Computational characterization of the mtORF of pocilloporid corals:
insights into differences in protein structure and function among Stylophora lineages from contrasting environments

*Eulalia Banguera-Hinestroza,1,2, Yvonne Sawall3, Jean-Francois Flot1,2

1. Université libre de Bruxelles, Evolutionary Biology and Ecology, Belgium.
2. Interuniversity Institute of Bioinformatics in Brussels – (IB)2, Brussels, Belgium

*Corresponding authors: ebanguer@ulb.ac.be; eulalia.banguera@gmail.com
Abstract

More than a decade ago, a new mitochondrial Open Reading Frame (mtORF) was discovered in corals of the family Pocilloporidae, which turn out to be an effective barcode gene for these corals. However, its function remains unknown. Recently, this gene revealed the existence of a hybrid \textit{Stylophora} lineage (\textit{RS\_LinA}) inhabiting in sympatry along the environmental gradient of the Red Sea (18.5°C to 33.9°C) with its parental species (\textit{RS\_LinB}). Furthermore, in \textit{RS\_LinB}, the mtORF uncovered phylogeographic patterns that were strongly correlated with environmental variations. This was similar to the patterns unraveled by \textit{hsp70}, suggesting that mtORF too might be involved in thermal adaptation. Here we used computational approaches to characterize the mtORF and to identify its potential role. Results showed that this gene encodes a transmembrane protein (0.97<P< 1.00) involved in transport (0.80<P< 0.87), regulation of metabolic processes (0.70<P<0.85), and likely in the cell-surface receptor signaling pathway (0.56<P<0.80). Predicted protein functions differed among \textit{Stylophora} lineages and interestingly, in \textit{RS\_LinB} only, the protein was intrinsically disordered and displayed domains involved in cellular complexes and stress response (0.0001< P <0.001). These characteristics, exclusive of an endemic lineage adapted to extreme environmental fluctuations, support a role of the mtORF in stress response, speciation and adaptation.

\textbf{Key words}: protein function, mtORF, \textit{Stylophora}, thermal adaptation, disordered amino acids residues, pocilloporid corals, a transmembrane protein, pocilloporid corals.
1. Introduction

New mitochondrial open reading frames (mtORFs) are occasionally reported in the literature; in some cases, they have been proven to encode functional proteins, with a genetic variability sufficient to unravel patterns of differentiation in populations and species from different geographic origin (e.g., unicellular algae of the class Raphidophyceae [1]). In corals, a mtORF apparently unique to three genera of the family Pocilloporidae (Pocillopora, Seriatopora and Stylophora) was reported by Flot and Tillier 2007 [2]. This mtORF is also found in the genus Madracis (this study) –of the same coral family [3]–, but it was annotated as a control region in the only mitogenome of this genus available to date [4], leading to some confusion in the literature.

The mtORF gene has been useful for the delimitation of species within Pocillopora [2,5,6] and has allowed the identification of paraphyletic morphospecies in Seriatopora, as well as fine-scale genetic structure in species belonging to this genus [7,8]. Furthermore, it has revealed strong phylogenetic and phylogeographic patterns in Stylophora [9,10]. This indicates that in the absence of highly variable mtDNA genes in corals [11,12], the mtORF may therefore be a useful mitochondrial barcode gene for pocilloporids [2,6].

mtDNA genes, in particular barcode genes (e.g., cytochrome oxidase subunit I gene cox1 [13]), often unravel patterns of genetic variation that are highly correlated with changes in environmental conditions, mostly related to temperature and altitude ([13–17]. This is because they are involved in key molecular mechanisms related to mitochondrial bioenergetics. For example, they are key players in the mito-nuclear protein complexes of the oxidative phosphorylation [OXPHOS] system, which transform ADP (adenosine 5′-diphosphate) and phosphate to the cellular energy carrier adenosine 5′-triphosphate (ATP) ([18,19]. Hence, mtDNA genes play an important role in the thermal tolerance of an organism by regulating the availability of ATP whose demand is often increased in organisms thriving under environmental extremes [20–23].
Mutations at the mtDNA OXPHOS genes, enhanced by natural selection, has been found to have an important impact in the ecology and evolutionary history of multiple taxa [14–16,21,24–26]. These mutations produce mito-nuclear incompatibilities between genotypes adapted to markedly different environments, creating reproductive barriers and increasing genetic differentiation among populations (particularly in species distributed across broad geographical ranges), which are the main processes leading to speciation [13,21,23,27–29].

The involvement of mtDNA genes in adaptation to the environment and in speciation seems to be supported by the evolutionary patterns revealed by the mtORF barcode gene in the genus *Stylophora*. Recently, Banguera-Hinestroza et al. (2018) [9] analysed 827 samples of *Stylophora* covering a broad geographical range of this genus, including the full latitudinal (12 latitudes) and environmental gradient of the Red Sea (i.e. minimum and maximum Sea Surface Temperatures – SST– have been recorded between 18.5°C-29.9°C in northern areas and 23.5°C-33.9°C in southern regions [30]). They found a high number of mutational changes in the mtORF gene as well as in adjacent genes (*nd6* and *atp6*). Interestingly, the three genes are part of a recombinant region that strongly differentiates a hybrid lineage (*RS_LinA*) from its parental species (*RS_LinB*), both inhabiting the entire environmental gradient of the Red Sea. Furthermore, in *RS_LinB*, the mtORF uncovered the existence of two well-differentiated populations restricted respectively to the colder northern regions or to the warmer central-southern Red Sea. The same phylogeographic pattern was observed for the *hsp70* gene encoding a heat-shock protein well known for its role in stress response and adaptation to different climatic conditions [31]. This suggests that both genes may have played a similar function in the adaptation of the ancestral and endemic *Stylophora* lineage to the different temperature regimes of the Red Sea including extremely warm conditions in the Southern region.

