

Article

# Computational Characterization of the mtORF of Pocilloporid Corals: Insights into Protein Structure and Function in *Stylophora* Lineages from Contrasting Environments

Eulalia Banguera-Hinestroza <sup>1,2,\*</sup>, Evandro Ferrada <sup>3,†</sup>, Yvonne Sawall <sup>4</sup> and Jean-François Flot <sup>1,2,\*</sup>

<sup>1</sup> Evolutionary Biology and Ecology, Université libre de Bruxelles, B-1050 Brussels, Belgium

<sup>2</sup> Interuniversity Institute of Bioinformatics in Brussels—(IB)2, 1050 Brussels, Belgium

<sup>3</sup> Center for Genomics and Bioinformatics, Universidad Mayor, Santiago, Chile; evandro.ferrada@mayor.cl

<sup>4</sup> Coral Reef Ecology, Bermuda Institute of Ocean Sciences (BIOS), St. George's GE 01, Bermuda; yvonne.sawall@bios.edu

\* Correspondence: eulalia.banguera@gmail.com or ebanguer@ulb.ac.be (E.B.-H.); jflot@ulb.ac.be (J.-F.F.)

† These authors share first authorship.

Received: 21 January 2019; Accepted: 23 April 2019; Published: 27 April 2019



**Abstract:** More than a decade ago, a new mitochondrial Open Reading Frame (mtORF) was discovered in corals of the family Pocilloporidae and has been used since then as an effective barcode for these corals. Recently, mtORF sequencing revealed the existence of two differentiated *Stylophora* lineages occurring in sympatry along the environmental gradient of the Red Sea (18.5 °C to 33.9 °C). In the endemic Red Sea lineage *RS\_LinB*, the mtORF and the heat shock protein gene *hsp70* uncovered similar phylogeographic patterns strongly correlated with environmental variations. This suggests that the mtORF too might be involved in thermal adaptation. Here, we used computational analyses to explore the features and putative function of this mtORF. In particular, we tested the likelihood that this gene encodes a functional protein and whether it may play a role in adaptation. Analyses of full mitogenomes showed that the mtORF originated in the common ancestor of *Madracis* and other pocilloporids, and that it encodes a transmembrane protein differing in length and domain architecture among genera. Homology-based annotation and the relative conservation of metal-binding sites revealed traces of an ancient hydrolase catalytic activity. Furthermore, signals of pervasive purifying selection, lack of stop codons in 1830 sequences analyzed, and a codon-usage bias similar to that of other mitochondrial genes indicate that the protein is functional, i.e., not a pseudogene. Other features, such as intrinsically disordered regions, tandem repeats, and signals of positive selection particularly in *Stylophora RS\_LinB* populations, are consistent with a role of the mtORF in adaptive responses to environmental changes.

**Keywords:** mtORF; *Stylophora*; adaptation; disordered residues; tandem repeats; pocilloporid corals; selection; transmembrane proteins

## 1. Introduction

The mitochondrial genome of metazoans (multicellular animals) usually contain 37 genes with no introns, among which only 13 are protein coding [1]. However, deviations from this rule exists and new functional mitochondrial open reading frames (mtORFs) have been reported in several marine taxa, such as brachiopods [2], molluscs [3–5], cnidarians [6], and sponges (reviewed in [7]). In some cases, these new mtORFs have been shown to encode proteins with a genetic variability sufficient to

unravel patterns of differentiation in marine populations and species from different geographic origins (e.g., in unicellular algae of the class Raphidophyceae [8]).

In corals, a novel mtORF was reported by Flot and Tillier [9] in three genera of the family Pocilloporidae (*Pocillopora*, *Seriatopora* and *Stylophora*). This mtORF gene has been useful for the delimitation of species within *Pocillopora* [10,11] and has allowed the identification of cryptic species and of fine-scale genetic structure in *Seriatopora* [12–14]. Furthermore, it has revealed strong phylogenetic and phylogeographic patterns in *Stylophora* [15,16]. This indicates that in contrast to the absence of highly variable mtDNA genes in most corals [17,18], the mtORF may be a suitable mitochondrial barcode gene for pocilloporids [9,11], revealing genetic variation useful for distinguishing lineages evolving under different environmental conditions.

mtDNA genes, in particular barcode genes (e.g., the cytochrome *c* oxidase subunit I gene *cox1* [19]), often unravel patterns of genetic variation that are highly correlated with changes in environmental conditions, mostly related to temperature and altitude [19–23]. 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 transforms ADP (adenosine 5'-diphosphate) and phosphate into the cellular energy carrier adenosine 5'-triphosphate (ATP) [24,25]. Hence, mtDNA genes play an important role in the thermal tolerance of organisms by regulating the availability of ATP, the demand of which is often increased under environmental extremes [26–29].

Positively selected mutations in mtDNA OXPHOS have been found to have an important impact on the ecology and evolutionary history of multiple taxa [20–22,27,30–32]. 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 some of the main processes known to lead to speciation [19,27,33–35].

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*. In 2018, Banguera-Hinestroza et al. [15] analyzed 827 samples of *Stylophora* covering a broad geographical range of this genus, including the full latitudinal (12 degrees of latitude) and environmental gradients of the Red Sea. They found a high number of mutational changes in the mtORF as well as in its adjacent genes *nd6* and *atp6*. Interestingly, these genes are part of an apparently recombinant region that strongly differentiates a putative hybrid lineage (*RS\_LinA*) from its sympatric 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 [15], which encodes a heat-shock protein well known for its role in stress response and climatic adaptation [36–40]. Results from Banguera-Hinestroza et al. [15] therefore suggest that the mtORF as well as *hsp70* may have both been involved in the adaptation of the ancestral and endemic *Stylophora* lineage to the different environmental regimes of the Red Sea, including extremely warm conditions in the Southern region.

On a broad geographical scale, pocilloporid corals are found in tropical areas worldwide. Confirmed records indicate that *Stylophora* and *Seriatopora* are restricted to the Pacific and Indian oceans, while *Madracis* and *Pocillopora* have wider distributions and are also found in the Atlantic Ocean and the Caribbean Sea [41]. *Stylophora* lineages occur throughout the entire Red Sea in regions with strong environmental differences. The northernmost 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 lowest nutrient input, while the southernmost region (Farasan islands, 20°N) is characterized by the warmest temperature regime (23.5 °C–33.9 °C), lowest salinity (37 PSU) and highest nutrient input [42–44]. Furthermore, the multiple environmental and geological changes recorded in the Red Sea since its formation around 30–14 Mya [45,46] have strongly influenced the patterns of endemism

and diversification of its fauna [47,48]. Some of these changes include hypersaline conditions, extreme variations in sea levels, extended periods of isolation from the Indian Ocean, and strong temperature fluctuations (up to 35 °C) [49,50]. The latter is considered an extreme condition for the growth and survival of coral reefs [43,51].

As mentioned above, there is mounting evidence that the mtORF reveals consistent phylogenetic and phylogeographic patterns in pocilloporids as well as patterns of population structure linked to environmental conditions, as barcode genes do. However, this mitochondrial region has been suggested to be a pseudogene [52,53], a hypothesis that has not been tested so far. Testing this hypothesis might help us elucidate whether evolutionary changes in the mtORF recapitulate adaptation and speciation in pocilloporids. Therefore, the aim of the current study was twofold: (a) to test whether the mtORF encodes and expresses a functional protein, by investigating the mtORF origin and exploring clues about its putative structure and function in pocilloporids; and (b) to search for signatures of selection in the mtORF of *Stylophora* species inhabiting a broad range of environmental conditions in the Red Sea.

If not functional or expressed, long ORFs (more than 600 bp in the present case) are expected to accumulate mutations [54]. Indeed, one of the clearest signals of pseudogenization in a putative ORF is the presence of multiple stop codons or frameshift mutations [55,56]. Hence, as a first step we analyzed the hundreds of available mtORF sequences from *Pocillopora*, *Seriatopora*, and *Stylophora*, in search for stop codons. In addition, we performed analyses of codon usage bias and tested for signatures of natural selection in representative pocilloporid sequences.

As a second step, we carried out a whole mitogenome comparative analysis to identify the first appearance of this mtORF in the evolutionary history of scleractinian corals, and used computational methods [57–60] to analyze its structure and function in *Madracis*, *Pocillopora*, *Seriatopora*, and *Stylophora*, including the two *Stylophora* lineages previously identified in the Red Sea [15]. Computational methods have been shown to predict reliably protein structure and function in non-model organisms [61]. Finally, we tested whether the mtORF may be involved in local adaptation by searching for signals of positive selection in the sequences of *Stylophora* lineages, particularly those distributed over a wide temperature gradient in the Red Sea.

## 2. Materials and Methods

### 2.1. Evolutionary Origin of the mtORF

Complete mitochondrial genomes of all pocilloporids (*sensu* [62]), including *Madracis mirabilis* (NC011160), *Stylophora pistillata* (accession: EU400214), *Seriatopora hystrix* (accession: EF633600), *Seriatopora caliendrum* (EF633601), *Pocillopora damicornis* (accession: EU400213), and *Pocillopora eydouxi* (accession: EF526303) were downloaded from the NCBI database and aligned with mitogenomes of the two Red Sea *Stylophora* lineages identified in [15] (Supplementary Data S1): the endemic *Stylophora* lineage from the Red Sea (*RS\_LinB*) and its putative hybrid with *S. pistillata* (*RS\_linA*).

Full mitogenome alignments were performed using the software LASTZ [63] and visualized using AliTV [64]. The alignments were contrasted against a phylogeny of pocilloporids built using the most variable mitochondrial genes in these species (*nd2*, *nd6*, *atp6*, mtORF, and *nd4*) [15]. This phylogenetic tree was constructed using the maximum-likelihood approach implemented in PhyML [65] available at <http://www.phylogeny.fr/index.cgi> [66,67]. The genera *Madrepora*, *Polycyathus* and *Astrangia* were included as outgroups, following the scleractinian phylogeny of Chuang et al. [68]. Branch support was evaluated using the approximate Likelihood Ratio Test (aLRT) [69].

### 2.2. Characterization of the mtORF Gene

To test for signals of pseudogenization, a total of 1830 mtORF nucleotide sequences (of complete and partial lengths) of *Pocillopora*, *Seriatopora*, and *Stylophora* from the NCBI database, including more than 600 unpublished sequences, were translated using the coelenterate mitochondrial code in the program MEGA [70] and scanned for stop codons (accession numbers are provided in Supplementary Table S1).

The start and end codons of the putative mtORF were identified by comparison with the protein sequences reported by Flot and Tillier [9]. To ensure the reliability of the data, sequences of *Stylophora* and *Pocillopora* from the Red Sea were taken exclusively from our collection [15,71]. Furthermore, we used the program Tandem Repeats Finder [72] to search for repeats in representative sequences.

### 2.3. Annotation of Protein Domains and Detection of Intrinsic Disorder

mtORF sequences extracted from full mitochondrial genomes (see above; Supplementary Data S2) were analyzed using a set of protein prediction approaches available at the PSIPRED Protein Sequence Analysis Workbench ([www.bioinf.cs.ucl.ac.uk/psipred](http://www.bioinf.cs.ucl.ac.uk/psipred) [73]) and the InterProScan data base [59,74,75]. These methods allow testing for transmembrane domains (TMDs) and different degrees of disorder in protein sequences.

The presence of TMDs was investigated using several predictors, including: (i) TMHMM Server 2.0, which predicts transmembrane helices based on a Hidden Markov Model (HMM). This method returns the posterior probabilities that each residue sits inside of the cell, outside of it or in a transmembrane helix (TMH), with low probabilities meaning weak support for TMHs [57,58]; (ii) the PSIPRED approach (Protein Structure Prediction) by Jones [76], a highly accurate method that uses two feed-forward neural networks to predict secondary structure in outputs generated using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool) [60,77]; (iii) the MEMSAT3 method (MEMbrane protein Structure And Topology 3), which predicts the secondary structure of transmembrane protein using multiple alignments produced by PSI-BLAST and by scoring log-likelihoods ratios through different topological models to gather the consensus [61,73,78,79]; and (iv) the MEMSAT-SVM approach using Support Vectors Machines, which are binary classifiers used to categorize residue preferences before combining them into a probabilistic framework [61].

To distinguish TM regions from signal peptides (SP), we used signalIP 4.1 [80,81], available at [www.expasy.org](http://www.expasy.org). Furthermore, the likelihood of a transmembrane helix being involved in the formation of pore-lining regions was calculated using MEMSAT. These regions run parallel to transmembrane helices and are vital for biological processes such as the transport of ions and molecules across membranes [61].

Disordered regions were identified using the consensus of several methods including DISOPRED [82–84], IUPred2A [85], and MobiDB [86]. DISOPRED implements a neural network combined with support vector machines (SVM), and in addition to disordered regions, it also provides information about possible binding of intrinsically disordered proteins to substrates. IUPred is a knowledge-based approach that provides a per-residue profile of the degree of disorder. In contrast, MobiDB integrates annotations of disordered regions through several other databases and methods [86]. For all methods, probabilities larger than 0.5 correspond to 95% confidence of identifying a true positive disordered region.

### 2.4. Structural and Functional Annotation

In order to identify homologs of mtORF among sequences of known structure and/or function, we used HMMER (<http://www.ebi.ac.uk/Tools/hmmer>) [87] and the FFAS server (<http://ffas.godziklab.org/>) [88]. HMMER uses profile HMMs to explore multiple profile data bases such as Pfam [89]. This approach has a high sensitivity for detecting remote homologs. Furthermore, a multiple-sequence alignment was build using Promals [90] and input to the FFAS server. The FFAS method integrates PSI-BLAST [91] and fold-recognition methods to search homologous proteins in several databases of model organisms, including Pfam [89], SCOP [92] and the non-redundant database (nr85), which contains sequence information from several sources [93,94].

Functional annotation was performed using FFPred 2.0 [95], which relies on the GOA (Gene Ontology Annotation data base) to predict Gene Ontology (GO) terms. The GO classification distinguishes among macromolecular interactions, biological processes, molecular function and cellular components. The FFPred method reports the posterior probability of a functional annotation in terms

of reliability, which includes sensitivity, specificity and precision [96], by searching throughout a large data set of already known proteins and annotations from eukaryotes [95].

Furthermore, the prediction of structural domains for mtORF sequence of each genus and lineage was carried out using pDomTHREADER [97]. This method retrieves structural hits with CATH domains annotated in the CATH protein data base using sequence and structural data [98]. The confidence in the existence of a given domain in the query proteins is reflected by its  $p$ -values. Low hits correspond to  $p$ -values  $\leq 0.1$ , medium hits to  $p$ -values  $\leq 0.01$ , high hits to  $p$ -values  $\leq 0.001$  and highly accurate hits to  $p$ -values  $\leq 0.0001$  [60].

