Genomics_ 2007; 8:B85.

44. Drickamer K. Multiplicity of lectin-carbohydrate interactions. _Nature Struct Biol_ 1995; 2:437-439.

45. Loris R. Principles of structures o animal and plant lectins. _BMC Biochem Acta_ 2002; 1572:198-208.

46. Chotia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. _EMBO J_ 1986; 5:823-6.

47. Creighton TE, Darby NJ. Functional evolutionary divergence of proteolytic enzymes and their inhibitors. _Trends Biochem Sci_ 1989; 14:319-24.

48. Wul AM, Singh T, Liu J-H, Kaczeminski M, Russwurm R, Siebert H-C, Bonvin AMJ, Andre S, Cabu HJ. Activity - structure correlations in divergent lectin evolution: fine specificity of chicken galectin CC-14 and computational analysis of flexible ligand docking for CG-14 and the closely related CG-16. _Cytology_ 2007, 17:165-184.

49. Ford MJ. Applications of selective neutrality tests to molecular ecology. _Mol Ecol_ 2002; 1:1245-1262.

50. Schwarz RS, Hodes-Villem D, Fitzpatrick KA, Fain MG, Hughes AL, Cadavid KA, Fain MG, Hughes AL, Cadavid KA. A gene family of putative immune recognition molecules in the _Hydrodynam_ _amphora._ _Immunogenetics_ 2007; 59:1572-1580.

51. De Mita S, Santoni S, Hochu J, Ronfort J, Bataillon T. Molecular evolution and positive selection of the symbiotic gene NORK in _Medicago truncatula._ _J Mol Evol_ 2006; 62:234-244.

52. De Mita S, Ronfort J, McNab H, Ponceet C, El Malek R, Bataillon T. Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in _Medicago truncatula._ Genetics 2007; 177:2123-2133.

53. Martinez-Romero E. Coevolution in _Rhizobium-legume_ symbiosis? _DNA Cell Biol_ 2009; 28:361-370.

54. Voolstra CR, Sunagawa S, Schwarz JA, Coffroth MA, Yellowlees D, Leggat W, Medina M. Evolutionary analysis of orthologous _cDNA_ sequences from cultured and symbiotic _dinoflagellate_ symbionts of reef-building corals (_Dinophyceae: Symbiodinium._ Comp Biochem Physiol D 2009; 4:67-74.

55. Baird A, Maynard MA. Coral adaptation in the face of climate change. Science 2008; 320:315-315.

56. Hoeh-Gulberg O, Mumb PJ, Hooten AT, Steneck RS, Greenfield P, Gomez E, Harvall DR, Sale PF, Edwards AJ, Caldeka K, Knowlton N, Eakin CM, Iglesias-Prieto R, Muthiga N, Bradbury RH, Dubi A, Hatzios ME. Coral adaptation in the face of climate change - Response. Science 2008; 320:315-316.

57. Reed JK, Keong CC, Shepard AN. Impacts of bottom-trawling on a deepwater _Oculina_ coral ecosystem off Florida. _Bull Mar Sci_ 2007; 81:481-496.

58. Vollmer SV, Kline DL. Natural disease resistance in threatened staghorn corals. PLoS One 2008; 3:e3718.

59. Moeller DA, Tiffin P. Genetic diversity and the evolutionary history of plant immunity genes in two species of _Zea._ _Mol Biol Evol_ 2005; 22:2480-2490.

60. Moeller DA, Tiffin P. Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. _Evolution_ 2008; 62:3069-3081.

61. Shinner D, Nickle DC, Jensen MA, Mullins J. Potential impact of recombination on sitewise approaches for detecting positive natural selection. _Genet Res_ 2003; 81:115-121.

62. Hey J, Wakeley J. A coalescent estimator of the population recombination rate. _Genetics_ 1997; 145:833-846.

63. Anisimova M, Nielsen R, Yang ZH. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. _Science_ 2003; 304:1229-1236.

64. Kaczkowski-Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDK. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 2006; 23:1891-1901.

65. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. _Genetics_ 1985, 111:147-164.

66. Schlenke TA, Begun DJ. Natural selection drives _Drosophila_ immune system evolution. _Genetics_ 2003; 164:1471-1480.

67. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubsiz MJ, Fledel-Alon A, Tenenbaum DM, Covello D, White TJ, Snisny JS, Adams MD, Cargill M. A scan for positively selected genes in the genomes of _humans_ and _chimpanzees._ PLoS Biol 2005; 3:e170.

68. Rukke EG, Toomejan C, Kretman M, Bergelson J. Genome-wide survey of _R_ gene polymorphisms in _Arabidopsis._ _Plant Cell_ 2006; 18:1803-1818.

69. Meyer E, Agliamoya GV, Wang S, Buchanan-Carter J, Abrego D, Colboune JK, Willis BL, Matz MV. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFP. _BMC Genomics_ 2009; 10:219.

70. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the _Phyre_ server. Nat Protoc 2009; 4:363-371.

71. Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA. Exploring the extremes of _sequence_ _structure_ space with ensemble fold recognition in the program _Phyre_. _Proteins_ 2008; 70:611-625.

72. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. _Electrophoresis_ 1997; 18:2714-2723.

73. Fraczkiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. _J Comp Chem_ 1998; 19:319-333.

74. Clark NL, Swanson WJ. Pervasive adaptive evolution in primate sensory proteins. _PLoS Genet_ 2005; 1:e35.

75. Yang Z. _PAML_ 4. Phylogenetic analysis by maximum likelihood. _Mol Biol Evol_ 2007; 24:1586-1591.

76. Wong WSW, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. _Genetics_ 2004; 168:1041-1051.

77. Yang Z, Nielsen R, Goldman N, Pedersen AMK. Codon-substitution models for heterogeneous selection pressure at amino acid sites. _Genetics_ 2000; 155:431-449.
82. Yang Z, Wong WSW, Nielsen R: Bayesian inference of amino acid sites under positive selection. Mol Biol Evol 2005, 22:1107-1118.
83. Posada D: jModelTest: Phylogenetic model averaging. Mol Biol Evol 2008, 25:1253-1256.
84. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. In PhD dissertation The University of Texas at Austin, 2006.
85. Aagaard JE, Phillips P: Accuracy and power of the likelihood ratio test for comparing evolutionary rates among genes. J Mol Evol 2004, 60:426-433.
86. Swofford DL: PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). In Version 4 Sinauer Associates, Sunderland, Massachusetts, 2003.
87. Librado P, Rozas J: DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009, 25:1451-1452.
88. McGuire G, Wright F: TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. Bioinformatics 2000, 16:130-134.
89. Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G: TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics 2004, 20:1806-7.
90. Kosakovsky Pond SL, Frost SDW: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 2005, 21:2531-2533.
91. Hudson RR: Estimating the recombination parameter of a finite population-model without selection. Genet Res 1987, 50:245-250.

doi: 10.1186/1471-2148-10-150
Cite this article as: Hayes et al., High amino acid diversity and positive selection at a putative coral immunity gene (tachylectin-2) BMC Evolutionary Biology 2010, 10:150