So far, the pocilloporids’ mtORF gene has not been fully characterized and its function in the mitogenome is currently unknown, which limits our understanding of its possible role in adaptation and speciation. Therefore, the aim of the current study, is to characterize the mtORF and unravel its function, as well as to identify signals of natural selection in the mtORF proteins of species inhabiting different environments, including the extreme environments of the Red Sea.
Previous studies suggested that the mtORF gene might be a pseudogene [32,33], in which case multiple stop codons or frameshift mutations should be observed, leading to a non-functional protein [34,35]. Hence, as a first step, we analyzed hundreds of mtORF sequences from *Pocillopora, Seriatopora* and *Stylophora*, freely available in the NCBI data base and searched for stop codons in the translated sequences. In this analysis, we also included the only mtORF sequence available from the mitogenome of the genus *Madracis*. This approach was additionally used for identifying whether this gene encodes a functional protein, because long ORF (more than 600 bp in the case of this gene) are rarely maintained in genome, if they are not expressed [36].

As a second step, we used computational prediction methods available at the TMHMM Server [37,38] and at the PSIPRED Workbench [39] to analyze the structure and function of mtORF in *Madracis, Pocillopora, Seriatopora*, and *Stylophora*, including the two *Stylophora* lineages previously identified from the Red Sea [9] —note that mtORF in italic refers to the protein, while non-italic notation of ORF refers to the gene—. Both workspaces have been shown to be reliable for predicting protein structure and function for non-model species [40]. After characterizing the mtORF protein, we tested whether it may be involved in local adaptation by searching for signals of selection in the sequences of *Stylophora* lineages [9], particularly those from the Red Sea. As mentioned above, these lineages occurred throughout the entire Red Sea in regions with strong environmental differences. The most northern region of the Red Sea (Gulf of Aqaba, 32°N) is characterized by the coldest temperature regime (18.5°C-28.8°C), highest salinity (41 PSU) and generally low nutrient input, while the most southern region (Farasan islands, 20°N) is characterized by an exceptionally warm temperature regime (23.5°C-33.9°C), lower salinity (37 PSU) and high nutrient input [30,41,42].

2. Material and methods

2.1. Characterization of the mtORF gene
A total of 1830 mtORF sequences of *Pocillopora*, *Seriatopora* and *Stylophora* publicly available in the NCBI data base [4,6–8,10,33,43,44] were translated using the coelenterate mitochondrial code implemented in the program MEGA [45] and scanned for stop codons. In addition, to identify whether this mtORF was also present in the genus *Madracis*, its mitogenome was downloaded from the NCBI data base (accession number: EU400212) and aligned with mitogenomes of the other three genera. After probing homology, the mtORF was extracted, translated and searched for stop codons. Samples of *Stylophora* and *Pocillopora* from the Red Sea were taken exclusively from our collection [9,46] to ensure the reliability of the data.

Structural differences among mtORF sequences and dissimilarities in length and composition of tandem repeats were identified using the Program Tandem repeats finder of Benson (1999) [47]. Sequences from *Stylophora* lineages included in this analysis correspond to the *Stylophora* hybrid (*RS_linA*), and sequences resembling those of the two parental species that gave origin to the hybrid lineage, this is: sequences of *Stylophora* from the Red Sea (*RS_LinB*) and sequences from *Stylophora* from the Indo-Pacific (referred here as *Stylophora pistillata*).

Second, the mtORF extracted from the full mitochondrial genomes of *Madracis mirabilis* (EU400212), *Stylophora pistillata* (accession: EU400214), *Seriatopora hystrix* (accession: EF633600), *Pocillopora damicornis* (accession: EF526303), and from *RS_LinA* and *RS_LinB* [9] were analysed using a set of protein prediction approaches available at the PSIPRED Protein Sequence Analysis Workbench (www.bioinf.cs.ucl.ac.uk/psipred [48]) and at the TMHMM Server v. 2.0. [37,38]. These methods included: (i) the PSIPRED approach (Protein Structure Prediction) by Jones (1999) [49], a highly accurate method that uses two feed-forward neural networks to predict secondary structure in outputs generated using PSI-BLAST (Position Specific Iterated BLAST [39,50]); (ii) the MEMSAT3 method (MEMbrane protein Structure And Topology 3) that predicts secondary structure of membrane protein by using multiple alignments produced by PSI-BLAST and by scoring log-likelihoods ratios through different topological models to gather the consensus [40,48,51,52]; and (iii) the MEMSAT-SVM approach using Support Vectors Machines, which are binary classifiers used to categorize residue preferences before combining them into a probabilistic framework [40]. The outputs of these methods were
compared with the TMHMM approach that predicts transmembrane helices based on a hidden Markov model [37,38].

Further analyses using the PSIPRED webserver included: (i) The prediction of the likelihood of a transmembrane helix (TMH) being involved in the formation of pore lining regions. Those run parallel to transmembrane helices and are vital for biological processes, such as the transport of ions and molecules across the membrane [40]. (ii) The identification of disordered regions using the DISOPRED approach [53–55]. (iii) Classifying functional assignments using FFPred 2.0 [56], which uses the GOA (Gen Ontology Annotation data base) to predict Gene Ontology (GO), including macromolecular interactions, biological processes, molecular function and cellular components, searching throughout a large data set of already known proteins and annotations from Eukaryotes [56]. For each prediction, the posterior probabilities are given in terms of reliability that is composed of three characteristics: sensitivity, specificity and precision [57]. (iv) The prediction of structural domains involved in protein-protein interaction using pDomTHREADER [58]. (v) The distinction of TM regions from signal peptides (SP), using the signalIP software 4.1 [59,60] available at www.expasy.org. For a full description of the approaches see the PSIPRED web server [57].

2.2. Signatures of adaptive evolution in the mtORF of Stylophora inhabiting different environments.

The potential role of the mtORF protein in local adaptation was tested by searching for signals of selection in the translated mtORF sequences. For this, mtORF haplotypes belonging to the Red Sea Stylophora lineages, as well as to Stylophora species from other oceanic regions were used. Samples included in these analyses (N= 827) belong to the same set of samples analysed in Banguera-Hinestroza et al. (2018) [9], from which also the definition of Stylophora lineages, clades, subclades and populations, was taking. Briefly, two Stylophora lineages were distinguished within two highly divergent clades: Clade 1 included the lineage RS_LinA (~50% of Stylophora corals collected within the Red Sea and Gulf of Aden) and Stylophora specimens from the Indo-Pacific, Madagascar, Arabian Gulf and Gulf of Aden regions. Clade 2 included
exclusively *Stylophora* specimens of the lineage *RS_LinB* (the remaining individuals from the Red Sea and Gulf of Aden areas). This Clade was further divided into two well defined groups (subclades), each grouping specimens distribute either in the northern areas of the Red Sea or in the central-southern regions.

