



Article

# Phylogenomics Provides New Insights into Gains and Losses of Selenoproteins among Archaeplastida

Hongping Liang<sup>1,2,3,†</sup>, Tong Wei<sup>2,3,4,†</sup>, Yan Xu<sup>1,2,3,†</sup>, Linzhou Li<sup>2,3,5</sup>, Sunil Kumar Sahu<sup>2,3,4</sup>, Hongli Wang<sup>1,2,3</sup>, Haoyuan Li<sup>1,2</sup>, Xian Fu<sup>2,3</sup>, Gengyun Zhang<sup>2,4</sup>, Michael Melkonian<sup>6</sup>, Xin Liu<sup>2,3,4</sup>, Sibow Wang<sup>2,4,7,\*</sup> and Huan Liu<sup>2,4,7,\*</sup>

- <sup>1</sup> Beijing Genomics Institute (BGI) Education Center, University of Chinese Academy of Sciences, Beijing 100049, China; lianghongping@genomics.cn (H.L.); xuyan@genomics.cn (Y.X.); wanghongli1@genomics.cn (H.W.); lihaoyuan@genomics.cn (H.L.)
  - <sup>2</sup> Beijing Genomics Institute (BGI) Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China; weitong@genomics.cn (T.W.); lilinzhou@genomics.cn (L.L.); sunilkumarsahu@genomics.cn (S.K.S.); fuxian1@cngb.org (X.F.); zhanggengyun@genomics.cn (G.Z.); liuxin@genomics.cn (X.L.)
  - <sup>3</sup> China National Gene Bank, Institute of New Agricultural Resources, BGI-Shenzhen, Jinsha Road, Shenzhen 518120, China
  - <sup>4</sup> State Key Laboratory of Agricultural Genomics, Beijing Genomics Institute (BGI) Shenzhen, Shenzhen 518083, China
  - <sup>5</sup> School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China
  - <sup>6</sup> Botanical Institute, Cologne Biocenter, University of Cologne, D-50674 Cologne, Germany; michael.melkonian@uni-koeln.de
  - <sup>7</sup> Department of Biology, University of Copenhagen, DK-1165 Copenhagen, Denmark
- \* Correspondence: wangsibo1@genomics.cn (S.W.); liuhuan@genomics.cn (H.L.)  
† These authors contributed equally to this work.

Received: 5 May 2019; Accepted: 18 June 2019; Published: 20 June 2019



**Abstract:** Selenoproteins that contain selenocysteine (Sec) are found in all kingdoms of life. Although they constitute a small proportion of the proteome, selenoproteins play essential roles in many organisms. In photosynthetic eukaryotes, selenoproteins have been found in algae but are missing in land plants (embryophytes). In this study, we explored the evolutionary dynamics of Sec incorporation by conveying a genomic search for the Sec machinery and selenoproteins across Archaeplastida. We identified a complete Sec machinery and variable sizes of selenoproteomes in the main algal lineages. However, the entire Sec machinery was missing in the Bangiophyceae-Florideophyceae clade (BV) of Rhodoplantae (red algae) and only partial machinery was found in three species of Archaeplastida, indicating parallel loss of Sec incorporation in different groups of algae. Further analysis of genome and transcriptome data suggests that all major lineages of streptophyte algae display a complete Sec machinery, although the number of selenoproteins is low in this group, especially in subaerial taxa. We conclude that selenoproteins tend to be lost in Archaeplastida upon adaptation to a subaerial or acidic environment. The high number of redox-active selenoproteins found in some bloom-forming marine microalgae may be related to defense against viral infections. Some of the selenoproteins in these organisms may have been gained by horizontal gene transfer from bacteria.

**Keywords:** evolution; horizontal gene transfer; phylogenomics; selenoproteins; selenocysteine; Sec machinery

## 1. Introduction

Selenium (Se) is an essential trace element for human health and its deficiency leads to various diseases, such as Keshan and Kashin-Beck diseases, and affects the immune system and promotes cancer development [1,2]. An essential Se metabolism is present in many organisms, including bacteria, archaea, and eukaryotes [1,3,4]. However, higher concentrations of Se are toxic by functioning as a pro-oxidant, which affects the intracellular glutathione (GSH) pool leading to an enhanced level of Reactive oxygen species (ROS) accumulation [5,6]. Se is essential for growth and development of numerous algal species but not for terrestrial plants (embryophytes), although it accumulates in certain plant species and can serve as dietary sources for Se uptake [3,7–9].