## 2.5. Signatures of Selection in a Family Framework and Analysis of Codon Usage

### 2.5.1. Selection

The hypothesis that the mtORF of pocilloporid corals encodes a functional protein was also tested by performing an evolutionary rate analysis across species. If this hypothesis is true, some form of selection should be observed in the mtORF-encoded putative protein. For instance, the effect of the constraints due to the preservation of structure, function and/or gene expression should overall be manifest as negative or purifying selection.

Patterns of selection in the mtORF of Pocilloporidae were investigated using the PAML package [99]. The overall nonsynonymous/synonymous substitution ratio ( $\omega = dN/dS$ ) as well as the relative rate of substitution between pairs of mtORF orthologs were calculated by constructing a maximum-likelihood tree, as described above, as well as a codon-based multiple-sequence alignment using the software pal2nal [100].  $dN/dS$  was calculated using the program codeml with the following parameters: model = 0, Nsites = 0, icode = 4.

In addition, the evolutionary models implemented in PAML were used to test site-specific variation in the mtORF evolutionary rate. Specifically, three likelihood ratio tests (LRTs) were carried out comparing first a neutral model allowing only  $\omega = 0$  and  $\omega = 1$  with uniform selective pressure among sites (M0) versus a model that allows variable selection among sites (M3), and second a model that assumes a beta distribution for  $\omega$  allowing purifying selection (M7) versus a model that allows for positive selection (M8). Sites evolving under positive selection were identified using the Bayes Empirical Bayes approach [101]. By contrasting these models, the likelihood of identifying whether multiple classes of rates provide a better explanation of the data than a single class is improved.

### 2.5.2. Codon Usage Bias

In order to quantify codon usage bias in the mtORF, we calculated the codon adaptation index (CAI) [102]. The CAI quantifies the optimality of the codon composition of a gene by measuring the bias of its set of codons towards the codon frequencies in a reference table computed from a set of highly expressed genes. In the absence of gene expression data specific for mitochondrial genes in pocilloporids, this reference table was constructed for each species by simply counting the codons of all protein coding genes in their mitogenomes. This procedure relies on the assumption that most mitochondrial genes are optimally adapted for gene expression [103,104].

To assert whether the CAI was high enough to be considered biased toward codon optimality, we used a method to estimate the expected CAI (eCAI) [104]. This was done by simulating sequences of identical amino acid composition with respect to the query sequence, while accounting for the relative frequency of synonymous codons in the reference table. The resulting null distribution provides an upper confidence interval for significantly biased CAI. eCAI values reported here were calculated at the 99% confidence interval. Because circular genomes are known to have strong context-dependent nucleotide compositional biases, CAI analyses were also performed for mitochondrial genes located at neighboring genomic positions with respect to the mtORF.

## 2.6. Signatures of Adaptive Evolution in the mtORF of *Stylophora* Inhabiting Different Environments

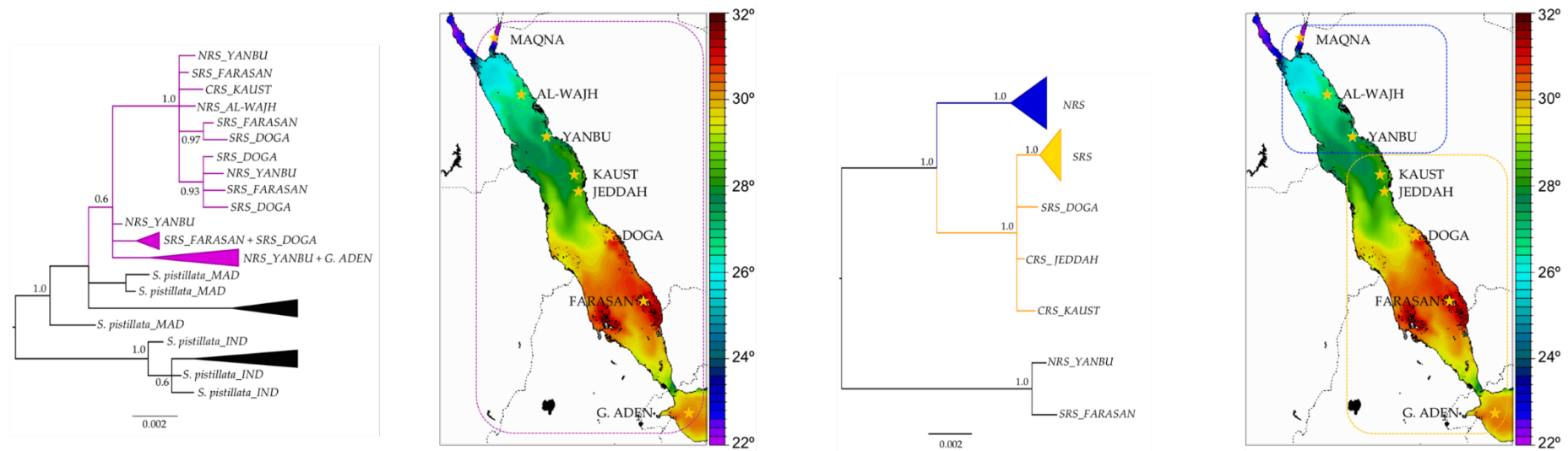
The potential involvement of the mtORF protein in local adaptation was tested by searching for signals of selection in the translated mtORF sequences. For this, mtORF haplotypes belonging to Red Sea *Stylophora* lineages as well as to *Stylophora* species from other oceanic regions were used. Samples included in these analyses (N = 827) belonged to the same set of samples analysed in Banguera-Hinestroza et al. [15], from which the definitions of *Stylophora* lineages, clades, subclades and populations were taken. Briefly, two *Stylophora* lineages were distinguished within two highly divergent clades: Clade 1 included 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 subclades, each grouping specimens distributed either in the northern areas of the Red Sea or in the central-southern regions (Figure 1). For analyses of selection, all *Stylophora* sequences were first evaluated 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 *RS\_LinA* and also in sequences from the northern and southern populations of *RS\_LinB*.

Signatures of selection at individual sites on the mtORF-encoded protein were inferred using several approaches implemented in the Datamonkey webserver ([www.datamonkey.org](http://www.datamonkey.org)) [105,106]. (i) The Fixed Effects Likelihood method (FEL) detects selection by first estimating branch lengths and substitution rate parameters in a given phylogeny when the corresponding coding-region alignment is provided. After the calculation of these two parameters, the method works 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 [107]. (ii) The Mixed Effects Model of Evolution (MEME) identifies episodic positive selection at individual sites [108] by combining both fixed models and random-effects models. (iii) The Branch-Site REL model (BS-REL) uses likelihood computations to infer the variation of dN and dS over branches and per site [109,110]. Statistical significance levels were set at a *p*-value of 0.1 for MEME and FEL and of 0.5 for BS-REL as recommended by the authors.

Furthermore, the M8 model of Yang et al. [111] and the Mechanistic–Empirical Combination (MEC) model from Doron-Faigenboim and Pupko [112] were applied to the mtORF protein using the Selecton webserver [113,114]. These approaches find the model that best fits the data by performing a likelihood ratio test (LRT) over the whole alignment. Here, a null model in which positive selection is not allowed (i.e., M8a; no-codons with dN/dS > 1) was compared against a general model that assumes positive selection (i.e., dN/dS > 1) as in the M8 and the MEC model. If positive selection was detected, a Bayes Empirical Bayes (BEB) approach was used to calculate the posterior probabilities that sites underwent positive selection [101]. 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 was used to compare both models and the significance test was considered passed whenever the AIC score of M8 or MEC was 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. [114].

The last test used to detect selection was the codon-based Z-test of positive selection of Nei and Gojobori [115] implemented in MEGA 7 [70]. This test allows discrimination between positive and purifying selection by performing pairwise comparisons of the relative rates of synonymous (dS) and non-synonymous substitutions (dN) and their variances [70] among all haplotypes, thereby enabling inferences of selection when specific clades or subclades are compared. The null hypothesis of dN = dS and the alternative hypotheses of dN > dS (positive/diversifying selection) and dN < dS (purifying selection) are tested using a 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 [116].



**Figure 1.** Phylogenetic position of Red Sea haplotypes and their distribution along the latitudinal and environmental gradients of the Red Sea. Graphs have been modified from those published by Banguera-Hinestroza et al. [15]. (a) Clade 1: *Stylophora RS\_LinA* (magenta), *Stylophora pistillata* from Madagascar (MAD), and *Stylophora pistillata* from the Indo-Pacific region (IND). (b) Clade 2. *Stylophora RS\_LinB*. Northern and southern populations are highlighted in blue and yellow respectively. Yellow stars indicate geographical locations. Haplotypes are named according to the sampling sites in which they showed the highest frequency (in the southern (SRS), northern (NRS), or central (CRS) Red Sea).

### 3. Results

#### 3.1. The mtORF Occurs in All Pocilloporids and Does Not Exhibit Stop Codons

Phylogenetic analyses of the complete mitogenomes of corals within Pocilloporidae (genera *Madracis*, *Pocillopora*, *Seriatopora*, and *Stylophora*) in a phylogenetic framework and using three outgroups (*Madrepora*, *Polyciathus*, and *Astrangia*) showed that the mtORF likely appeared in the common ancestor of the family Pocilloporidae (Figure 2), which according to the calibrated coral phylogeny of Chuang et al. [68] probably existed 150–250 Mya. Stop codons were not found in the translated mtORF sequences of *Madracis*, *Pocillopora*, *Seriatopora*, and *Stylophora* (N = 1830), except in 3 out of 431 sequences from *Seriatopora* (accessions KR150027, KR150037, and KR150052 [13]; Supplementary Table S1). Examination of the chromatograms kindly provided by the authors showed that the apparent stop codons were the result of base calling errors, hence none of the 1830 sequences examined contains actually stop codons.

#### 3.2. The mtORF-Encoded Proteins of Pocilloporids Vary in Length and in Aliphatic Indices

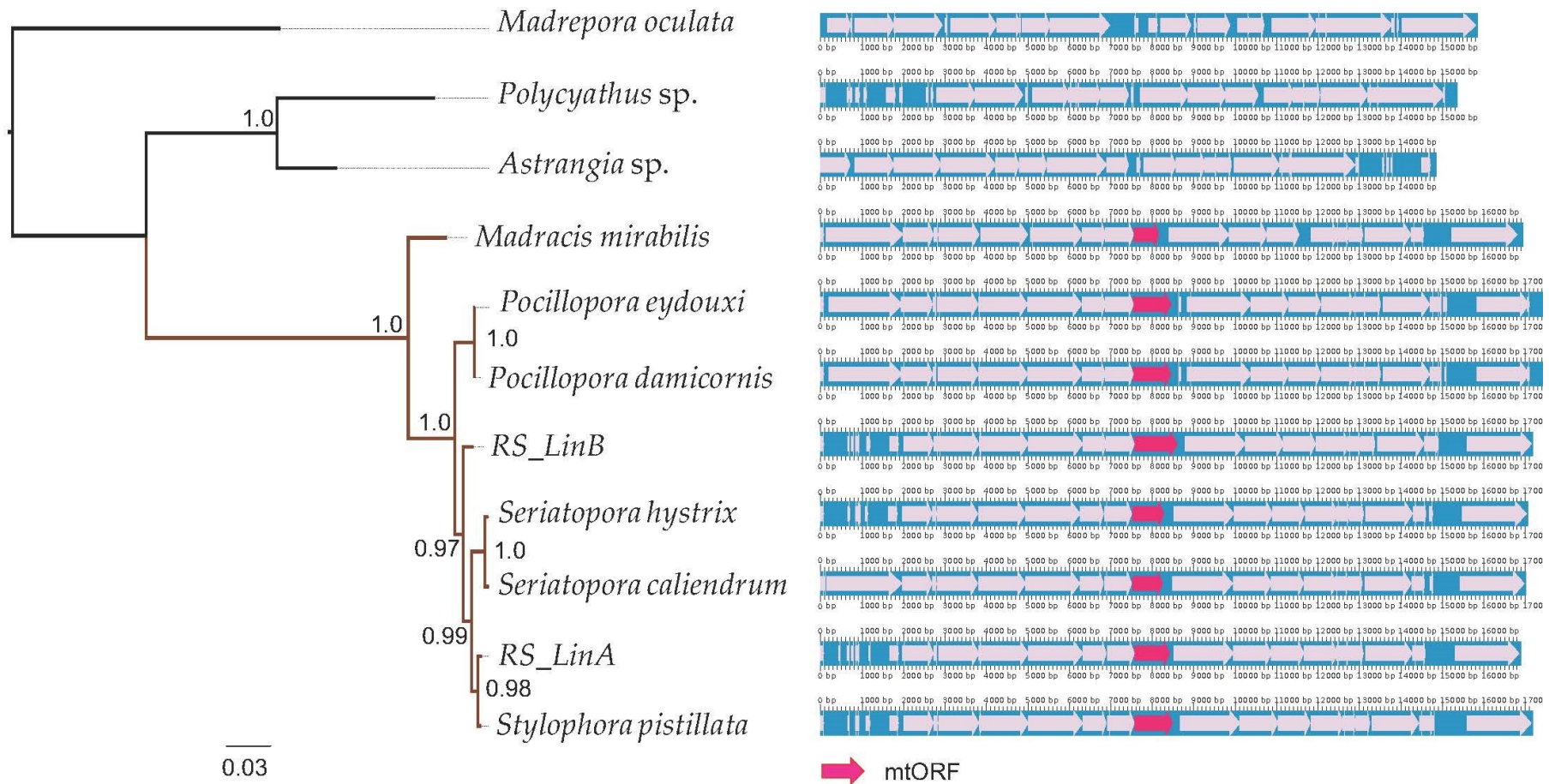
Translating the mtORF sequences of each genus resulted in proteins of different lengths. The shortest was found in *Madracis* with 221 amino acids and the longest was found in *RS\_LinB* with 362 amino acids, followed by *Stylophora pistillata* with 309 amino acids, *Pocillopora* with 302, *Stylophora RS\_LinA* with 301, *Seriatopora caliendrum* with 282, and *Seriatopora hystrix* with 269. The alignment of the conserved regions of the mtORF-encoded protein among genera is shown in Figure 3. Differences in length among and within *Seriatopora* and *Stylophora* sequences were found to be associated with the presence of duplicated and polymorphic tandem repeats (TRs) (Table 1) that generated long insertion-deletion (indels) in the alignments. In *Stylophora*, some of these TRs were exclusively found in sequences from Red Sea specimens of *RS\_LinB* and motifs differed in length, number and/or nucleotide composition between southern and northern populations as well as among *Stylophora* lineages (Supplementary Table S2).