For analyses of selection, first, all *Stylophora* sequences were analysed using regions that were unambiguously aligned among all haplotypes. In addition, sequences from Clade 1 (*RS_LinA* plus *Stylophora* from other oceanic basins) and Clade 2 (*RS_LinB*) were analysed separately. Last, selection was tested in sequences from the hybrid lineage (*RS_LinA*) and also in sequences from the northern and southern populations of the parental Red Sea lineage (*RS_LinB*).

Signatures of selection at individual sites on the mtORF protein were inferred using several approaches implemented in the Datamonkey webserver (www.datamonkey.org) [61,62]. (i) The Fixed Effects Likelihood method (FEL), which detects selection by first estimating branch lengths and substitution rates parameters in a given phylogeny, when its corresponding coding alignment is provided. After the calculation of these two parameters the method work with fixed values to infer nonsynonymous (dN) and synonymous (dS) substitution rates per sites allowing the identification of selection along the branches of the phylogenetic tree [63]. (ii) The Mixed Effects Model of Evolution (MEME), which identifies episodic positive selection at individual sites [64] by combining both, fixed models and random effect models. (iii) The Branch-Site REL model (BS-REL -aBSREL-), a method that uses maximum likelihood computations to infer the variation of dN and dS over branches and from site to site [65,66]. Statistical significance levels were set at a *p*-value of 0.1 for MEME and FEL and at 0.5 for aBSREL as recommended by the authors.

Furthermore, the M8 model of Yang et al. (2000) [67] and the Mechanistic–Empirical Combination (MEC) model from Doron-Faigenboim and Pupko, 2006 [68] were applied to the mtORF protein via the Selecton webserver [69,70]. Under these approaches, the best model that fits the data is found by performing a likelihood ratio test (LRT) over the whole alignment among several sequences. Here, a null model in which positive selection is not allowed (i.e. M8a; non-
codons with dN/dS > 1) is compared against a general model that assumes positive selection (i.e. dN/dS > 1) such as the M8 and the MEC model. If positive selection is detected, a Bayes empirical Bayes (BEB) approach is used to calculate the posterior probabilities of sites undergoing positive selection [71]. Statistical significance was tested comparing the AIC (Akaike Information Content) scores between the MEC model or M8 model against the M8a model. Note that the likelihood ratio test is used to compare both models and the significance test is passed whenever the AIC score of M8 or MEC is lower than the M8a score, using a threshold of 0.05.

For details about the advantages of the MEC model, refer to Stern et al. (2007) [70].

The last test used to detect selection was the codon-based Z-test of positive selection of Nei and Gojobori, (1986) [72] implemented in MEGA 7 [45]. This test allows discrimination between positive and purifying selection by performing pairwise comparisons among all haplotypes within a specific clade/subclade against the others. This is done by comparing the relative abundance of synonymous and non-synonymous substitutions and their variances, by calculating dS and dN [45]. The null hypothesis of dN = dS and the alternative hypotheses of dN > dS (positive/diversifying selection) and dN < dS (purifying selection) are tested using one-tailed Z-test. Although this method is comparatively simple for testing selection hypotheses, it is considered to perform as well as more complex methods [73].

3. Results

3.1. General features of the mtORF-encoded protein

Stop codons were not found in the translated mtORF sequences of *Pocillopora*, *Seriatopora* and *Stylophora* (N=1830), except in 3 out of 431 sequences from *Seriatopora* (Accessions KR150027, KR150037 and KR150052), which showed an insertion of one nucleotide in position 165, 206 and 397 respectively, when aligned with all other sequences. We interpreted this as errors during base calls in the published sequences. The mtORF extracted from the mitogenome of *Madracis* was not interrupted by stop codon, as well, and showed to be homologous to the mtORF of other pocilloporids (the alignment of the mtORF of this genus with the other genera is
shown in Supplementary Figure S1). This suggests that the mtORF gene is fully functional and not a pseudogene, which was also corroborated by subsequent analyses of this study (see below).

Translating the mtORF sequences of each genus resulted in proteins of different lengths. The longest was found in RS_LinB with 362 amino acids, followed by Stylophora pistillata with 309 amino acids, Pocillopora with 302, Stylophora RS_LinA with 301, Seriatopora with 269, and Madracis with 221 amino acids. Within Stylophora, differences in length, particularly within the Red Sea group, were found to be associated with the presence of duplicated tandem repeats (TR), which are recorded in Table 1.

3.2. Prediction of transmembrane helices (TMH) and protein structures

Two TMH were predicted in the mtORF encoded protein (abbreviated as mtORF-pro thereafter) of Stylophora, Seriatopora and Pocillopora and three TMH in the mtORF-pro of Madracis by the TMHMM approach, with high posterior probabilities (P > 0.97) and with low probabilities of the N-term being in the cytoplasmic site of the membrane (0.09 < P < 0.48) (Figure 1). The presence of these helices was confirmed by the MEMSAT3 and MENSAT-SVM approaches, except for Seriatopora, in which three TMHs were predicted by MEMSAT3 and MENSAT-SVM. Furthermore, transmembrane topology predicted by MENSAT-SVM suggested a lack of pore-lining residues in Madracis, but the existence of two pore-lining helixes in Seriatopora and one in Pocillopora and Stylophora. No evidence for signal peptides, which are often mistaken with transmembrane helixes, was found using MEMSAT or signalP.