Se is incorporated into nascent polypeptides in the form of selenocysteine (Sec), the 21st amino acid [10]. Se incorporation requires a specialized machinery and Sec insertion sequence (SECIS) elements present in selenoprotein mRNAs [11,12]. In eukaryotes, it consists of Sec synthesis and Sec incorporation. Sec synthesis starts with tRNA<sup>Sec</sup>, aminoacylated with serine, which is phosphorylated by O-phosphoserine-transfer tRNA<sup>Sec</sup> kinase (PSTK) and then catalyzed by Sec synthase (SecS) to produce selenocysteinyl-tRNA<sup>Sec</sup> from selenophosphate [10–13]. The Sec donor, selenophosphate, is generated from selenide by selenophosphate synthetase 2 (SPS2), which is often a selenoprotein itself [14,15]. During Sec incorporation, SECIS-binding protein 2 (SBP2) recognizes the SECIS elements in the 3'-untranslated region (3'-UTR) and recruits the Sec-specific elongation factor (eEFSec) that delivers selenocysteinyl-tRNA<sup>Sec</sup> to the ribosome at the in-frame Sec-coding UGA (amber codon) stop codon. Bacteria possess a similar machinery including *selB* (Sec-specific elongation factor), *selC* (tRNA<sup>Sec</sup>) and *selD* (selenophosphate synthase), except that Sec synthesis is catalyzed by a single bacterial Sec synthase, *SelA* [10].

Although selenoproteins constitute only a small fraction of the proteome in any living organism, they play important roles in redox regulation, antioxidation, and thyroid hormone activation in animals including humans [16]. Sec incorporation has been well documented in animals, bacteria, and archaea, while the largest selenoproteome was reported in algae. In the pelagophyte alga *Aureococcus anophagefferens*, 59 selenoproteins were identified in its genome, compared with 25 selenoproteins in humans [17,18]. The green alga *Chlamydomonas reinhardtii* has at least ten selenoproteins, whereas the picoplanktonic, marine green alga *Ostreococcus lucimarinus* harbors 20 selenoprotein genes in its genome [3,19]. Considering that Se is essential for growth in at least 33 algal species that belong to six phyla, Sec incorporation is thought to be universal in diverse algal lineages [20]. In a previous study, no selenoproteins were found in any land plants [7], suggesting a complete loss of Sec incorporation after streptophyte terrestrialization. Exploring the Sec machinery across the Archaeplastida, especially in algae, would provide insight into its evolutionary dynamics in this important lineage of photosynthetic eukaryotes. Here in this study, we searched 38 plant genomes, including 33 algal species that represent the major algal lineages, for the Sec machinery and selenoproteins.

## 2. Results

### 2.1. Sec Machinery in Algae

To cover the plant tree of life, we selected 33 genomes of algal species and five embryophyte species with a focus on Archaeplastida, the major group of photosynthetic eukaryotes with primary plastids (Supplementary Figure S1). The 33 algal species include one glaucophyte, six rhodophytes, 16 chlorophytes, and seven streptophyte algae. Another three species, the pelagophyte *A. anophagefferens*, the diatom *Thalassiosira pseudonana*, and the coccolithophorid *Emiliania huxleyi*, were also included to represent other distinct algal lineages (Supplementary Table S1A).

The Sec machinery was searched in 38 genome assemblies using Selenoprofiles [21] (See Methods). As shown in Figure 1, embryophytes lack the entire Sec machinery as previously reported (Figure 2a [7]). Interestingly, the Sec machinery is not intact in all tested algal species. Among 33 algae,





that did not predict SECIS elements in the downstream region, especially in Mamiellophyceae, e.g., *Bathycoccus prasinus*, which may be because of lineage-specific characteristics or incomplete assembly [28]. The presence of selenoproteins in each assembled genome agrees with the intactness of the Sec machinery. In the rhodophyte clade that lacks the machinery or in the algae that miss one of the components, none of the known selenoproteins and SECIS elements were found in their genomes (except for *C. paradoxa*, in which the unidentified SBP2 protein may be incompletely assembled or other proteins replace the function of SBP2).