Furthermore, the mtORF-encoded proteins for these genera differed also in their aliphatic indices, higher values of which indicate a higher thermostability [117]: calculated values were 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*.

#### 3.3. mtORF-Encoded Proteins Contain Transmembrane Domains and Intrinsically Disordered Regions

Exploration of domain architecture using five predictors of secondary structure and transmembrane domains (TMDs) suggested that the mtORF encodes a transmembrane protein. From this point on we shall refer to this protein as TMP362 (short for TransMembrane Protein 362 amino acids long, as in the case of *RS\_LinB*) and to the corresponding gene as *tmp362*. In *Madracis*, TMP362 comprises three well-supported TMDs (one at the N-terminal side and two at the C-terminal side), versus two (one at each side) in the other pocilloporids genera (Figure 4 and Supplementary Figure S1). Often signal peptides can be misannotated as TMDs, but there was no evidence for signal peptides in TMP362 using MEMSAT or signalIP, suggesting that our predictions indeed correspond to TMDs. These annotations are not surprising since most mitochondrial protein-coding genes are associated to the mitochondrial membrane. A third TMD was predicted with high probability at the C-terminal in *Seriatopora* by the MEMSAT3 approach but was not supported by the other methods. If we assume that this third TMD really exists, it suggests that TMP362 comprised three TMDs in the common ancestor of all pocilloporids.





**Figure 2.** Family-level phylogeny showing that the mitochondrial Open Reading Frame (mtORF) likely appeared in the common ancestor of all pocilloporid corals. **Left:** maximum-likelihood tree based on a concatenation of most variable mtDNA genes (*nd2*, *nd6*, *atp6*, mtORF, and *nd4*). **Right:** full mitogenome alignments with the mtORF highlighted in red.

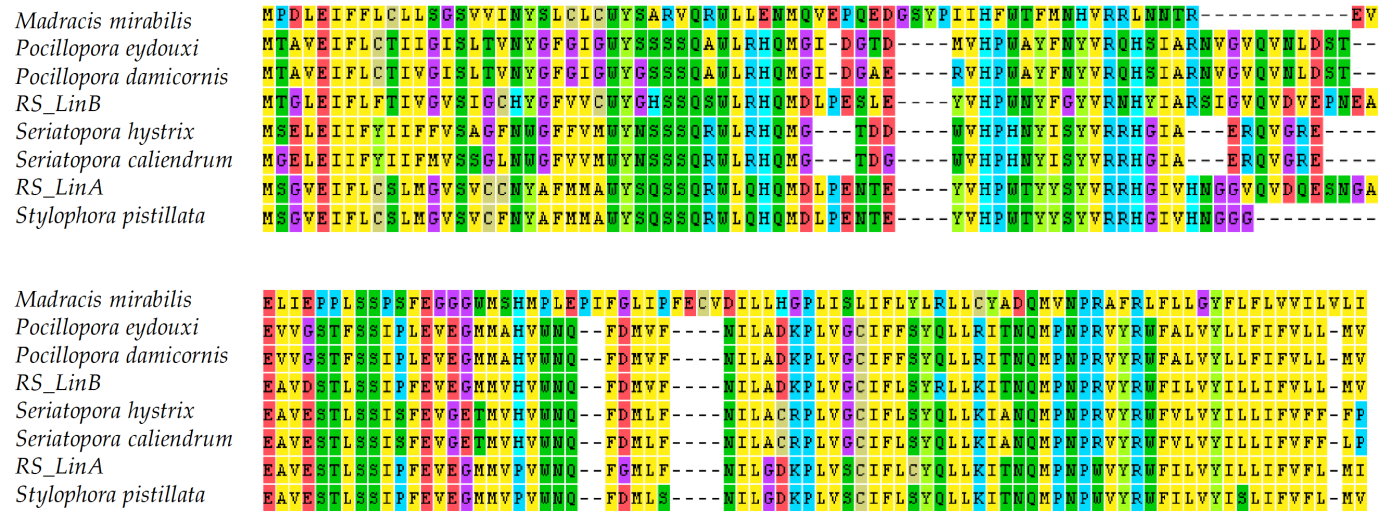


Figure 3. Conserved blocks in the mtORF-encoded protein of pocilloporid corals.

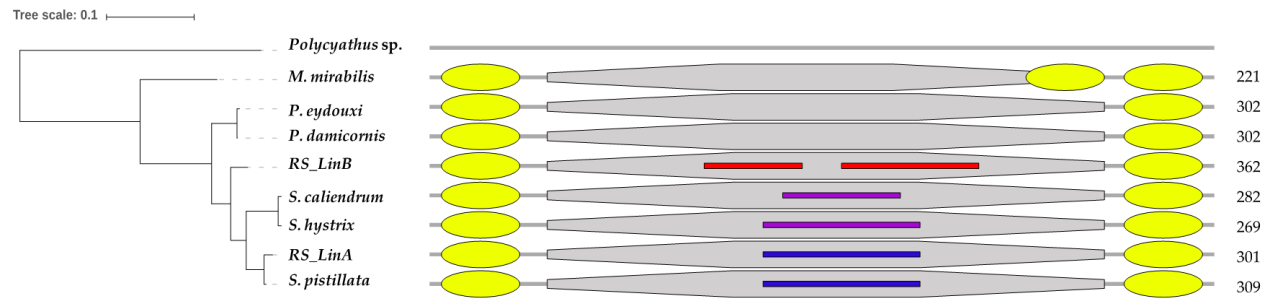
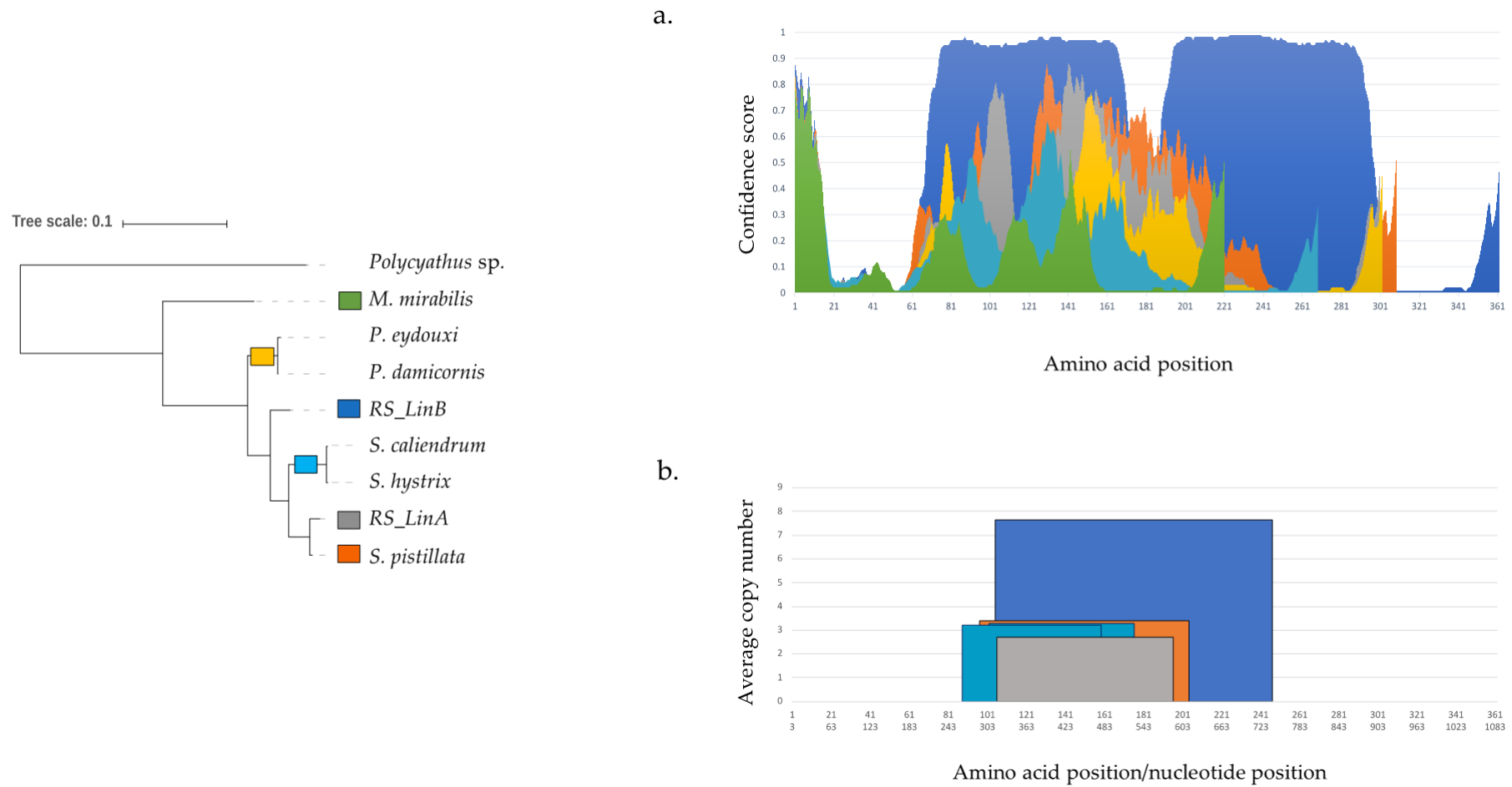


Figure 4. Domain organization. **Left:** maximum-likelihood phylogeny of the family Pocilloporidae based on the concatenation of *nd2*, *nd6*, *atp6*, *tmp362*, and *nd4*. **Right:** schematic representation of the TMP32 orthologs and their annotated domains. Sequences, domains, and IDRs differ in length, but for illustration purposes they are shown aligned. Yellow ellipses represent transmembrane domains (TMDs); grey octagons, hydrolase domain (HDs); thin rectangles, disordered regions found as the consensus of several methods (red) or only reported by DISOPRED (blue), or InterproScan and IUPred2A (magenta). Numbers on the right-hand side indicate the length of each protein.

**Table 1.** Tandem repeats in the mtORF sequences of representative lineages of *Stylophora* and *Seriatopora* corals.

	Indices	Period Size	Copy Number	Consensus Size	Percent Indels
<i>RS_LinB</i>	315–741	27	16.9	27	11
	318–741	54	8.4	54	13
	371–613	75	3.3	75	5
	501–634	21	6.1	21	10
	486–741	69	3.5	69	9
<i>S. hystrix</i>	310–487	51	3.5	51	0
	352–430	24	3.2	24	10
	310–528	51	4.3	51	1
	306–525	102	2.2	102	1
<i>S. caliendrum</i>	265–436	51	3.4	51	0
	265–477	102	2.1	102	1
	265–477	51	4.2	51	1
<i>RS_LinA</i>	317–424	39	2.6	42	4
	365–440	21	3.8	21	10
	350–444	39	2.4	39	7
	480–587	51	2.1	51	0
<i>S. pistillata</i>	290–417	39	3.2	42	6
	338–413	21	3.8	21	10
	453–611	51	3.1	51	0

In addition, analyses revealed intrinsically disordered regions (IDRs) in TMP362 (Figure 5). In *RS\_LinB* IDRs span 216 amino acids, from residue 68 to 173 and from residue 188 to 297. Results from DISOPRED, which is specifically trained to identify disordered regions by their intrinsic characteristics [118], indicated that IDRs are not fully conserved among *Stylophora* lineages. In *Stylophora pistillata*, 79 amino acids were predicted as IDRs (from residue 123 to 201), whereas in the protein of *RS\_LinA* IDRs cover two small regions with a total of 27 amino acids, from residue 134 to 153 and from 160 to 166 (Figure 5a). These residues were undetectable by other predictors centered on sequence information, such as IUPred. This predictor is based on a pairwise energy-like parameter derived from sequences in globular proteins [85] and suggested disordered regions in *Seriatopora* and *RS\_LinB* but not in other *Stylophora* lineages (Figure 4). Noticeably, in all species the predicted IDRs coincide with a region rich in duplicated and polymorphic TRs (Table 1, Figure 5b, Supplementary Table S2).



**Figure 5.** Intrinsic disorder profile and tandem repeats. **(a).** Intrinsically disordered regions in pocilloporid corals predicted by DISOPRED3. Amino acids with a confidence score above 0.5 are considered disordered, predicted at a false positive rate threshold of 0.05 [83,85]. **(b).** Tandem repeats in representative sequences of *tmp362* of *Stylophora* and *Seriatopora* species.

### 3.4. Remote Homologs Suggest that mtORF has a Hydrolase Domain

Despite the presence of disordered regions, a large fraction of internal domain of TMP362 seems to be structured. The FFAS method detected a far, although significant homology with a known structure from the Protein Data Base (PDB ID: 3b57, chain A) that corresponds to a bacterial protein known as lin9899, not characterized experimentally till now but with a hydrolase (HD) domain annotated based on homology. The HD domain belongs to the all- $\alpha$  structural class that is highly conserved across Bacteria, Archaea, and Eukarya and is involved in nucleic acid metabolism and signaling [119,120]. This domain is often associated to a metal-dependent phosphohydrolase function (EC 3.1).

Interestingly, the HD domain covers a large fraction of the internal domain of the TMP362, spanning approximately 100 residues (residues 25 to 132 in 3b57, chain A) close to the TMP362 C-terminal end (a schematic view is presented in Figure 4). The canonical catalytic site in the HD domain involves histidine (H) and aspartic (D) residues responsible for metal binding. A multiple sequence alignment with the homolog of the known structure (3b57) suggests that a fully conserved histidine (H, at position 23 with respect to 3b57), as well as several relatively conserved aspartic residues (D, e.g., at position 104 with respect to 3b57), might still be involved in a hydrolase activity or might be the relict of a former HD functional domain. The finding of a HD domain homolog is supported by annotations carried out using the Gene Ontology Annotation system (GOA), which found GO terms uncovering 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 activity ( $0.59 < P < 0.68$ ). Furthermore, the catalytic function of a hydrolase was supported with largely significant posterior probabilities ( $P > 0.75$ ) (Table 2).

### 3.5. Biological Processes and Cellular Localization Support the Hypothesis of a Mitochondrial Transmembrane Protein

Two main biological processes were suggested by GO terms for the role of TMP362 in all the pocilloporid genera and lineages examined (Table 3): transport ( $0.80 < P < 0.87$ ) and regulation of metabolic processes ( $0.71 < P < 0.85$ ), except in *RS\_LinB* where the 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 TMP362 of *Madracis* and *Seriatopora* (i.e., ion transmembrane transport and nucleotide metabolic process; Table 3). 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 moderate to high probabilities of being part of an organelle membrane (i.e., mitochondria;  $0.70 < P < 0.80$ ). Overall, these analyses support the hypothesis that TMP362 is a mitochondrial transmembrane protein with possible transport, signaling and/or metabolic functions.