Differences in the mtORF-pro among genera were mainly found in the predictions of secondary structure (Figure 2) and intrinsically disordered regions –IDRs– (Figure 3). Only one disordered residue was predicted in Madracis, Pocillopora and Seriatopora, but multiple unstructured amino acid residues were found in the mtORF-pro of Stylophora lineages: in the ancestral RS_LinB, the IDR expands 234 amino acids (from residue 63 to 296), in Stylophora pistillata 79 amino acids (from residue 123 to 201), and in the mtORF-pro of the hybrid (RS_LinA) it covers two small regions with a total of 27 amino acids (from residue 134 to 153 and from 160 to 166). Noticeably,
Supplementary Figure S1. Conserved blocks in the mtORF-pro of pocilloporid corals
Table 1. Polymorphic Tandem Repeats (TR) in the mtORF of *Stylophora*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Consensus sequence</th>
<th>Period Size</th>
<th>Copy Number</th>
<th>Consensus Size</th>
<th>Percent Indels</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS_LinA</td>
<td>AAGTGACGTCAGAAGATGAGGATTTTTGAGGGAAAGTTTTTT</td>
<td>51</td>
<td>2.1</td>
<td>51</td>
<td>0</td>
<td>38</td>
<td>2</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>RS_LinA</td>
<td>AAGTGACGTCAGAAGATGAGGATTTTTGAGGGAAAGTTTTTT</td>
<td>51</td>
<td>4.1</td>
<td>51</td>
<td>0</td>
<td>39</td>
<td>2</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>Madagascar</td>
<td>AAGTGACGTCAGAAGATGAGGATTTTTGAGGGAAAGTTTTTT</td>
<td>51</td>
<td>2.1</td>
<td>51</td>
<td>0</td>
<td>38</td>
<td>2.3</td>
<td>33</td>
<td>24-25</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>AAGTGACGTCAGAAGATGAGGATTTTTGAGGGAAAGTTTTTT</td>
<td>2.1</td>
<td>51</td>
<td>0</td>
<td>37</td>
<td>1</td>
<td>35</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>RS_LinA</td>
<td>ATTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>39</td>
<td>2.6</td>
<td>42</td>
<td>4</td>
<td>22</td>
<td>7</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Madagascar</td>
<td>TTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>39</td>
<td>2.4</td>
<td>39</td>
<td>7</td>
<td>21</td>
<td>6</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>Madagascar</td>
<td>ATTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>39</td>
<td>2.7</td>
<td>39</td>
<td>4</td>
<td>23</td>
<td>7</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>ATTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>39</td>
<td>3.2</td>
<td>42</td>
<td>6</td>
<td>22</td>
<td>6</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>ATTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>39</td>
<td>2.6</td>
<td>39</td>
<td>0</td>
<td>24</td>
<td>4</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>RS_LinA</td>
<td>TTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>21</td>
<td>3.7</td>
<td>20</td>
<td>12</td>
<td>20</td>
<td>6</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>RS_LinA</td>
<td>TTCAGAAGTTTTGTATGAAAATGTTTTGTTTGTC</td>
<td>18-21</td>
<td>2.8-5.8</td>
<td>21</td>
<td>10-16</td>
<td>21.23</td>
<td>1-3</td>
<td>30-32</td>
<td>42-43</td>
</tr>
<tr>
<td>RS_LinA</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>18</td>
<td>2.4</td>
<td>21</td>
<td>13</td>
<td>29</td>
<td>4</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Indo-Pacific</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>18</td>
<td>1.9</td>
<td>18</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>31</td>
<td>42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region within the Red Sea</th>
<th>Consensus Sequence</th>
<th>Period Size</th>
<th>Copy Number</th>
<th>Consensus Size</th>
<th>Percent Indels</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>21</td>
<td>6-11</td>
<td>21</td>
<td>10</td>
<td>32.34</td>
<td>9</td>
<td>29-31</td>
<td>26</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>27</td>
<td>8-20</td>
<td>27</td>
<td>5-12</td>
<td>34.35</td>
<td>9</td>
<td>24-27</td>
<td>28-30</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>27</td>
<td>4-7</td>
<td>27</td>
<td>7-15</td>
<td>34.35</td>
<td>10-11</td>
<td>27-29</td>
<td>25-26</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>3.1</td>
<td>47</td>
<td>19</td>
<td>34</td>
<td>7</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>6-7</td>
<td>48</td>
<td>16-17</td>
<td>33.34</td>
<td>9-10</td>
<td>27-28</td>
<td>27-28</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>5.9</td>
<td>54</td>
<td>13-17</td>
<td>34.35</td>
<td>9-10</td>
<td>25-27</td>
<td>28-29</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>51</td>
<td>1.9</td>
<td>52</td>
<td>6</td>
<td>36</td>
<td>10</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>69</td>
<td>2.6</td>
<td>69</td>
<td>0</td>
<td>33.34</td>
<td>8</td>
<td>29-30</td>
<td>27</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>69</td>
<td>3.5</td>
<td>69</td>
<td>9</td>
<td>34</td>
<td>9</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>69</td>
<td>4.6</td>
<td>69</td>
<td>3-11</td>
<td>33.34</td>
<td>9-10</td>
<td>26-27</td>
<td>28</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>75</td>
<td>3.3</td>
<td>75</td>
<td>5</td>
<td>33</td>
<td>9</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>75</td>
<td>5.3</td>
<td>69</td>
<td>6</td>
<td>33</td>
<td>9</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Northern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>75</td>
<td>5.9</td>
<td>75</td>
<td>8</td>
<td>34</td>
<td>10</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>21</td>
<td>8.9</td>
<td>24</td>
<td>16</td>
<td>33</td>
<td>10</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>27</td>
<td>8.1</td>
<td>27</td>
<td>7-10</td>
<td>34.35</td>
<td>10</td>
<td>24-27</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>2.5</td>
<td>48</td>
<td>0</td>
<td>34</td>
<td>10</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>4.8</td>
<td>48</td>
<td>12</td>
<td>33</td>
<td>9</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>5.8</td>
<td>48</td>
<td>17</td>
<td>34</td>
<td>9</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>5.6</td>
<td>46</td>
<td>7</td>
<td>34</td>
<td>9</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>5.8</td>
<td>48</td>
<td>5</td>
<td>34</td>
<td>9</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>6.8</td>
<td>47</td>
<td>16</td>
<td>34</td>
<td>10</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>7.4-7.9</td>
<td>54</td>
<td>13-14</td>
<td>34</td>
<td>10</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>48</td>
<td>7.9</td>
<td>54</td>
<td>12</td>
<td>33</td>
<td>10</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>69</td>
<td>3.9</td>
<td>69</td>
<td>2</td>
<td>33</td>
<td>10</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>69</td>
<td>5</td>
<td>74</td>
<td>7</td>
<td>34</td>
<td>10</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>75</td>
<td>4.7</td>
<td>75</td>
<td>7</td>
<td>34</td>
<td>10</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>75</td>
<td>2.7</td>
<td>91</td>
<td>10</td>
<td>34</td>
<td>9</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Southern and Southern</td>
<td>AATGGGGATAAGTTTGGAGGAT</td>
<td>96</td>
<td>3.7</td>
<td>96</td>
<td>8</td>
<td>10</td>
<td>27</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Top: predicted TM segments by the TMHHM approach. This plot shows the overall probability that a residue sits in a helix, inside of it, or outside (Sonnhammer et al., 1998; Krogh et al., 2001). Bottom: MEMSAT3 and MEMSAT-SVM predictions for the query sequence. The curve represent a raw output for the support vector machines. Dashed line shows the prediction threshold (PSIPRED server; Buchan et al., 2013; Nugent, T. & Jones, D.T. (2009).
Figure 2. Secondary structure predictions by PSIPRED (Jones 1999)
Figure 3. Left: Intrinsic disorder profile predicted by DISOPRED3. Disordered amino acids are indicated by a blue line above the grey dashed line (confidence score higher than 0.5). The confidence of disordered protein binding residues are denoted by an orange line (Buchan et al., 2013). Right: TM helix Map predicted by MEMSAT.
in RS_LinB the IDR coincides with a core region revealing a high number of duplicated polymorphic Tandem Repeats (TR) that were different to those found in RS_LinA, in length and nucleotide composition (Table 1).

Three protein domains were identified in RS_LinB with 4218 PSIBLAST hits that were not predicted in other members of this coral family, including other Stylophora lineages. Searches against the 95 million protein domains available in the CATH data base, using the pDomTHREADER approach, showed that the mtORF-pro of RS_LinB had the best matches (levels of confidence: 0.0001 < P < 0.001) with annotated domains involved in: (i) the structural integrity of a complex or its assembly within or outside a cell as per molecular function determined by Gene Ontology –GO = 0005198– (CATH domain = 1s58A00; p-value = 8 x 10^{-5}; from amino acid 1 to 299), (ii) host-bacteria/virus interactions (CATH domain = 1ocyA02, p-value = 5 x 10^{-5}; from amino acid 147-286), and (iii) domains that play a role in cell to cell, cell to matrix interactions and response to stress (CATH domain = 1ux6A01; p-value = 1 x 10^{-4}; from amino acid 136 to 292).

3.3. Predictions of molecular function, biological process, and cellular localization

Three molecular functions: catalytic activity (0.53 < P < 0.77), hydrolase activity acting on acid anhydrides (0.56 < P < 0.92), and cytoskeletal protein binding (0.59 < P < 0.68) were predicted for the mtORF-pro in all pocilloporids. However, broad differences were found in their posterior probabilities (Table 2). For example, in the mtORF-pro of Stylophora RS_LinB, catalytic function was predicted with low probabilities (P = 0.53) and the same was true for hydrolase activity acting on acid anhydrides in the mtORF-pro of Seriatopora (P = 0.56).

Noticeably, in most genera the highest probabilities (P > 0.7) were given to different molecular function. In Seriatopora, ATP binding presented the highest probability (P = 0.832). In Madracis, the highest scores were found for ion channel activity (P = 0.913), hydrolase activity acting on acid anhydrides (P = 0.902) and ion transmembrane transporter activity (P = 0.88).
### Table 2. Prediction of molecular function and biological processes by Psipred

<table>
<thead>
<tr>
<th>Molecular Function</th>
<th>GO term</th>
<th>Posterior Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005216 ion channel activity</td>
<td>Madracis 0.913</td>
<td>Pocillopora 0.862  Seriatopora 0.628</td>
</tr>
<tr>
<td>GO:0016817 hydrolase activity, acting on acid anhydrides</td>
<td>Stylophora RS LinB 0.902</td>
<td>Stylophora RS LinA 0.824</td>
</tr>
<tr>
<td>GO:0015075 ion transmembrane transporter activity</td>
<td>Stylophora RS LinA 0.864</td>
<td></td>
</tr>
<tr>
<td>GO:0022890 inorganic cation transmembrane transporter activity</td>
<td>Stylophora pistillata 0.858</td>
<td></td>
</tr>
<tr>
<td>GO:0008324 cation transmembrane transporter activity</td>
<td>Stylophora 0.833  RS LinB 0.832</td>
<td></td>
</tr>
<tr>
<td>GO:0005524 ATP binding</td>
<td>Stylophora 0.79</td>
<td></td>
</tr>
<tr>
<td>GO:0008092 cytoskeletal protein binding</td>
<td>Stylophora 0.748  RS LinB 0.762</td>
<td></td>
</tr>
<tr>
<td>GO:000166 nucleotide binding</td>
<td>Stylophora 0.707  RS LinB 0.602  Seriatopora 0.602</td>
<td></td>
</tr>
<tr>
<td>GO:0001882 nucleoside binding</td>
<td>Stylophora 0.706  RS LinB 0.56  Seriatopora 0.696</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 cation channel activity</td>
<td>Stylophora 0.688</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 transporter activity</td>
<td>Stylophora 0.67  RS LinB 0.548  Seriatopora 0.684</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 purine ribonucleoside triphosphate binding</td>
<td>Stylophora 0.639  RS LinB 0.538  Seriatopora 0.706</td>
<td></td>
</tr>
<tr>
<td>GO:0008092 cytoskeletal protein binding</td>
<td>Stylophora 0.588  RS LinB 0.592  Seriatopora 0.598</td>
<td></td>
</tr>
<tr>
<td>GO:000166 nucleotide binding</td>
<td>Stylophora 0.582  RS LinB 0.682  Seriatopora 0.582</td>
<td></td>
</tr>
<tr>
<td>GO:0001882 nucleoside binding</td>
<td>Stylophora 0.587  RS LinB 0.648  Seriatopora 0.648</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 cation channel activity</td>
<td>Stylophora 0.587  RS LinB 0.648  Seriatopora 0.648</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 transporter activity</td>
<td>Stylophora 0.587  RS LinB 0.648  Seriatopora 0.648</td>
<td></td>
</tr>
<tr>
<td>GO:0005215 purine ribonucleoside triphosphate binding</td>
<td>Stylophora 0.587  RS LinB 0.648  Seriatopora 0.648</td>
<td></td>
</tr>
</tbody>
</table>