**Table 2.** Predicted molecular functions for TMP362 of pocilloporid corals.

GO Term	Molecular Function	Posterior Probabilities					
		<i>Madracis</i>	<i>Pocillopora</i>	<i>Seriatopora</i>	<i>Stylophora</i> <i>RS_LinB</i>	<i>Stylophora</i> <i>RS_LinA</i>	<i>Stylophora</i> <i>pistillata</i>
GO:0005216	ion channel activity	0.913	-	0.628	-	-	0.581
GO:0016817	<b>hydrolase activity, acting on acid anhydrides</b>	<b>0.902</b>	<b>0.862</b>	<b>0.561</b>	<b>0.824</b>	<b>0.695</b>	<b>0.626</b>
GO:0015075	ion transmembrane transporter activity	0.88	-	0.654	-	-	-
GO:0022890	inorganic cation transmembrane transporter activity	0.864	-	-	-	-	-
GO:0008324	cation transmembrane transporter activity	0.858	-	-	-	-	-
GO:0005524	ATP binding	0.833	-	0.832	-	0.739	0.502
GO:0046873	metal ion transmembrane transporter activity	0.787	-	-	-	-	-
GO:0015077	monovalent inorganic cation transmembrane transporter activity	0.749	-	-	-	-	-
GO:0003824	<b>catalytic activity</b>	<b>0.748</b>	<b>0.762</b>	<b>0.771</b>	<b>0.528</b>	<b>0.739</b>	<b>0.681</b>
GO:0016818	<b>hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides</b>	<b>0.707</b>	<b>0.602</b>	<b>0.602</b>	<b>0.666</b>	<b>0.695</b>	<b>0.619</b>
GO:0022857	transmembrane transporter activity	0.706	0.56	0.696	-	-	0.502
GO:0005261	cation channel activity	0.688	-	-	-	-	-
GO:0005215	transporter activity	0.67	0.548	0.684	-	0.561	0.515
GO:0035639	purine ribonucleoside triphosphate binding	0.639	0.538	0.706	-	0.733	0.538
GO:0008092	<b>cytoskeletal protein binding</b>	<b>0.588</b>	<b>0.592</b>	<b>0.598</b>	<b>0.682</b>	<b>0.62</b>	<b>0.676</b>
GO:0000166	nucleotide binding	-	0.52	0.71	-	0.72	-
GO:0001882	nucleoside binding	-	-	0.69	-	0.752	-
GO:0032549	ribonucleoside binding	-	-	0.682	-	0.779	-
GO:0017076	purine nucleotide binding	-	-	0.604	-	0.779	0.557
GO:0022891	substrate-specific transmembrane transporter activity	-	-	0.68	-	-	-
GO:0030554	adenyl nucleotide binding	-	-	0.648	-	0.547	-
GO:0016301	kinase activity	-	0.53	0.549	0.617	0.857	0.655

**Table 3.** Predicted biological processes for TMP362 of pocilloporid corals.

GO Term	Biological Process	Posterior Probabilities					
		<i>Madracis</i>	<i>Pocillopora</i>	<i>Seriatopora</i>	<i>Stylophora</i> <i>RS_LinB</i>	<i>Stylophora</i> <i>RS_LinA</i>	<i>Stylophora</i> <i>pistillata</i>
GO:0006810	<b>transport</b>	<b>0.862</b>	<b>0.847</b>	<b>0.874</b>	<b>0.803</b>	<b>0.824</b>	<b>0.82</b>
GO:0034220	ion transmembrane transport	0.845	-	0.689	-	-	-
GO:0019222	<b>regulation of metabolic process</b>	<b>0.813</b>	<b>0.853</b>	<b>0.795</b>	<b>0.711</b>	<b>0.849</b>	<b>0.791</b>
GO:0007166	<b>cell surface receptor signaling pathway</b>	<b>0.804</b>	<b>0.63</b>	<b>0.564</b>	<b>0.79</b>	<b>0.662</b>	<b>0.717</b>
GO:0009117	nucleotide metabolic process	0.728	-	0.697	-	-	-
GO:0051649	<b>establishment of localization in cell</b>	<b>0.694</b>	<b>0.692</b>	<b>0.721</b>	<b>0.538</b>	<b>0.698</b>	<b>0.659</b>
GO:0051641	<b>cellular localization</b>	<b>0.647</b>	<b>0.631</b>	<b>0.64</b>	<b>0.632</b>	<b>0.643</b>	<b>0.636</b>

### 3.6. Strong Signatures for Positive and Negative Selection are Detected in the mtORF of Pocilloporid Corals

If the mtORF encodes a functional protein, then some form of selection for the preservation of expression and/or function should be at play. In order to test this hypothesis, an evolutionary rate analysis of the *tmp362* gene across species was carried out. First, the evolutionary rate of pairs of *tmp362* orthologs (i.e., from the 8 species of pocilloporids included here) was evaluated under the assumption that there was a single average rate along the sequence. This analysis revealed that on average *tmp362* has experienced negative selection (i.e.,  $dN/dS < 1.0$ ). For instance, with respect to *Madracis*, *Pocillopora* is on average evolving at a  $dN/dS = 0.53$ , *Seriatopora* at 0.44, and *RS\_LinB* and *RS\_LinA* at 0.43 and 0.6, respectively. The assumption of a single rate along an entire gene often averages out signatures of selection operating in few sites along the sequence. Thus, it is possible that while some sites are evolving under strong negative selection, some are positively selected ( $dN/dS > 1.0$ ) or experience neutral evolution (i.e.,  $dN/dS = 1.0$ ). To explore these alternatives, a likelihood ratio test (LRT) was implemented. This test contrasts a model that assumes a single constant rate (M0), versus a model that allows for a set of discrete classes of sites (M3). By comparing these two models, it can be asserted whether multiple classes of rates do indeed provide a better explanation of the data than a single class. Here, the LRT provided strong evidence in favor of the M3 model, suggesting that distinct groups of sites in *tmp362* are evolving at considerably different rates (M3 & M0:  $2\Delta\text{lnL} = 95.48$ ;  $df = 4$ ,  $p\text{-value} < 0.001$ ).

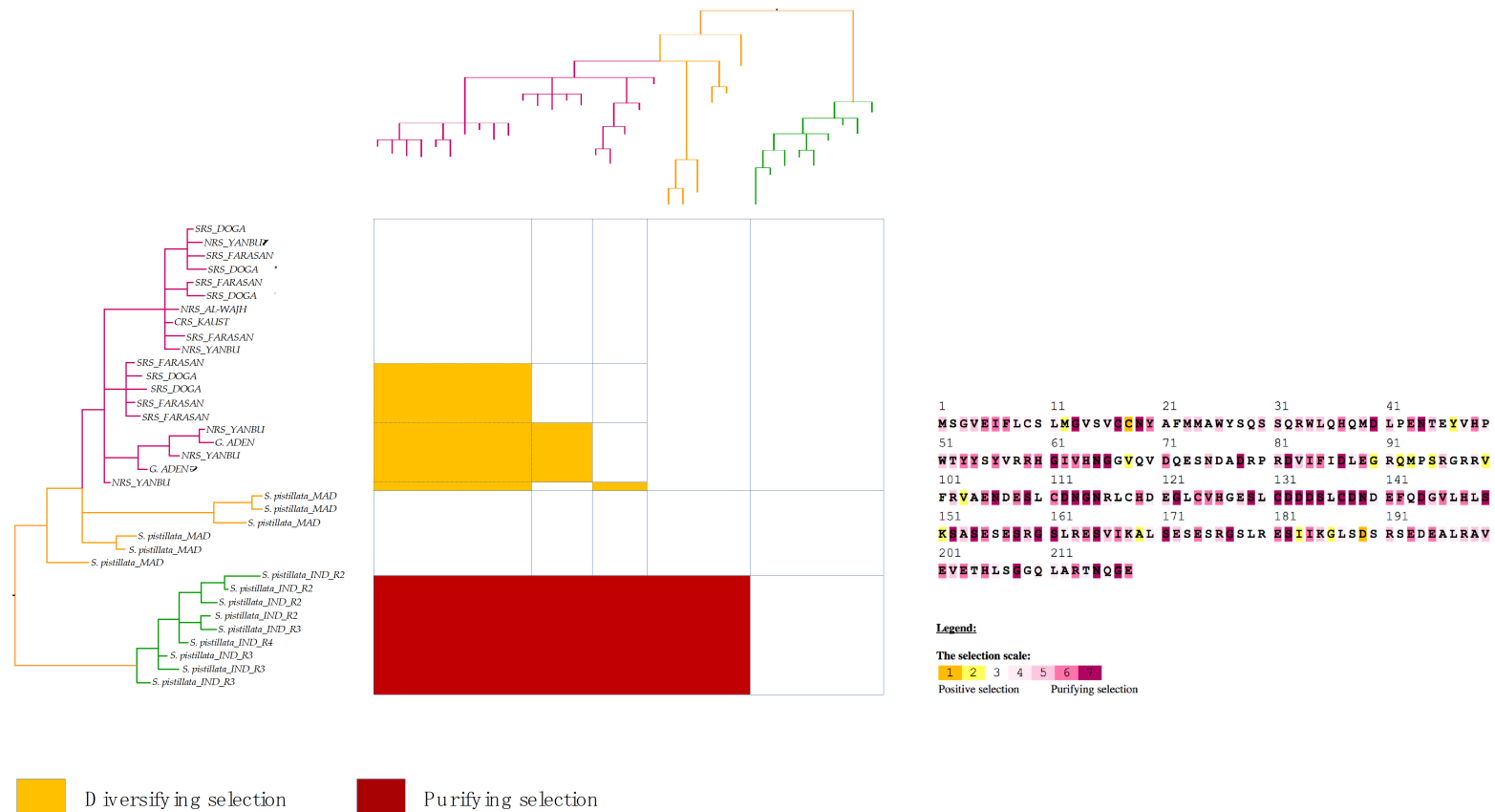
In addition, a comparison of M3 models using an increasing number of rate classes provided significant support for three versus two, but not for four versus three classes (M3 (K = 3 and K = 2):  $2\Delta\text{lnL} = 7.95$ ;  $df = 2$ ,  $p\text{-value} < 0.05$ ). This analysis showed that 36% of sites are under strong purifying selection ( $dN/dS = 0.1$ ), whereas 18% are experiencing positive selection ( $dN/dS = 4.3$ ). The remaining 46% are evolving in a relatively neutral way. A Bayes Empirical Bayes approach indicated that four of the positively selected sites (i.e., 27A, 34I, 47V and 65V, all of which located close to the N-terminal end of the protein) are associated to high posterior probabilities ( $P > 0.97$ ).

A codon bias analysis revealed that the codon usage in the *tmp362* of all pocilloporid coral species examined in this work is significantly biased with 99% confidence. The eCAI analysis gave evidence supporting that mitochondrial genes under similar context-dependent mutations are subject to similarly significant biases in codon usage as *tmp362*. This suggests that *tmp362* is fully functional and not a pseudogene, in accordance with the results of subsequent analyses (see below).

### 3.7. *Stylophora* Corals in the Red Sea Exhibit Signatures of Selection in the Their mtORF-Encoded Protein Along a Latitudinal Gradient

To test whether *tmp362* might play a role in adaptation to the environment, the hypothesis of selection was tested on populations of *Stylophora* corals adapted to a range of different environmental conditions in the Red Sea. *Stylophora* specimens from other regions were included for comparison. Amino acid sites under positive/diversifying selection at two sites were identified with the MEME approach when sequences from Clade 1 –*RS\_LinA* and *Stylophora pistillata* (i.e., specimens from Pacific and Indian Oceans, including Madagascar)– were evaluated together. Positive selection (at 5 sites) was also found when the analyses were restricted to Red Sea sequences from *RS\_LinA* and *RS\_LinB* (the phylogenetic position and distribution of these lineages are illustrated in Figure 1). However, positive selection was not detected when the alignment included the small section of TMP362 aligned among all *Stylophora* sequences (evaluating only unambiguously aligned regions) or when sequences from each *Stylophora* lineage were analysed separately.

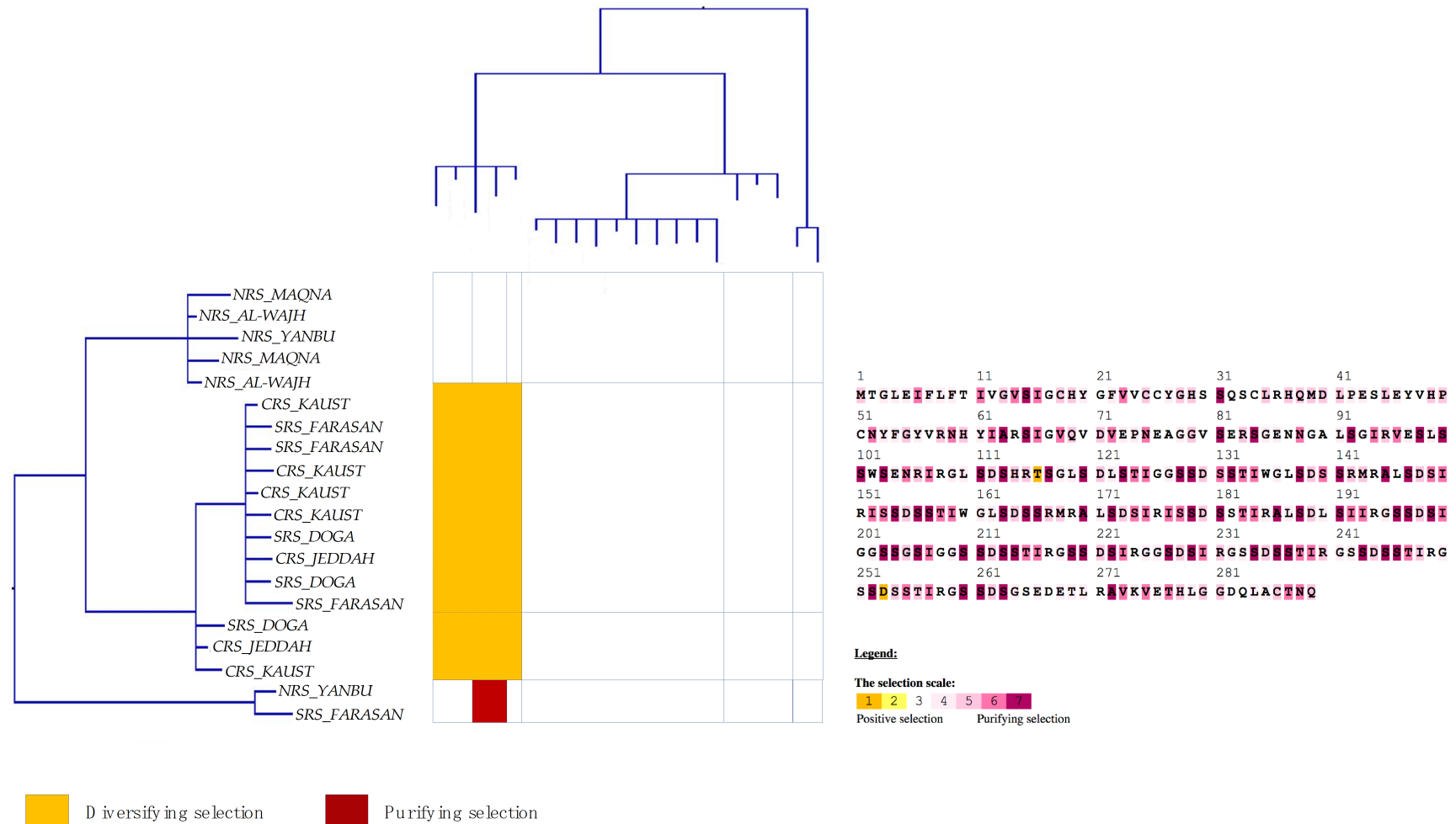
The M8 and MEC model recovered the highest number of single amino acid sites under positive selection in most cases (Figures 6 and 7). Nevertheless, only the MEC model showed significant values for most comparisons, namely: (i) in the TMP362 of *Stylophora RS\_LinA* and *Stylophora pistillata* at one site (the AIC score for MEC: 2846.897 was lower than for M8a: 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., MEC score = 1737.074; M8a score = 1745.86).