### Biological Process

<table>
<thead>
<tr>
<th>GO term</th>
<th>Posterior Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0006810 transport</td>
<td>Madracis 0.862  Pocillopora 0.847  Seriatopora 0.874  Stylophora RS LinB 0.803  Stylophora RS LinA 0.824  Stylophora pistillata 0.82</td>
</tr>
<tr>
<td>GO:0034220 ion transmembrane transport</td>
<td>Madracis 0.845  Pocillopora 0.689</td>
</tr>
<tr>
<td>GO:0019222 regulation of metabolic process</td>
<td>Madracis 0.813  Pocillopora 0.853  Seriatopora 0.795  Stylophora RS LinB 0.711  Stylophora RS LinA 0.849  Stylophora pistillata 0.791</td>
</tr>
<tr>
<td>GO:0007166 cell surface receptor signaling pathway</td>
<td>Madracis 0.804  Pocillopora 0.63  Seriatopora 0.564  Stylophora RS LinB 0.79  Stylophora RS LinA 0.662  Stylophora pistillata 0.717</td>
</tr>
<tr>
<td>GO:0009117 nucleotide metabolic process</td>
<td>Madracis 0.728  Pocillopora 0.697</td>
</tr>
<tr>
<td>GO:000166 establishment of localization in cell</td>
<td>Madracis 0.694  Pocillopora 0.692  Seriatopora 0.721  Stylophora RS LinB 0.538  Stylophora RS LinA 0.698  Stylophora pistillata 0.659</td>
</tr>
<tr>
<td>GO:000166 cellular localization</td>
<td>Madracis 0.647  Pocillopora 0.631  Seriatopora 0.64  Stylophora RS LinB 0.632  Stylophora RS LinA 0.643  Stylophora pistillata 0.636</td>
</tr>
</tbody>
</table>
In *Pocillopora* and *RS_LinB*, the highest score was given to hydrolase activity acting on acid anhydrides (P = 0.862 and P = 0.814 respectively). In *RS_LinA*, kinase activity (P = 0.857), purine nucleotide binding (P = 0.779), and ribonucleoside binding (P = 0.779) had the highest scores. In *Stylophora pistillata* catalytic activity and cytoskeletal protein binding were predicted as the most probable functions, with moderated probabilities (both P = 0.68). Ion transmembrane transport activities were predicted for *Madracis* and *Seriatopora*, but not for other genera. Moreover, kinase activity was suggested for the mtORF-pro of all genera, except *Madracis*, while ribonucleoside binding was predicted only for *RS_LinA* and *Seriatopora* (Table 2).

The mtORF-pro for these genera differed also in their aliphatic indices, a measure of the thermostability of the protein, where a higher value indicates a higher thermostability [74]: 112.4 for *Madracis*, 94.4 for *Pocillopora*, 88.3 for *RS_LinA*, 87.2 for *RS_LinB*, 85.4 for *Stylophora pistillata* and 84.0 for *Seriatopora*.

Two main biological processes were suggested for the role of mtORF-pro in all genera and *Stylophora* lineages (Table 2): transport (0.80 < P < 0.87) and regulation of metabolic processes (0.71 < P < 0.85), except in *RS_LinB* where cell surface receptor signaling pathway had the highest posterior probability after transport (P = 0.79). Other processes were predicted with moderated probabilities and two were only suggested for the mtORF-pro of *Madracis* and *Seriatopora* (i.e. ion transmembrane transport and nucleotide metabolic process; Table 2). Furthermore, predictions for cellular localization yielded a high probability and reliability for this protein being an intrinsic and integral component of the membrane (0.97 < P < 1.00) in all genera, with moderated to high probabilities of being part of an organelle membrane (i.e. mitochondria; 0.70 < P < 0.80).

**3.4. Signatures of selection in the mtORF-pro of *Stylophora* corals**

Amino acid sites under positive/diversifying selection were identified with the MEME approach for Clade 1 sequences (i.e. sequences from *RS_LinA* plus specimens from Madagascar, Indo-Pacific and Indian Ocean; at 2 sites) and when sequences from *RS_LinA* and *RS_LinB* were
evaluated together (at 5 sites). However, positive selection was not detected when the alignment included all *Stylophora* sequences (Clade 1 and Clade 2; evaluating only unambiguously aligned regions) or when sequences from each *Stylophora* lineage were analysed separately. In contrast, the M8 and MEC model recovered the highest number of single amino acid sites under positive selection in most cases. However, only the MEC model showed significant values for most comparisons, this is: (i) in the mtORF-pro of *Stylophora* specimens within Clade 1 –at 1 site– (AIC score for MEC: 2846.897 was lower than the M8a score: 2852.35), (ii) when the alignment included sequences from both Red Sea lineages (*RS*_LinA and *RS*_LinB; AIC score for MEC: 1468.55; AIC score for M8a: 1472.37), and (iii) when sequences from all *Stylophora* specimens were analysed (i.e. the MEC score = 1737.074; M8a score = 1745.86).

In addition, the analyses of sequences from the southern and northern group of *RS*_LinB, separately, indicated multiple sites under positive selection in the mtORF-pro of the southern group (AIC score for MEC: 2313.488, AIC score for M8a: 2317.705), which was corroborated by the analyses of the Nei-Gojobori method with high-probabilities (0.003 < *P* < 0.042). Positive selected sites were not found in the northern group of this lineage nor in sequences from *RS*_LinA alone (except in the mtORF-pro of specimens which haplotypes have the highest frequencies in the southern Red Sea). A graphical view of positively and negatively selected sites using the MEC approach, as well as a graphical summary for the outputs of the Nei-Gojobori method are included in Figure 4 (Full matrices are available from the authors).

Amino acid sites under pervasive negative/purifying selection were found in all alignments of the mtORF-pro using the FEL method: at 11 sites for Clade 1 plus Clade 2 sequences; at 7 sites for Clade 1 sequences, at 6 sites for Clade 2 sequences (*RS*_LinB) and also at 6 sites when *RS*_LinA and *RS*_LinB sequences were placed together. One site was found under purifying selection in the mtORF-pro of the northern group of *RS*_LinB, as well as in the southern group. Furthermore, purifying selection was also detected in the mtORF-pro of Indo-Pacific specimens by the Nei-Gojobori method (0.003 < *P* < 0.042). aBSREL did not detect sites under selection along the branches of the phylogenetic tree of *Stylophora*. 
Stylophora CLADE 1 (RS_LinA plus Stylophora pistillata)

Stylophora CLADE 2 (RS_LinB)

Figure 4. Left: a graphical view of the Codon-based test of purifying and positive/diversifying selection (Nei and Gojobori 1986) between the sequences belonging to haplotypes in each Stylophora Clade. Right: Sites under positive and purifying selection predicted by the MEC approach (Doron-Faigenboim and Pupko, 2006); for clade 1 (above) and for the southern group of Clade 2 (below).
4. Discussion

4.1 mtORF protein structure in the coral family Pocilloporidae

This study confirms that the mtORF gene encodes a transmembrane protein in pocilloporid corals, which is in agreement with its preliminary characterization by Flot and Tillier (2007) [2]. No stop codons were found in the translated sequences and all protein prediction approaches applied (TMHMM, MEMSAT and MEMSAT-SVM) predicted two or three different TMHs in all four investigated pocilloporid genera, with high posterior probabilities (P > 0.98 in the TMHMM approach). This finding was reinforced by Gene Ontology (GO terms) analysis predicting that this protein is an integral (P > 0.97) and intrinsic component of the membrane (P > 0.99), likely of the mitochondria (P > 0.7).

The secondary structure of the mtORF-pro differed among genera and particularly among Stylophora lineages (Figure 2) and the prediction of IDRs varied largely among pocilloporids. For example, IDRs were absent in the proteins of Madracis, Pocillopora and Seriatopora (only one amino acid residue was predicted as disordered), but a long stretch of IDRs were predicted in the protein of Stylophora RS_LinB (234 disordered amino acid residues out of 362). This contrasted with the number of IDRs predicted for Stylophora pistillata (79 residues out of 309) and RS_LinA (27 residues out of 301). Furthermore, in RS_LinB, IDRs coincided with a region rich in duplicated TR (Table 1) that generated long insertion-deletions (indels) among sequences.

Long IDRs, as the predicted here for RS_LinB, are recognized to facilitate macromolecular interactions, participating actively in the assembly of signaling complexes by interacting with structured domains in other proteins [54,75–77]. They are also essential for increasing the involvement of a range of proteins in the diversity of cellular processes required during adaptation, particularly to variable environmental conditions [78]. In line with this, one of the most outstanding findings in this study was the presence of protein domains (0.0001< P < 0.001) that were exclusively found in the mtORF-pro of RS_LinB (likely the oldest Stylophora within the Red Sea [9]) and were predicted to be involved in the assembly of cellular complexes (CATH...
domain code: 1s5sA00), cell-to-cell and cell-to-matrix interactions, and response to stress
(CATH domain code: 1ux6A0). This is highly congruent with the occurrence of IDRs in the
protein of this lineage, suggesting a key role of the mtORF-pro in mitochondrial complexes,
which are key in the regulation of multiple processes related with thermal adaptation [22].
Furthermore, in most metazoans IDRs are associated with regulation of cellular processes,
including intracellular signaling, membrane fusion and transport, and signal transduction [54,77],
which were the main processes predicted for biological function of the mtORF-pro of RS_LinB
(i.e. transport and cell surface receptor signaling pathway [P = 0.8]).
Our findings, particularly those associated to the mtORF-pro of Stylophora RS_LinB, therefore,
support the hypothesis that differences in protein structure among genera and Stylophora lineages
are likely the results of dissimilarities in the environmental pressures that each group faced
during its evolution. Although experimental studies would be necessary to understand the role of
IDRs and TR in the genome of RS_LinB, previous studies in other eukaryotes have suggested that
IDRs and indels not only play a role in genome evolution but are also key in the regulation of
gene expression during the exposition of organisms to different stressors [79,80]. Moreover, TRs
are a common characteristic of the genome of organisms evolving under strong selective
pressures, environmental stress, drastic environmental changes, global warming and/or that are
exposed to new environments [81–83]. In humans, for example, mutations that affect TR number
have been found to increase the fitness of the cell when exposed to stressful conditions (e.g.,
cold, heat, hypoxia, oxidative stress) by adjusting the regulatory network which enhances protein
activity and gene expression [84]. All these previous findings support our assumption of a key
role of the mtORF-pro and of the mitogenome in adaptation to changing environments, as
suggested by Banguera-Hinestroza et al. (2018) [9].

4.2. Differences in mtORF-pro function among pocilloporids

The molecular function of the mtORF-pro was predicted to differ among genera and among
Stylophora lineages. Most functions related with ion and cation transmembrane transporter
activities (0.75 < P < 0.91; Table 2), were only predicted for the mtORF-pro of Madracis. Moreover, for this genus, as well as for Pocillopora and RS_LinB, hydrolase activity acting on acid anhydrides, catalyzing the transmembrane movement of substances, was predicted with the highest probability (for all P > 0.82; only second after ion channel activity in Madracis). This function, however, showed low probabilities for Seriatopora, RS_LinA and Stylophora pistillata (0.56 < P < 0.69). In fact, the analyses of the mtORF-pro of these genera yielded the highest probabilities for different molecular functions: ATP binding (P = 0.83) for Seriatopora; kinase activity (P = 0.86) for Stylophora RS_LinA; and catalytic and cytoskeletal protein binding activities (both P = 0.68) for Stylophora pistillata.