Probabilities:  $0.004 < p < 0.043$

**Figure 6.** Left: graphical view of the codon-based test of purifying and positive/diversifying selection (Nei and Gojobori [115]) in pairwise comparisons between haplotypes of *Stylophora* within Clade 1 (*RS\_LinA* plus Indian and Pacific Oceans samples). Right: sites under positive and purifying selection predicted by the MEC approach. The phylogenetic tree has been modified from Banguera-Hinestroza et al. [15]. For simplicity, haplotypes have been named according to their region of origin (see Figure 1).





Probabilities:  $0.004 < p < 0.043$

**Figure 7.** Left: graphical view of the codon-based test of purifying and positive/diversifying selection (Nei and Gojobori [115]) in pairwise comparisons between haplotypes of *Stylophora* within Clade 2 (*RS\_LinB*). Right: sites under positive and purifying selection predicted by the MEC approach. The phylogenetic tree has been modified from Banguera-Hinestroza et al. [15]. For simplicity, haplotypes have been named according to their region of origin (see Figure 1).

Analyses of sequences from the southern and northern group of *RS\_LinB* independently indicated sites under positive selection in the TMP362 of the southern group (AIC score for MEC: 2313.488, AIC score for M8a: 2317.705). This was corroborated by the results of the Nei-Gojobori method with highly significant  $p$ -values ( $0.003 < p < 0.043$ , Figure 6). Furthermore, positively selected sites were not found in the northern group of this lineage nor in sequences of *RS\_LinA* alone (except in the TMP362 of specimens whose haplotypes had the highest frequencies in the southern Red Sea or were restricted to YANBU, a sampling site found at the boundary between northern and central Red Sea oceanic provinces; Figure 1). 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 Figures 6 and 7 (the full matrices are available upon request).

Congruent with the finding in Section 3.6, amino acid sites under pervasive negative/purifying selection were found in all alignments of TMP362 using the FEL method: at 11 sites for all *Stylophora* lineages; at 7 sites for *Stylophora pistillata* and *RS\_LinA*, at 6 sites for *RS\_LinB* and also at 6 sites when *RS\_LinA* and *RS\_LinB* sequences were placed together. When the southern and northern group of *RS\_LinB* were analyzed separately, one site was found under purifying selection in the TMP362 of each group. Furthermore, purifying selection was also detected in mtORF-encoded protein of *Stylophora pistillata* using the Nei-Gojobori method ( $0.003 < p < 0.043$ ). BS-REL did not detect sites under selection along the branches of the phylogenetic tree of this genus.

#### 4. Discussion

The mitochondrial genome of plants and animals have been shown to be prone to loss and/or acquisition of new genes [7,121,122]. Some of these genes do not show homology with proteins of known function, but are fully functional [1,4,7]. In this work, we carried out a computational characterization of a putative protein-coding gene previously found in the mitochondrial genome of corals of the family Pocilloporidae. To test the hypothesis that the mtORF of pocilloporids is a functional gene, we examined single-gene and whole mitochondrial genome alignments, ran several predictors of protein features, performed structural and functional annotations, looked for signatures of selection, and investigated codon-usage biases. Particularly, we focused on previously reported *Stylophora* lineages [15] evolving along the full latitudinal and temperature gradients of the Red Sea, one of the hottest marine basins on Earth that experiences southern temperatures above 30 °C [43,44,123], the thermal limit for corals in most parts of the world [124].

Here we discuss our analyses broadly, first in the context of protein structure and function and second in light of the current knowledge on the role of structural changes in adaptation. In this regard, we explore whether the patterns of nucleotide substitutions and structural changes in the mtORF-encoded protein dubbed here TMP362 might be explained by the current knowledge of the role of mitochondrial proteins, tandem repeats (TRs) and intrinsically disordered regions (IDRs) in adaptation. This, we hope, will inform the current efforts to understand the signatures of environmental pressures on the mitochondrial genome of marine species, particularly on reef-building corals.

##### 4.1. Origin and Conservation of *tmp362*

Our analyses present substantial evidence suggesting that the mtORF gene *tmp362* arose in the most recent common ancestor of the family Pocilloporidae around 150–250 Mya. The consensus of several prediction methods revealed that *tmp362* encodes a transmembrane protein in all genera studied here, in agreement with its preliminary characterization by Flot and Tillier [9]. Overall, the competing hypothesis that *tmp362* is a pseudogene was not supported by our analyses. Indeed, an extensive analysis including 1830 partial and complete sequences of this gene, representative of several Pocilloporidae species, did not show a single stop codon in the translated protein. Furthermore, comparisons of evolutionary models revealed pervasive purifying selection, several sites under positive selection, and a bias in codon usage comparable to or even stronger than that showed by other mitochondrial genes in neighboring genomic regions.

#### 4.2. Putative Homology with a Bacterial Hydrolase Domain

Homology searches revealed the existence of a handful of significant far homologs of known structure. The top-ranking homologs correspond to bacterial proteins annotated with a hydrolase domain, which spans the internal region of TMP362. However, homology detection at low levels of significance is error-prone, hence this result should be interpreted with caution. First, although the hydrolase homology was detected in all mtORF orthologs when using a multi-alignment strategy, it was not detected in all pocilloporid mtORF sequences when analyzing them separately. Second, some predictors suggested the presence of  $\beta$ -sheets, but the hydrolase domain present in the far homolog is associated to an all- $\alpha$  fold.

Multiple disordered regions associated to tandem repeats (see discussion below), as predicted by our analyses in *Stylophora* and likely in *Seriatopora*, might have introduced several structural changes into the ancestral domain, and probably also explain the variations in secondary structure among genera. Although this suggests that the hydrolase domain might not be fully conserved among pocilloporids, an independent analysis of functional annotation using Gene Ontology also predicted a hydrolase catalytic function with highly significant posterior probabilities for all TMP362 orthologs studied here (Tables 2 and 3). This additional evidence supporting structural annotation, plus the fact that reef-building corals are holobionts harboring multiple symbiotic organisms including photosynthetic algae, viruses, and bacteria [125,126], lead us to propose that *tmp362* is of bacterial origin and was horizontally transferred to the mitochondrial genome of the common ancestor of all pocilloporid corals. Mutation accumulation mainly in form of tandem repeats, intragenomic recombination (as suggested by Banguera-Hinestroza [15]), and possibly, adaptation to varying environmental conditions, resulted in strong genotypic variation. Despite the nucleotide-level and structural diversity of *tmp362* among pocilloporid corals, its encoded protein remains apparently functional in all the genera and lineages examined.

#### 4.3. Putative Role of *tmp362* in Environmental Adaptation

##### 4.3.1. Tandem Repeats (TRs) and Intrinsically Disordered Regions (IDRs)

TRs were the main source of variation among sequences of *tmp362* in most pocilloporids. Noticeably, the highest number of repeat units was found in the lineage endemic to the Red Sea (*RS\_LinB*), in which copy number varies between 3 and 17 (Table 1). TRs were also found, although to a lesser extent, in *Seriatopora* and *Stylophora* lineages with broader distribution: in *RS\_LinA* (likely restricted to the Red Sea, Arabian Gulf, Gulf of Oman, and Gulf of Aden) and *Stylophora pistillata*, *S. caliendrum*, and *S. hystrix* (widely distributed in the Indo-Pacific region). Refer to [15] and [41,127] for the distribution of Red Sea and Indo-Pacific *Stylophora* lineages respectively. Moreover, in pocilloporids, TRs were associated with predicted IDRs in TMP362, except in *Pocillopora* and *Madracis* in which neither TRs nor IDRs were found (Figure 5).

Analysis of a large number of *tmp362* sequences (N = 827) from Banguera-Hinestroza et al. [15] revealed differences in number, length, and nucleotide composition of repeat motifs among *Stylophora* lineages inhabiting distinct oceanic basins. Conversely, indels and TRs motifs in *tmp362* sequences were found to differentiate northern and southern populations of *RS\_LinB* (restricted to the coldest and hottest areas of the Red Sea, respectively; Figure 1, Supplementary Table S2).

Interestingly, indels and TRs have been found to be associated with patterns of evolution, ecological diversification, and adaptation in a range of marine and terrestrial organisms, including viruses [128,129], bacteria [130] (reviewed in [131,132]), plants [133–135], and animals [136–138]. Indeed, the genome of organisms evolving under strong selective pressures often exhibit a proliferation of TRs (reviewed in [139]), which has been explained as the results of rapid adaptive changes [140–142] arisen in response to the environmental stress imposed by the colonization of new environments, drastic environmental changes, or global warming [132,134,139,143,144]. TRs are also involved in protein function [139,145], increase functional variability [146] and play key roles in the regulation of

gene expression in organisms growing under different stressors [132,147]. 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 that enhances protein activity and gene expression [148].

On the other hand, IDRs as predicted here for TMP362 of *Stylophora* and likely *Seriatopora* lineages, (Figure 5) have been found to be a common feature of the genome of eukaryotes (reviewed in [84,149,150]). These regions evolve mostly through the expansion of functional TRs (reviewed in [151]) and indels [152], and their degree of polymorphism has been hypothesized to be the result of the strength of evolutionary forces leading to adaptive changes [153,154]. Furthermore, IDRs confer functional advantages to a protein, providing the adaptability that allows them to bind to different partners (i.e., functional promiscuity) with different regions (domains) often found associated to different functions [149,155,156]. In fact, the functional promiscuity as well as the conformational diversity found in some enzymes allow them to evolve new activities rapidly [153].

IDRs appear to be highly represented in proteins of organisms evolving in extreme environments (i.e., prokaryotes [157]). Their role in response to abiotic stress have been mainly recorded in plants [158,159]. However, it has been shown that they perform important functions to protect the cell from thermal stress in other organisms, including cold conditions in bacteria [160] and high temperatures in animals (e.g., under desiccation conditions in tardigrades [161]). This is in agreement with early studies showing that flexibility is a key characteristic of proteins involved in thermal adaptation (reviewed in [162]).

On the other hand, there is a mounting evidence showing that similar to TRs, IDRs are essential for a diversity of cellular processes (see review in [163]), some of them key for adaptation to the environment. For instance, these regions facilitate macromolecular interactions by participating actively in the assembly of signaling complexes, interacting with structured domains in other proteins [84,163,164]. This is particularly relevant in mitochondrial proteins that interact closely with their nuclear partners in mito-nuclear interactions and have been found to be key players not only in adaptive responses but also in speciation [27,33,35].

Noticeably, the main processes predicted for biological function of the TMP326 of *RS\_LinB* (i.e., transport and cell surface receptor signaling pathway [ $P = 0.8$ ]) were similar to those processes found in disordered proteins of other metazoans, the IDRs of which have been associated with regulation of cellular processes, including intracellular signaling, membrane fusion and transport, and signal transduction [84,164]. However, these results should be taken with caution as a fully experimental characterization of this protein will be needed to unravel its function in the mitochondrial genomes of pocilloporids, particularly in species evolving in environments considered extreme for reef-building corals.

#### 4.3.2. Structural Diversification of TMP362

TMP362 presents three well-supported transmembrane domains in *Madracis* versus two in the other pocilloporids (but a third transmembrane domain may be present in *Seriatopora*; Supplementary Figure S1). Interestingly, differences in the number of TMDs between homologous proteins have been recorded in plants and others photosynthetic organisms. For example, changes in amino acid sequences leading to differences in secondary structures have been reported in two rice lines (i.e., Huhan1A and Huhan 1B) that differ in their tolerance to stress [165]. In an experimental study, authors found a reduction of TMDs in the proteins encoded by the mitochondrial genes *ccmB* and *ccmC* of Huhan1A and Huhan 1B, from 5 to 3 and 6 to 4 respectively, when analyzing gene expression under different environmental settings. These differences were interpreted as a consequence of deficiencies in RNA editing due to oxidative stress, which is also associated to other stressors such as drought, heat and cold [166].

At a broader evolutionary scale, changes in TMDs have been also reported in homologous proteins of the light-harvesting complex (LHC)-like family, which exhibits a broad functional diversity and

has acquired novel functions throughout the evolution of photosynthetic organisms [167,168]. During evolution, these proteins have gained and lost TMDs multiple times. The ancestral protein likely evolved from one to two TMDs and acquired two extra TMDs by duplication, with one domain loss (from 4 to 3) in the homologous protein of extant lineages (e.g., cyanobacteria and photosynthetic eukaryotes; reviewed in [169,170]). Remarkably, LHC homologs are transmembrane proteins involved in photo-protective responses, protecting against the photo-oxidative damage caused by excessive production of reactive oxygen species, particularly in periods of long exposure to sunlight [171–173]. They are also induced as a response to other environmental stressors such as increases in salinity and temperature, and during drastic decreases in nutrients availability (reviewed in [174]).

Retracing the more likely path for the evolutionary changes in the structure of TMP326 orthologs is out of the scope of the present work. However, based on the studies outlined here, a possible explanation could be that the structural differences among TMP326 orthologs evolved as a response to different adaptive pressures, possibly related to environmental changes. Studies in reef-building corals have shown that the response to environmental stressors, such as heat stress, may differ among taxa [175]. In some corals, however, this response is under a strong maternal influence and is characterized by the overexpression of set of mitochondrial proteins [176]. Such studies are still lacking for reef-building corals and will require experimental characterizations of mtDNA-encoded proteins involved in stress response to understand their significance in an evolutionary context.

#### 4.3.3. Signatures of Selection in the TMP362 of Stylophora Lineages along Latitudinal and Environmental Gradients

A third line of evidence in support of the hypothesis of a role of the TMP362 protein in adaptation is the signal of positive selection in sequences of specimens inhabiting the warmest regions of the Red Sea, particularly in sequences from *RS\_LinB* (Figures 6 and 7), versus an absence of positively selected sites in TMP362 of specimens restricted to the coldest northern areas. Noticeably, TMP362 of individuals from both lineages, distributed along the full environmental gradient of the Red Sea, showed multiples sites under purifying selection (Figures 6 and 7). This finding was supported by all methods (see results Section 3.7 and Figure 1 for reference to the geographic distribution of these lineages in the Red Sea).

It is broadly accepted that coral reef ecosystems accompanied the evolution of the Red Sea basin since its incipient stage in the early Miocene [177], and that pocilloporid corals likely entered the region as early as the Pliocene-Pleistocene [177–181]. During these periods they faced a variety of environmental stressors, such as extended periods of hyper-salinity, changes in nutrient availability, and strong fluctuations in oceanic temperatures and sea levels [50,182,183] that lead to the extinction of some lineages, while others survived in refugia in the northern and southern Red Sea (reviewed in DiBattista et al. [48]). These environmental settings might explain the signals of adaptive evolution as well as the unique characteristics of the mtORF-encoded protein of the endemic lineage *RS\_LinB*, which may have experienced more extreme fluctuations in environmental conditions than corals evolving under less stressful conditions.

On the other hand, the lack of positive selection in the mtORF-encoded protein of *RS\_LinA* contrasts with the multiple amino acid sites found to be under purifying selection in this lineage. This could be explained by its relatively recent appearance in the Red Sea, in agreement with its hybrid origin [15]. In fact, more recent *Stylophora* lineages likely entered the Red Sea within the last 4–7k years, when environmental conditions were similar to those recorded nowadays [184]. They overcame the bottleneck of high temperature waters in the southern Red Sea (up to 34 °C) and then spread north along with other coral species [185].

Overall, these findings are congruent with results based on protein structure and function, discussed in the previous sections, suggesting that the mtORF-encoded protein may indeed have a role in environmental adaptation, likely to high temperatures, highlighting also its functional importance.

However, as already mentioned, this hypothesis will require further testing using experimental data and a combination of proteomics, genomics and ecological approaches.

## 5. Conclusions

Our study not only offers insights into the coding nature of the mtORF gene in pocilloporid corals, but adds to our knowledge of the main characteristics of the mtORF-encoded protein. Striking dissimilarities were found among the TMP362 proteins of pocilloporid corals, with the greatest differences found among *Stylophora* lineages, including changes in structure. The most divergent protein was that of *RS\_LinB*, a lineage that was hypothesized to have adapted to the extreme environments of the Red Sea during an early colonization, likely predating that of other *Stylophora* lineages. Several features found in the TMP362 of the *RS\_LinB* (i.e., high frequency of TRs, long stretches of disordered residues, as well as signals of selection) show trends similar to proteins of organisms surviving under severe conditions and/or exposed to damage by oxidative stress, as most reef-building corals under thermal stress. Therefore, our findings suggest that *tmp362* and likely other genes involved in mitochondrial interactions have played a key role in the adaptation of *Stylophora* corals to extreme environments and fluctuating conditions.

This study, therefore, opens the door for further studies looking at the role of mitochondria and mitochondrial genes in the adaptation of coral species during their 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 interactions at the mitochondrial level of particular relevance, since they appear to play a key role in coral adaptation to strong environmental changes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/5/324/s1>, Figure S1: Transmembrane domains and pore-lining regions in pocilloporid corals; Table S1: Accession numbers of mtORF sequences analyzed in this study; Table S2: TRs motifs in *Stylophora* lineages and populations; Data S1: Full mitogenomes of *RS\_LinA* and *RS\_LinB*; Data S2: Translated mtORF sequences of Pocilloporidae corals.

**Author Contributions:** Conceptualization, E.B.-H., Y.S. and J.-F.F.; Formal analysis, E.B.-H., E.F.; Funding acquisition, J.-F.F.; Investigation, E.B.-H., Y.S., and J.-F.F.; Methodology, E.B.-H. and E.F.; Writing—original draft, E.B.-H.; Writing—review and editing, E.B.-H., Y.S., E.F., and J.-F.F.

**Funding:** 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) and the bilateral project "The Jeddah Transect" of the King Abdulaziz University (KAU) in Saudi Arabia, and the Helmholtz Center for Ocean Research (GEOMAR) in Germany (YS).

**Acknowledgments:** We thank Christian Voolstra and the Bioscience Core Lab at KAUST for sharing their facilities, Abdulmohsin Al-Sofyani at KAU for supporting coral sampling, and Patricia Warner for giving access to the chromatograms of *Seriatopora* species. Thanks also to Sandra Cervantes Arango, Dario Ojeda Alayon, Patrick Mardulyn, and Javier Fuertes Aguilar for useful discussions and advice. We are also grateful to the reviewers for their detailed checks of the manuscript and their useful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Szafranski, P. Evolutionarily recent, insertional fission of mitochondrial *cox2* into complementary genes in bilaterian Metazoa. *BMC Genomics* **2017**, *18*, 269. [[CrossRef](#)] [[PubMed](#)]
2. Endo, K.; Noguchi, Y.; Ueshima, R.; Jacobs, H.T. Novel repetitive structures, deviant protein-encoding sequences and unidentified ORFs in the mitochondrial genome of the brachiopod *Lingula anatina*. *J. Mol. Evol.* **2005**, *61*, 36–53. [[CrossRef](#)]
3. Breton, S.; Beaupré, H.D.; Stewart, D.T.; Piontkivska, H.; Karmakar, M.; Bogan, A.E.; Blier, P.U.; Hoeh, W.R. Comparative mitochondrial genomics of freshwater mussels (*Bivalvia*: unionoida) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames and putative origins of replication. *Genetics* **2009**, *183*, 1575–1589. [[CrossRef](#)]

4. Breton, S.; Milani, L.; Ghiselli, F.; Guerra, D.; Stewart, D.T.; Passamonti, M. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet.* **2014**, *30*, 555–564. [[CrossRef](#)] [[PubMed](#)]
5. Wu, X.; Li, X.; Li, L.; Xu, X.; Xia, J.; Yu, Z. New features of Asian *Crassostrea* oyster mitochondrial genomes: A novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* **2012**, *507*, 112–118. [[CrossRef](#)]
6. Kayal, E.; Bentlage, B.; Collins, A.G.; Kayal, M.; Pirro, S.; Lavrov, D.V. Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol. Evol.* **2012**, *4*, 1–12. [[CrossRef](#)]
7. Gissi, C.; Iannelli, F.; Pesole, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **2008**, *101*, 301–320. [[CrossRef](#)] [[PubMed](#)]
8. Higashi, A.; Nagai, S.; Salomon, P.S.; Ueki, S. A unique, highly variable mitochondrial gene with coding capacity of *Heterosigma akashiwo*, class Raphidophyceae. *J. Appl. Phycol.* **2017**, *29*, 2961–2969. [[CrossRef](#)]
9. Flot, J.-F.; Tillier, S. The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. *Gene* **2007**, *401*, 80–87. [[CrossRef](#)]
10. Flot, J.-F.; Magalon, H.; Cruaud, C.; Couloux, A.; Tillier, S. Patterns of genetic structure among Hawaiian corals of the genus *Pocillopora* yield clusters of individuals that are compatible with morphology. *C. R. Biol.* **2008**, *331*, 239–247. [[CrossRef](#)] [[PubMed](#)]
11. Johnston, E.C.; Forsman, Z.H.; Flot, J.-F.; Schmidt-Roach, S.; Pinzón, J.H.; Knapp, I.S.S.; Toonen, R.J. A genomic glance through the fog of plasticity and diversification in *Pocillopora*. *Sci. Rep.* **2017**, *7*, 5991. [[CrossRef](#)]
12. Flot, J.-F.; Licuanan, W.Y.; Nakano, Y.; Payri, C.; Cruaud, C.; Tillier, S. Mitochondrial sequences of *Seriatopora* corals show little agreement with morphology and reveal the duplication of a tRNA gene near the control region. *Coral Reefs* **2008**, *27*, 789–794. [[CrossRef](#)]
13. Warner, P.A.; Van Oppen, M.J.H.; Willis, B.L. Unexpected cryptic species diversity in the widespread coral *Seriatopora hystrix* masks spatial-genetic patterns of connectivity. *Mol. Ecol.* **2015**, *24*, 2993–3008. [[CrossRef](#)] [[PubMed](#)]
14. Nakajima, Y.; Nishikawa, A.; Iguchi, A.; Nagata, T.; Uyeno, D.; Sakai, K.; Mitarai, S. Elucidating the multiple genetic lineages and population genetic structure of the brooding coral *Seriatopora* (Scleractinia: Pocilloporidae) in the Ryukyu Archipelago. *Coral Reefs* **2017**, *36*, 415–426. [[CrossRef](#)]
15. Banguera-Hinestroza, E.; Sawall, Y.; Al-Sofyani, A.; Mardulyn, P.; Fuertes-Aguilar, J.; Cardenas-Henao, H.; Jimenez-Infante, F.; Voolstra, C.R.; Flot, J.-F. mtDNA recombination indicative of hybridization suggests a role of the mitogenome in the adaptation of reef building corals to extreme environments. *bioRxiv* **2018**, 462069.
16. Flot, J.-F.; Blanchot, J.; Charpy, L.; Cruaud, C.; Licuanan, W.Y.; Nakano, Y.; Payri, C.; Tillier, S. Incongruence between morphotypes and genetically delimited species in the coral genus *Stylophora*: phenotypic plasticity, morphological convergence, morphological stasis or interspecific hybridization? *BMC Ecol.* **2011**, *11*, 22. [[CrossRef](#)]
17. Shearer, T.L.; van Oppen, M.J.H.; Romano, S.L.; Wörheide, G. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Mol. Ecol.* **2002**, *11*, 2475–2487. [[CrossRef](#)] [[PubMed](#)]
18. Shearer, T.L.; Coffroth, M.A. DNA BARCODING: Barcoding corals: limited by interspecific divergence, not intraspecific variation. *Mol. Ecol. Resour.* **2008**, *8*, 247–255. [[CrossRef](#)] [[PubMed](#)]
19. Hill, G.E. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol. Evol.* **2016**, *6*, 5831–5842. [[CrossRef](#)] [[PubMed](#)]
20. Cheviron, Z.A.; Brumfield, R.T. Genomic insights into adaptation to high-altitude environments. *Heredity* **2012**, *108*, 354–361. [[CrossRef](#)]
21. Morales, H.E.; Pavlova, A.; Joseph, L.; Sunnucks, P. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol. Ecol.* **2015**, *24*, 2820–2837. [[CrossRef](#)]
22. Scott, G.R.; Schulte, P.M.; Egginton, S.; Scott, A.L.M.; Richards, J.G.; Milsom, W.K. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol. Biol. Evol.* **2011**, *28*, 351–363. [[CrossRef](#)]
23. Silva, G.; Lima, F.P.; Martel, P.; Castilho, R. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc. R. Soc. B Biol. Sci.* **2014**, *281*, 20141093. [[CrossRef](#)]
24. Saccone, C.; Lanave, C.; De Grassi, A. Metazoan OXPHOS gene families: evolutionary forces at the level of mitochondrial and nuclear genomes. *Biochim. Biophys. Acta* **2006**, *1757*, 1171–1178. [[CrossRef](#)]