The dissimilarities in the predicted protein function of mtORF-pro and the differences in posterior probabilities when several functions were predicted, are in line with differences found in protein structure among genera. This may be the result of mutational and structural changes occurring in this protein during the divergence of pocilloporids, also likely linked to different environmental settings (i.e. in Stylophora lineages). Although this hypothesis is highly speculative and experimental analyses are required to confirm these differences, it is well known that proteins may perform different functions depending on their biological context, such as cellular location, the substrate to which they bind and their interactions with other proteins to form molecular complexes [85]. In this respect, it is broadly accepted that proteins involved in mitochondrial complexes, which is likely the case of this mtORF-pro, are prone to structural changes, which not only impact protein function, but also enhance the protein flexibility that is required for adaptation to different conditions (see review by Fields, 2001 [86]).

4.3. Signatures of selection in the mtORF-pro of Stylophora lineages

Our data show that the mtORF-pro of Stylophora lineages carries signatures of both positive/diversifying and negative/purifying selection a multiple amino acid sites. Interestingly, the MEC approach detected signal of positive selection only in the protein of the southern group of RS_LinB (specimens inhabiting the warmest regions of the Red Sea), a signature that was confirmed by the Nei-Gojobori method (Figure 3). In contrast, no positive selected sites were
detected in the mtORF-pro of the northern group (i.e. restricted to the coldest areas). In this
group, as well as in sequences from RS_LinA specimens, multiple sites were found under
purifying selection (except in protein sequences of those haplotypes with the highest frequency in
the southern Red Sea, Figure 3). The findings of positive selection in the mtORF-pro of
specimens inhabiting the hottest areas of the Red Sea are congruent with the results based in
protein structure and function, supporting the hypothesis of a role of this protein in thermal
adaptation to high temperatures, highlighting also its functional importance.

As suggested by previous phylogenetic studies [9], RS_LinB represent an ancestral lineage,
endemic of the Red Sea, with an ancient history within this region. Although the first occurrence
of this lineage in the Red Sea is unknown, it is broadly accepted that coral reef ecosystems
accompanied the evolution of this basin since its incipient stage in the early Miocene and have
been affected by the multiple and strong climatic variations occurring in the region throughout
time [87]. Pocilloporid corals, as well as those from other coral families, inhabiting the Red Sea
are likely descendants of those that entered the region during the Pliocene-Pleistocene epochs
[87–91], which overcame the strong climatic fluctuations in salinity and temperature of the
glacial e interglacial periods [92–94] by surviving in refugia in the northern and southern Red Sea
(Reviewed in DiBattista et al., 2016 [95]).

The above explain the signals of adaptive evolution found in the mtORF-pro of RS_LinB, for
which the extreme fluctuation in temperatures in the southern Red Sea may have imposed strong
selective pressures. Moreover, the lack of positive selection in the protein of RS_LinA that
contrasted with the multiple amino acid sites found under purifying selection, may be related
with its recent history in the Red Sea, in agreement with its hybrid origin. In fact, more recent
Stylophora lineages likely entered the Red Sea within the last 4-7k years, when environmental
conditions were similar as today [96], passing the bottleneck of high temperature waters in the
southern Red Sea (up to 34°C) and then spreading north along with a range of other coral species
[97]. Purifying selection, therefore, likely acted as the main force against deleterious mutations in
the mtORF-pro of this lineage, shaping the mitochondrial diversity needed for its survival in such
range of environmental conditions.
Conclusions

Our study not only offers insights into the coding nature of the mtORF gene in pocilloporid corals, but add to our knowledge of the main characteristics of this protein. Striking differences were found among the mtORF proteins of pocilloporid corals, with the most noticeable being those among *Stylophora* lineages, which include differences in predicted structure and functions. The most differentiated protein was that of *RS_LinB*, a lineage that was hypothesized to have conquered the extreme environments of the Red Sea during an early colonization, which date is still unknown, but likely predating other *Stylophora* lineages. All the characteristics found in the mtORF-pro of the *RS_LinB* (i.e. high frequency of TR, long stretches of disordered residues and the presence CATH domains related to stress response and protein complexes), plus the signals of selection, support our hypothesis of the role of this protein, and therefore of the mitogenome, in the adaptation of *Stylophora* corals to extreme environments and fluctuating conditions.

This study opens the door for further studies looking at the role of mitochondria and mitochondrial genes in coral species adaptation and diversification. Corals and coral reefs are severely threatened by climate change (e.g. sea surface temperature rise), hence, more detailed studies on coral adaptive mechanisms are required to understand the potential responses of corals in both ecological and evolutionary scales. We consider the understanding of protein-protein interaction at the mitochondrial level of particular relevance, since they appear to play a key role in coral adaption to strong environmental changes.

**Funding statement:**

EBH’s postdoctoral position in Belgium was funded by an *Action de Recherche Concertée* (ARC) grant of the Fédération Wallonie-Bruxelles to JFF. The study was further supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia and the bi-lateral project “The Jeddah Transect” of the King Abdulaziz University (KAU), Saudi Arabia, and the Helmholtz Center for Ocean Research (GEOMAR), Germany (YS).
Acknowledgments: We thank Christian Voolstra and the Bioscience Core Lab at KAUST for sharing their facilities, and Abdulmoshin Al-Sofyani at KAU for supporting coral sampling. Thanks also to Sandra Cervantes Arango, Dario Ojeda Alayon and Patrick Mardulyn, for useful discussions and advices.

Author Contributions: Conceptualization, Eulalia Banguera-Hinestroza, Yvonne Sawall and Jean-François Flot; Formal analysis, Eulalia Banguera-Hinestroza; Funding acquisition, Jean-François Flot; Investigation, Eulalia Banguera-Hinestroza, Yvonne Sawall and Jean-François Flot; Methodology, Eulalia Banguera-Hinestroza; Writing – original draft, Eulalia Banguera-Hinestroza; Writing – review & editing, Eulalia Banguera-Hinestroza, Yvonne Sawall and Jean-François Flot.

REFERENCES


65. Kosakovsky Pond, S.L.; Murrell, B.; Fourment, M.; Frost, S.D.W.; Delport, W.; Scheffler,
K. A Random Effects Branch-Site model for detecting episodic diversifying selection. 