25. Saraste, M. Oxidative phosphorylation at the *fin de siècle*. *Science*. **1999**, *283*, 1488–1493. [[CrossRef](#)] [[PubMed](#)]
26. Ben Slimen, H.; Schaschl, H.; Knauer, F.; Suchentrunck, F. Selection on the mitochondrial *ATP synthase 6* and the *NADH dehydrogenase 2* genes in hares (*Lepus capensis* L., 1758) from a steep ecological gradient in North Africa. *BMC Evol. Biol.* **2017**, *17*, 46. [[CrossRef](#)]
27. Gershoni, M.; Templeton, A.R.; Mishmar, D. Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays* **2009**, *31*, 642–650. [[CrossRef](#)]
28. Lajbner, Z.; Pnini, R.; Camus, M.F.; Miller, J.; Dowling, D.K. Experimental evidence that thermal selection shapes mitochondrial genome evolution. *Sci. Rep.* **2018**, *8*, 9500. [[CrossRef](#)] [[PubMed](#)]
29. Sunnucks, P.; Morales, H.E.; Lamb, A.M.; Pavlova, A.; Greening, C. Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Front. Genet.* **2017**, *8*, 25. [[CrossRef](#)] [[PubMed](#)]
30. Baris, T.Z.; Wagner, D.N.; Dayan, D.I.; Du, X.; Blier, P.U.; Pichaud, N.; Oleksiak, M.F.; Crawford, D.L. Evolved genetic and phenotypic differences due to mitochondrial-nuclear interactions. *PLoS Genet.* **2017**, *13*, e1006517. [[CrossRef](#)] [[PubMed](#)]
31. Zhao, N.; Korkin, D.; Finch, T.M.; Frederick, K.H.; Eggert, L.S. Evidence of positive selection in mitochondrial complexes I and V of the African elephant. *PLoS One* **2014**, *9*, e92587.
32. Rand, D.M.; Haney, R.A.; Fry, A.J. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* **2004**, *19*, 645–653. [[CrossRef](#)]
33. Hill, G.E. Mitonuclear ecology. *Mol. Biol. Evol.* **2015**, *32*, 1917–1927. [[CrossRef](#)]
34. Pavlova, A.; Amos, J.N.; Joseph, L.; Loynes, K.; Austin, J.J.; Keogh, J.S.; Stone, G.N.; Nicholls, J.A.; Sunnucks, P. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution* **2013**, *67*, 3412–3428. [[CrossRef](#)] [[PubMed](#)]
35. Hill, G.E. The mitonuclear compatibility species concept. *Auk* **2017**, *134*, 393–409. [[CrossRef](#)]
36. Mayer, M.P.; Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **2005**, *62*, 670–684. [[CrossRef](#)]
37. Sørensen, J.G.; Kristensen, T.N.; Loeschcke, V. The evolutionary and ecological role of heat shock proteins. *Ecol. Lett.* **2003**, *6*, 1025–1037. [[CrossRef](#)]
38. Kvitt, H.; Rosenfeld, H.; Tchernov, D. The regulation of thermal stress induced apoptosis in corals reveals high similarities in gene expression and function to higher animals. *Sci. Rep.* **2016**, *6*, 30359. [[CrossRef](#)]
39. Narum, S.R.; Campbell, N.R.; Meyer, K.A.; Miller, M.R.; Hardy, R.W. Thermal adaptation and acclimation of ectotherms from differing aquatic climates. *Mol. Ecol.* **2013**, *22*, 3090–3097. [[CrossRef](#)] [[PubMed](#)]
40. Horowitz, M. Heat acclimation: phenotypic plasticity and cues to the underlying molecular mechanisms. *J. Therm. Biol.* **2001**, *26*, 357–363. [[CrossRef](#)]
41. Veron, C.; Stafford-Smith, M.; Turak, E.; DeVantier, L. Corals of the World. Version 0.01 Beta. Available online: <http://www.coralsoftheworld.org/page/home/> (accessed on February 2019).
42. Kürten, B.; Al-Aidaros, A.M.; Struck, U.; Khomayis, H.S.; Gharbawi, W.Y.; Sommer, U. Influence of environmental gradients on C and N stable isotope ratios in coral reef biota of the Red Sea, Saudi Arabia. *J. Sea Res.* **2014**, *85*, 379–394. [[CrossRef](#)]
43. Osman, E.O.; Smith, D.J.; Ziegler, M.; Kürten, B.; Conrad, C.; El-Haddad, K.M.; Voolstra, C.R.; Suggett, D.J. Thermal refugia against coral bleaching throughout the northern Red Sea. *Glob. Chang. Biol.* **2018**, *24*, e474–e484. [[CrossRef](#)] [[PubMed](#)]
44. Sawall, Y.; Al-Sofyani, A.; Hohn, S.; Banguera-Hinestroza, E.; Voolstra, C.R.; Wahl, M. Extensive phenotypic plasticity of a Red Sea coral over a strong latitudinal temperature gradient suggests limited acclimatization potential to warming. *Sci. Rep.* **2015**, *5*, 8940. [[CrossRef](#)]
45. Rasul, N.M.A.; Stewart, I.C.F.; Nawab, Z.A. Introduction to the Red Sea: its origin, structure, and environment. In *The Red Sea: The formation, morphology, oceanography and environment of a young ocean basin*; Rasul, N.M.A., Stewart, I.C.F., Eds.; Springer: Heidelberg, Germany, 2015; pp. 1–28.
46. Bruckner, A.; Rowlands, G.; Riegl, B.; Purkis, S.; Williams, A.; Renaud, P.; Khaled Bin Sultan Living Oceans Foundation. *Atlas of Saudi Arabian Red Sea Marine Habitats*; Panoramic Press: Phoenix, AZ, USA, 2012; ISBN 9780983561118.
47. DiBattista, J.D.; Roberts, M.B.; Bouwmeester, J.; Bowen, B.W.; Coker, D.J.; Lozano-Cortés, D.F.; Howard Choat, J.; Gaither, M.R.; Hobbs, J.-P.A.; Khalil, M.T.; et al. A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. *J. Biogeogr.* **2016**, *43*, 423–439. [[CrossRef](#)]



48. DiBattista, J.D.; Howard Choat, J.; Gaither, M.R.; Hobbs, J.-P.A.; Lozano-Cortés, D.F.; Myers, R.F.; Paulay, G.; Rocha, L.A.; Toonen, R.J.; Westneat, M.W.; et al. On the origin of endemic species in the Red Sea. *J. Biogeogr.* **2016**, *43*, 13–30. [[CrossRef](#)]
49. Siddall, M.; Smeed, D.A.; Hemleben, C.; Rohling, E.J.; Schmelzer, I.; Peltier, W.R. Understanding the Red Sea response to sea level. *Earth Planet. Sci. Lett.* **2004**, *225*, 421–434. [[CrossRef](#)]
50. Siddall, M.; Rohling, E.J.; Almogi-Labin, A.; Hemleben, C.; Meischner, D.; Schmelzer, I.; Smeed, D. Sea-level fluctuations during the last glacial cycle. *Nature* **2003**, *423*, 853–858. [[CrossRef](#)]
51. Moustafa, M.Z.; Moustafa, M.S.; Moustafa, Z.D.; Moustafa, S.E. Survival of high latitude fringing corals in extreme temperatures: Red Sea oceanography. *J. Sea Res.* **2014**, *88*, 144–151. [[CrossRef](#)]
52. Arrigoni, R.; Benzoni, F.; Terraneo, T.I.; Caragnano, A.; Berumen, M.L. Recent origin and semi-permeable species boundaries in the scleractinian coral genus *Stylophora* from the Red Sea. *Sci. Rep.* **2016**, *6*, 34612. [[CrossRef](#)]
53. Stefani, F.; Benzoni, F.; Yang, S.Y.; Pichon, M.; Galli, P.; Chen, C.A. Comparison of morphological and genetic analyses reveals cryptic divergence and morphological plasticity in *Stylophora* (Cnidaria, Scleractinia). *Coral Reefs* **2011**, *30*, 1033–1049. [[CrossRef](#)]
54. Pohl, M.; Theißen, G.; Schuster, S. GC content dependency of open reading frame prediction via stop codon frequencies. *Gene* **2012**, *511*, 441–446. [[CrossRef](#)]
55. Tutar, Y. Pseudogenes. *Comp. Funct. Genomics* **2012**, *2012*, 1–4. [[CrossRef](#)]
56. Xiao, J.; Sekhwal, M.K.; Li, P.; Ragupathy, R.; Cloutier, S.; Wang, X.; You, F.M. Pseudogenes and their genome-wide prediction in plants. *Int. J. Mol. Sci.* **2016**, *17*, 1991. [[CrossRef](#)] [[PubMed](#)]
57. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)] [[PubMed](#)]
58. Sonnhammer, E.L.; von Heijne, G.; Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 175–182.
59. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)] [[PubMed](#)]
60. McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **2000**, *16*, 404–405. [[CrossRef](#)]
61. Nugent, T.; Jones, D.T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **2009**, *10*, 159. [[CrossRef](#)]
62. JEN, V.; Pichon, M. Scleractinia of eastern Australia. Part I: Families Thamnasteriidae, Astrocoeniidae, Pocilloporidae. *Aust. Gov. Publ. Serv.* **1976**, 208.
63. Harris, R.S. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University, University Park, PA, USA, 2007. [[CrossRef](#)]
64. Ankenbrand, M.J.; Hohlfeld, S.; Hackl, T.; Förster, F. AliTV — interactive visualization of whole genome comparisons. *PeerJ. Comput. Sci.* **2017**, *3*, 1–10. [[CrossRef](#)]
65. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)]
66. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469. [[CrossRef](#)] [[PubMed](#)]
67. Dereeper, A.; Audic, S.; Claverie, J.-M.M.; Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **2010**, *10*, 8. [[CrossRef](#)] [[PubMed](#)]
68. Chuang, Y.; Kitahara, M.; Fukami, H.; Tracey, D.; Miller, D.J.; Chen, C.A. Loss and gain of group I introns in the mitochondrial *cox1* gene of the Scleractinia (Cnidaria; Anthozoa). *Zool. Stud.* **2017**, *56*.
69. Anisimova, M.; Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **2006**, *55*, 539–552. [[CrossRef](#)]
70. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]

71. Robitzsch, V.; Banguera-Hinestroza, E.; Sawall, Y.; Al-Sofyani, A.; Voolstra, C.R. Absence of genetic differentiation in the coral *Pocillopora verrucosa* along environmental gradients of the Saudi Arabian Red Sea. *Front. Mar. Sci.* **2015**, *2*. [[CrossRef](#)]
72. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
73. Buchan, D.W.A.; Ward, S.M.; Lobley, A.E.; Nugent, T.C.O.; Bryson, K.; Jones, D.T. Protein annotation and modelling servers at University College London. *Nucleic Acids Res.* **2010**, *38*, W563–W568. [[CrossRef](#)] [[PubMed](#)]
74. Mulder, N.; Apweiler, R. InterPro and InterProScan: Tools for protein sequence classification and comparison. *Methods Mol Biol.* **2007**, *396*, 59–70.
75. Mitchell, A.L.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.-Y.; El-Gebali, S.; Fraser, M.I.; et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2019**, *47*, D351–D360. [[CrossRef](#)]
76. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195–202. [[CrossRef](#)] [[PubMed](#)]
77. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
78. Jones, D.T. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **2007**, *23*, 538–544. [[CrossRef](#)] [[PubMed](#)]
79. Jones, D.T.; Taylor, W.R.; Thornton, J.M. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **1994**, *33*, 3038–3049. [[CrossRef](#)]
80. Nielsen, H. Predicting secretory proteins with SignalP. *Methods Mol Biol.* **2017**, *1611*, 59–73.
81. Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
82. Jones, D.T.; Ward, J.J. Prediction of disordered regions in proteins from position specific score matrices. *Proteins Struct. Funct. Genet.* **2003**, *53*, 573–578. [[CrossRef](#)]
83. Ward, J.J.; McGuffin, L.J.; Bryson, K.; Buxton, B.F.; Jones, D.T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004**, *20*, 2138–2139. [[CrossRef](#)]
84. Van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.; et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114*, 6589–6631. [[CrossRef](#)]
85. Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [[CrossRef](#)] [[PubMed](#)]
86. Piovesan, D.; Tabaro, F.; Paladin, L.; Necci, M.; Mieti, I.; Camilloni, C.; Davey, N.; Dosztányi, Z.; Mészáros, B.; Monzon, A.M.; et al. MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* **2018**, *46*, D471–D476. [[CrossRef](#)] [[PubMed](#)]
87. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)] [[PubMed](#)]
88. Jaroszewski, L.; Li, Z.; Cai, X.-H.; Weber, C.; Godzik, A. FFAS server: novel features and applications. *Nucleic Acids Res.* **2011**, *39*, W38–W44. [[CrossRef](#)]
89. Finn, R.D.; Bateman, A.; Clements, J.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: the protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [[CrossRef](#)] [[PubMed](#)]
90. Pei, J.; Grishin, N.V. PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* **2007**, *23*, 802–808. [[CrossRef](#)]
91. Bhagwat, M.; Aravind, L. PSI-BLAST Tutorial. **2007**, *395*, 177–186.
92. Murzin, A.G.; Brenner, S.E.; Hubbard, T.; Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247*, 536–540. [[CrossRef](#)]
93. Wheeler, D.L.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvermin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2007**, *36*, D13–D21. [[CrossRef](#)]
94. Jaroszewski, L.; Rychlewski, L.; Li, Z.; Li, W.; Godzik, A. FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res.* **2005**, *33*, W284–W288. [[CrossRef](#)]

95. Minneci, F.; Piovesan, D.; Cozzetto, D.; Jones, D.T. FFPred 2.0: improved homology-independent prediction of gene ontology terms for eukaryotic protein sequences. *PLoS One* **2013**, *8*, e63754. [[CrossRef](#)]
96. Buchan, D.W.A.; Minneci, F.; Nugent, T.C.O.; Bryson, K.; Jones, D.T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **2013**, *41*, W349–W357. [[CrossRef](#)]
97. Lobley, A.; Sadowski, M.I.; Jones, D.T. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* **2009**, *25*, 1761–1767. [[CrossRef](#)]
98. Knudsen, M.; Wiuf, C. The CATH database. *Hum. Genomics* **2010**, *4*, 207. [[CrossRef](#)]
99. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
100. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [[CrossRef](#)]
101. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [[CrossRef](#)]
102. Sharp, P.M.; Li, W.-H. The codon adaptation index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [[CrossRef](#)]
103. Jia, W.; Higgs, P.G. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Mol. Biol. Evol.* **2008**, *25*, 339–351. [[CrossRef](#)]
104. Puigbò, P.; Bravo, I.G.; Garcia-Vallvé, S. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* **2008**, *9*, 65. [[CrossRef](#)]
105. Delport, W.; Poon, A.; Frost, S.D.W.; Kosakovsky Pond, S.L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **2010**, *26*, 2455–2457. [[CrossRef](#)]
106. Kosakovsky Pond, S.L.; Frost, S.D.W. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **2005**, *21*, 2531–2533. [[CrossRef](#)]
107. Kosakovsky Pond, S.L.; Frost, S.D.W. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [[CrossRef](#)]
108. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **2012**, *8*, e1002764. [[CrossRef](#)] [[PubMed](#)]
109. 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. *Mol. Biol. Evol.* **2011**, *28*, 3033–3043. [[CrossRef](#)]
110. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [[CrossRef](#)]
111. Yang, Z.; Bielawski, J.P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **2000**, *15*, 496–503. [[CrossRef](#)]
112. Doron-Faigenboim, A.; Pupko, T.A. Combined empirical and mechanistic codon model. *Mol. Biol. Evol.* **2006**, *24*, 388–397. [[CrossRef](#)]
113. Doron-Faigenboim, A.; Stern, A.; Mayrose, I.; Bacharach, E.; Pupko, T.A. Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* **2005**, *21*, 2101–2103. [[CrossRef](#)]
114. Stern, A.; Doron-Faigenboim, A.; Erez, E.; Martz, E.; Bacharach, E.; Pupko, T.A. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.* **2007**, *35*, W506–W511. [[CrossRef](#)]
115. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418–426.
116. Hughes, A.L.; Friedman, R. Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes. *Immunogenetics* **2008**, *60*, 495–506. [[CrossRef](#)]
117. Merkler, D.J.; Farrington, G.K.; Wedler, F.C. Protein thermostability. Correlations between calculated macroscopic parameters and growth temperature for closely related thermophilic and mesophilic bacilli. *Int. J. Pept. Protein Res.* **1981**, *18*, 430–442. [[CrossRef](#)]
118. Jones, D.T.; Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863. [[CrossRef](#)] [[PubMed](#)]
119. Aravind, L.; Koonin, E.V. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.* **1998**, *23*, 469–472. [[CrossRef](#)]

120. Galperin, M.Y.; Natale, D.A.; Aravind, L.; Koonin, E.V. A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol.* **1999**, *1*, 303–305. [[PubMed](#)]
121. Kubo, T.; Newton, K.J. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* **2008**, *8*, 5–14. [[CrossRef](#)]
122. Galtier, N.; Kitazaki, K.; Kubo, T.; Ballard, J.; Whitlock, M.; Davila, J.; Arrieta-Montiel, M.; Wamboldt, Y.; Cao, J.; Hagmann, J.; et al. The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biol.* **2011**, *9*, 61. [[CrossRef](#)] [[PubMed](#)]
123. Raitzos, D.E.; Hoteit, I.; Prihartato, P.K.; Chronis, T.; Triantafyllou, G.; Abualnaja, Y. Abrupt warming of the Red Sea. *Geophys. Res. Lett.* **2011**, *38*. [[CrossRef](#)]
124. Weeks, S.J.; Anthony, K.R.N.; Bakun, A.; Feldman, G.C.; Guldborg, O.H. Improved predictions of coral bleaching using seasonal baselines and higher spatial resolution. *Limnol. Oceanogr.* **2008**, *53*, 1369–1375. [[CrossRef](#)]
125. Muller-Parker, G.; D’Elia, C.F.; Cook, C.B. Interactions between corals and their symbiotic algae. In *Coral Reefs in The Anthropocene*; Birkeland, C., Ed.; Springer: Dordrecht, The Netherlands, 2015; pp. 99–116.
126. Bourne, D.G.; Morrow, K.M.; Webster, N.S. Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* **2016**, *70*, 317–340. [[CrossRef](#)] [[PubMed](#)]
127. Keshavmurthy, S.; Yang, S.-Y.; Alamaru, A.; Chuang, Y.-Y.; Pichon, M.; Obura, D.; Fontana, S.; De Palmas, S.; Stefani, F.; Benzoni, F.; et al. DNA barcoding reveals the coral “laboratory-rat”, *Stylophora pistillata* encompasses multiple identities. *Sci. Rep.* **2013**, *3*, 1520. [[CrossRef](#)] [[PubMed](#)]
128. Maddamsetti, R.; Johnson, D.T.; Spielman, S.J.; Petrie, K.L.; Marks, D.S.; Meyer, J.R. Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution* **2018**, *72*, 2234–2243. [[CrossRef](#)]
129. Lin, T.-Y. Simple sequence repeat variations expedite phage divergence: mechanisms of indels and gene mutations. *Mutat. Res. Mol. Mech. Mutagen.* **2016**, *789*, 48–56. [[CrossRef](#)] [[PubMed](#)]
130. Rocha, E.P.C.B.A. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res.* **2002**, *30*, 2031–2042. [[CrossRef](#)]
131. Moxon, R.; Bayliss, C.; Hood, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **2006**, *40*, 307–333. [[CrossRef](#)]
132. Zhou, K.; Aertsen, A.; Michiels, C.W. The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol. Rev.* **2014**, *38*, 119–141. [[CrossRef](#)] [[PubMed](#)]
133. Schaper, E.; Anisimova, M. The evolution and function of protein tandem repeats in plants. *New Phytol.* **2015**, *206*, 397–410. [[CrossRef](#)]
134. Pollak, Y.; Zelinger, E.; Raskina, O. Repetitive DNA in the architecture, repatterning, and diversification of the genome of *Aegilops speltoides* Tausch (Poaceae, Triticeae). *Front. Plant Sci.* **2018**, *9*, 9. [[CrossRef](#)] [[PubMed](#)]
135. Hu, Y.; Zhang, L.; He, S.; Huang, M.; Tan, J.; Zhao, L.; Yan, S.; Li, H.; Zhou, K.; Liang, Y.; et al. Cold stress selectively unsilences tandem repeats in heterochromatin associated with accumulation of H3K9ac. *Plant Cell Environ.* **2012**, *35*, 2130–2142. [[CrossRef](#)]
136. Droma, Y.; Hanaoka, M.; Basnyat, B.; Arjyal, A.; Neupane, P.; Pandit, A.; Sharma, D.; Ito, M.; Miwa, N.; Katsuyama, Y. Adaptation to high altitude in Sherpas: association with the insertion/deletion polymorphism in the angiotensin-converting enzyme gene. *Wilderness Environ. Med.* **2008**, *19*, 22–29. [[CrossRef](#)]
137. Ren, J.; Hou, Z.; Wang, H.; Sun, M.; Liu, X.; Liu, B.; Guo, X. Intraspecific variation in mitogenomes of five *Crassostrea* species provides insight into oyster diversification and speciation. *Mar. Biotechnol.* **2016**, *18*, 242–254. [[CrossRef](#)]
138. Schüler, A.; Bornberg-Bauer, E. Evolution of protein domain repeats in Metazoa. *Mol. Biol. Evol.* **2016**, *33*, 3170–3182. [[CrossRef](#)]
139. Sharma, M.; Pandey, G.K. Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* **2016**, *6*, 1218. [[CrossRef](#)] [[PubMed](#)]
140. Kashi, Y.; King, D.G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **2006**, *22*, 253–259. [[CrossRef](#)]
141. King, D.G.; Soller, M.; Kashi, Y. Evolutionary tuning knobs. *Endeavour* **1997**, *21*, 36–40. [[CrossRef](#)]
142. Kashi, Y.; King, D.G. Has simple sequence repeat mutability been selected to facilitate evolution? *Isr. J. Ecol. Evol.* **2007**, *52*, 331–342. [[CrossRef](#)]

143. Gemayel, R.; Vinces, M.D.; Legendre, M.; Verstrepen, K.J. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* **2010**, *44*, 445–477. [[CrossRef](#)] [[PubMed](#)]
144. Hull, R.M.; Cruz, C.; Jack, C.V.; Houseley, J. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* **2017**, *15*, e2001333. [[CrossRef](#)]
145. Biscotti, M.A.; Olmo, E.; Heslop-Harrison, J.S. Repetitive DNA in eukaryotic genomes. *Chromosom. Res.* **2015**, *23*, 415–420. [[CrossRef](#)]
146. Verstrepen, K.J.; Jansen, A.; Lewitter, F.; Fink, G.R. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **2005**, *37*, 986–990. [[CrossRef](#)]
147. Sonay, T.B.; Carvalho, T.; Robinson, M.D.; Greminger, M.P.; Krützen, M.; Comas, D.; Highnam, G.; Mittelman, D.; Sharp, A.; Marques-Bonet, T.; et al. Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. *Genome Res.* **2015**, *25*, 1591–1599. [[CrossRef](#)] [[PubMed](#)]
148. Chatterjee, N.; Lin, Y.; Santillan, B.A.; Yotnda, P.; Wilson, J.H. Environmental stress induces trinucleotide repeat mutagenesis in human cells. *Proc. Natl. Acad. Sci.* **2015**, *112*, 3764–3769. [[CrossRef](#)] [[PubMed](#)]
149. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509–516. [[CrossRef](#)]
150. Ahrens, J.B.; Nunez-Castilla, J.; Siltberg-Liberles, J. Evolution of intrinsic disorder in eukaryotic proteins. *Cell. Mol. Life Sci.* **2017**, *74*, 3163–3174. [[CrossRef](#)]
151. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* **2003**, *25*, 847–855. [[CrossRef](#)] [[PubMed](#)]
152. Light, S.; Sagit, R.; Sachenkova, O.; Ekman, D.; Elofsson, A. Protein expansion is primarily due to indels in intrinsically disordered regions. *Mol. Biol. Evol.* **2013**, *30*, 2645–2653. [[CrossRef](#)]
153. James, L.C.; Tawfik, D.S. Conformational diversity and protein evolution – A 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **2003**, *28*, 361–368. [[CrossRef](#)]
154. Tokuriki, N.; Tawfik, D.S. Protein dynamism and evolvability. *Science* **2009**, *324*, 203–207. [[CrossRef](#)] [[PubMed](#)]
155. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [[CrossRef](#)]
156. Kriwacki, R.W.; Hengst, L.; Tennant, L.; Reed, S.I.; Wright, P.E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci.* **1996**, *93*, 11504–11509. [[CrossRef](#)] [[PubMed](#)]
157. Vicedo, E.; Schlessinger, A.; Rost, B. Environmental pressure may change the composition protein disorder in prokaryotes. *PLoS One* **2015**, *10*, e0133990. [[CrossRef](#)] [[PubMed](#)]
158. Pietrosevoli, N.; García-Martín, J.A.; Solano, R.; Pazos, F. Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. *PLoS One* **2013**, *8*, e55524. [[CrossRef](#)] [[PubMed](#)]
159. Mahjoubi, H.; Ebel, C.; Hanin, M. Molecular and functional characterization of the durum wheat TdRLL1, a member of the conserved Poaceae RSS1-like family that exhibits features of intrinsically disordered proteins and confers stress tolerance in yeast. *Funct. Integr. Genomics* **2015**, *15*, 717–728. [[CrossRef](#)]
160. Tantos, A.; Friedrich, P.; Tompa, P. Cold stability of intrinsically disordered proteins. *FEBS Lett.* **2009**, *583*, 465–469. [[CrossRef](#)] [[PubMed](#)]
161. Boothby, T.C.; Tapia, H.; Brozena, A.H.; Piskiewicz, S.; Smith, A.E.; Giovannini, I.; Rebecchi, L.; Pielak, G.J.; Koshland, D.; Goldstein, B. Tardigrades use intrinsically disordered proteins to survive desiccation. *Mol. Cell* **2017**, *65*, 975–984.e5. [[CrossRef](#)]
162. Fields, P.A. Review: protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2001**, *129*, 417–431. [[CrossRef](#)]
163. Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **2014**, *83*, 553–584. [[CrossRef](#)]
164. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)] [[PubMed](#)]
165. Yan, S.; Yu, X.; Luo, L.; Luo, Z.; Liu, G.; Liu, Y.; Xia, H.; Xiong, J.; Tao, T. RNA editing responses to oxidative stress between a wild abortive type male-sterile line and its maintainer line. *Front. Plant Sci.* **2017**, *8*, 1–12.
166. Millar, A.H.; Whelan, J.; Soole, K.L.; Day, D.A. Organization and regulation of mitochondrial respiration in plants. *Annu. Rev. Plant Biol.* **2011**, *62*, 79–104. [[CrossRef](#)] [[PubMed](#)]

167. Green, R.R.; Pichersky, E. Hypothesis for the evolution of three-helix Chl *a/b* and Chl *a/c* light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosynth. Res.* **1994**, *39*, 149–162. [[CrossRef](#)] [[PubMed](#)]
168. Koziol, A.G.; Borza, T.; Ishida, K.-I.; Keeling, P.; Lee, R.W.; Durnford, D.G. Tracing the evolution of the light-harvesting antennae in chlorophyll *a/b*-containing organisms. *Plant Physiol.* **2007**, *143*, 1802–1816. [[CrossRef](#)]
169. Heddad, M.; Engelken, J.; Adamska, I. Light stress proteins in viruses, cyanobacteria and photosynthetic eukaryota. In *Photosynthesis. Advances in Photosynthesis and Respiration, vol 34*; Eaton-Rye, J., Tripathy, B., Sharkey, T., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 299–317.
170. Heddad, M.; Adamska, I. The evolution of light stress proteins in photosynthetic organisms. *Comp. Funct. Genomics* **2002**, *3*, 504–510. [[CrossRef](#)]
171. Büchel, C. Evolution and function of light harvesting proteins. *J. Plant Physiol.* **2015**, *172*, 62–75. [[CrossRef](#)] [[PubMed](#)]
172. Hoffman, G.E.; Sanchez-Puerta, M.V.; Delwiche, C.F. Evolution of light-harvesting complex proteins from Chl *c*-containing algae. *BMC Evol. Biol.* **2011**, *11*, 101. [[CrossRef](#)]
173. Pascal, A.A.; Liu, Z.; Broess, K.; van Oort, B.; van Amerongen, H.; Wang, C.; Horton, P.; Robert, B.; Chang, W.; Ruban, A. Molecular basis of photoprotection and control of photosynthetic light-harvesting. *Nature* **2005**, *436*, 134–137. [[CrossRef](#)]
174. Rochaix, J.-D.; Bassi, R. LHC-like proteins involved in stress responses and biogenesis/repair of the photosynthetic apparatus. *Biochem. J.* **2019**, *476*, 581–593. [[CrossRef](#)] [[PubMed](#)]
175. Birkeland, C. *Coral Reefs in the Anthropocene*; Springer: Dordrecht, The Netherlands, 2015.
176. Dixon, G.B.; Davies, S.W.; Aglyamova, G.V.; Meyer, E.; Bay, L.K.; Matz, M.V. Genomic determinants of coral heat tolerance across latitudes. *Science* **2015**, *348*, 1460–1462. [[CrossRef](#)]
177. Taviani, M. Post-Miocene reef faunas of the Red Sea: glacio-eustatic controls. In *Sedimentation and Tectonics in Rift Basins Red Sea—Gulf of Aden*; Springer: Dordrecht, The Netherlands, 1998; pp. 574–582.
178. Bruggemann, J.H.; Buffler, R.T.; Guillaume, M.M.; Walter, R.C.; Von Cosel, R.; Ghebretensae, B.N.; Berhe, S.M. Stratigraphy, palaeoenvironments and model for the deposition of the Abdur Reef Limestone: context for an important archaeological site from the last interglacial on the Red Sea coast of Eritrea. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2004**, *203*, 179–206. [[CrossRef](#)]
179. Casazza, L.R. Pleistocene reefs of the Egyptian Red Sea: environmental change and community persistence. *PeerJ* **2017**, *5*, e3504. [[CrossRef](#)]
180. Dullo, W.C. Facies, fossil record, and age of Pleistocene reefs from the Red Sea (Saudi Arabia). *Facies* **1990**, *22*, 1–45. [[CrossRef](#)]
181. El-Sorogy, A.S. Contributions to the Pleistocene coral reefs of the Red Sea Coast, Egypt. *Arab Gulf J. Sci. Res.* **2008**, *26*, 63–85.
182. Rohling, E.J.; Grant, K.; Hemleben, C.; Siddall, M.; Hoogakker, B.A.A.; Bolshaw, M.; Kucera, M. High rates of sea-level rise during the last interglacial period. *Nat. Geosci.* **2008**, *1*, 38–42. [[CrossRef](#)]
183. Taviani, M.; López Correa, M.; Zibrowius, H.; Montagna, P.; McCulloch, M.; Ligi, M. Last glacial deep-water corals from the Red Sea. *Bull. Mar. Sci.* **2008**, *81*, 361–370.
184. Trommer, G.; Siccha, M.; Rohling, E.J.; Grant, K.; van der Meer, M.T.J.; Schouten, S.; Hemleben, C.; Kucera, M. Millennial-scale variability in Red Sea circulation in response to Holocene insolation forcing. *Paleoceanography* **2010**, *25*, 25. [[CrossRef](#)]
185. Fine, M.; Gildor, H.; Genin, A. A coral reef refuge in the Red Sea. *Glob. Chang. Biol.* **2013**, *19*, 3640–3647. [[CrossRef](#)]