The Sec machinery is absent in embryophytes including the liverwort *Marchantia polymorpha* and the moss *Physcomitrella patens* (Figure 1). The availability of genomes (or transcriptomes) of all major lineages of streptophyte algae, the phylogeny of which can now be regarded as basically resolved [29], allowed identification of the likely step in the evolution of streptophytes when the loss of the Sec machinery and of selenoproteins occurred. As a first attempt to address this question, we searched the transcriptomic data from the 1KP project (<http://www.onekp.com>) for the presence of the Sec machinery and selenoproteins (268 algal species, 70 species of non-vascular (liverworts, mosses, hornworts) plants, and 175 species of monilophytes, lycophytes, and conifers). The number of enzymes of the Sec machinery and the number of selenoproteins were computed for each group (Supplementary Table S2). The sec machinery was completely absent from hornworts with no Sec incorporation machinery enzyme and selenoproteins. In liverworts and mosses, only a few selenoproteins were detected (2 and 1 respectively), and only a few enzymes of the Sec machinery were randomly distributed (in no bryophyte species were more than two of the five components of the Sec machinery detected: PSTK and SecS were absent in hornworts and eEFsec and SPS were absent in mosses) (Supplementary Table S2). In vascular plants, the Sec machinery was absent in all transcriptomes of all plants and no selenoproteins were detected (Supplementary Table S2). In the sister group of embryophytes, the Zygnematophyceae, enzymes of the Sec machinery were more widely distributed compared to bryophytes (Supplementary Table S2). In Zygnematophyceae, none among the five genes of the Sec machinery was found in their transcriptomes (four of the five components of the Sec machinery were present in about one third of the 40 taxa). It might be a consequence of the fragmentary nature of transcriptomes (e.g., we could not detect a complete Sec machinery in the transcriptomes of "*Spirotaenia* sp." and *Mesotaenium endlicherianum*, although in both genomes the complete Sec machinery had been identified, Figure 1). Furthermore, selenoproteins were identified in only 15 of the 40 Zygnematophyceae and their number per species was low. Again, we did not detect selenoproteins in the transcriptomes of "*Spirotaenia* sp." and *M. endlicherianum*, although in their genomes a few genes encoding selenoproteins were identified (note that the number of selenoproteins, as well as components of the Sec machinery, is higher in "*Spirotaenia* sp." because of its recent genome triplication; Cheng et al. (unpublished observations)). In the other clades of the streptophyte algae (Coleochaetophyceae, Charophyceae, Klebsormidiophyceae and Mesostigmatophyceae) the situation is similar to that in Zygnematophyceae, a complete Sec machinery is present but the number of selenoproteins identified is low, especially in the subaerial taxa (two in *Klebsormidium nitens* and three in *Chlorokybus atmophyticus*), the only exception being the scaly flagellate *Mesostigma viride* with 9 identified selenoproteins (Table 1 and Supplementary Table S2).

**Table 1.** Number of enzymes involved in the Sec incorporation machinery and selenoproteins. The number of enzymes of the Sec incorporation machinery and selenoproteins are detected by Selenoprofiles across the sequenced algae, liverworts, mosses, hornworts and a part of lower embryophyte genomes and transcriptomes (from the 1 KP project).

1KP Group		Sec Machinery						Selenoproteins (Sec) & Homologues		
Group (513)	Clade/Order	Species Number	eEFSec	PSTK	SBP2	SecS	SPS	Sec	Cys	Other
Vascular (175)	Conifers	76	0	0	0	0	0	0	5256	2574
	Lycophytes	21	0	0	0	0	0	0	1473	678
	Eusporangiate	10	0	0	0	0	0	0	640	280
	Monilo-phytes	10	0	0	0	0	0	0	640	280
	Leptosporangiate Monilophytes	68	0	0	0	0	0	0	4999	2184
Non-Vascular (70)	Hornworts	6	0	0	0	0	0	0	245	120
	Mosses	39	0	1	5	17	0	1	3297	1485
	Liverworts	25	1	0	5	0	1	2	2184	1044
Algae (268)	Zygnematophyceae	40	51	7	27	34	25	15	2426	1298
	Coleochaetophyceae	4	3	2	1	1	1	1	217	120
	Charophyceae	2	1	2	1	1	2	4	103	65
	Klebsormidiophyceae	5	10	6	4	5	4	5	295	131
	Mesostigmatophyceae	4	4	2	4	4	1	12	199	124
	Chlorophyta	137	174	50	83	112	75	222	7518	4489
	Glaucochlorophyta	6	6	4	5	2	3	4	281	177
	Rhodophyta	35	6	4	4	9	5	5	1296	736
	Chromista (algae)	35	45	1	24	32	20	4	2100	1127

#### 2.4. Phylogenetic Analysis of the Enzymes Involved in the Sec Machinery

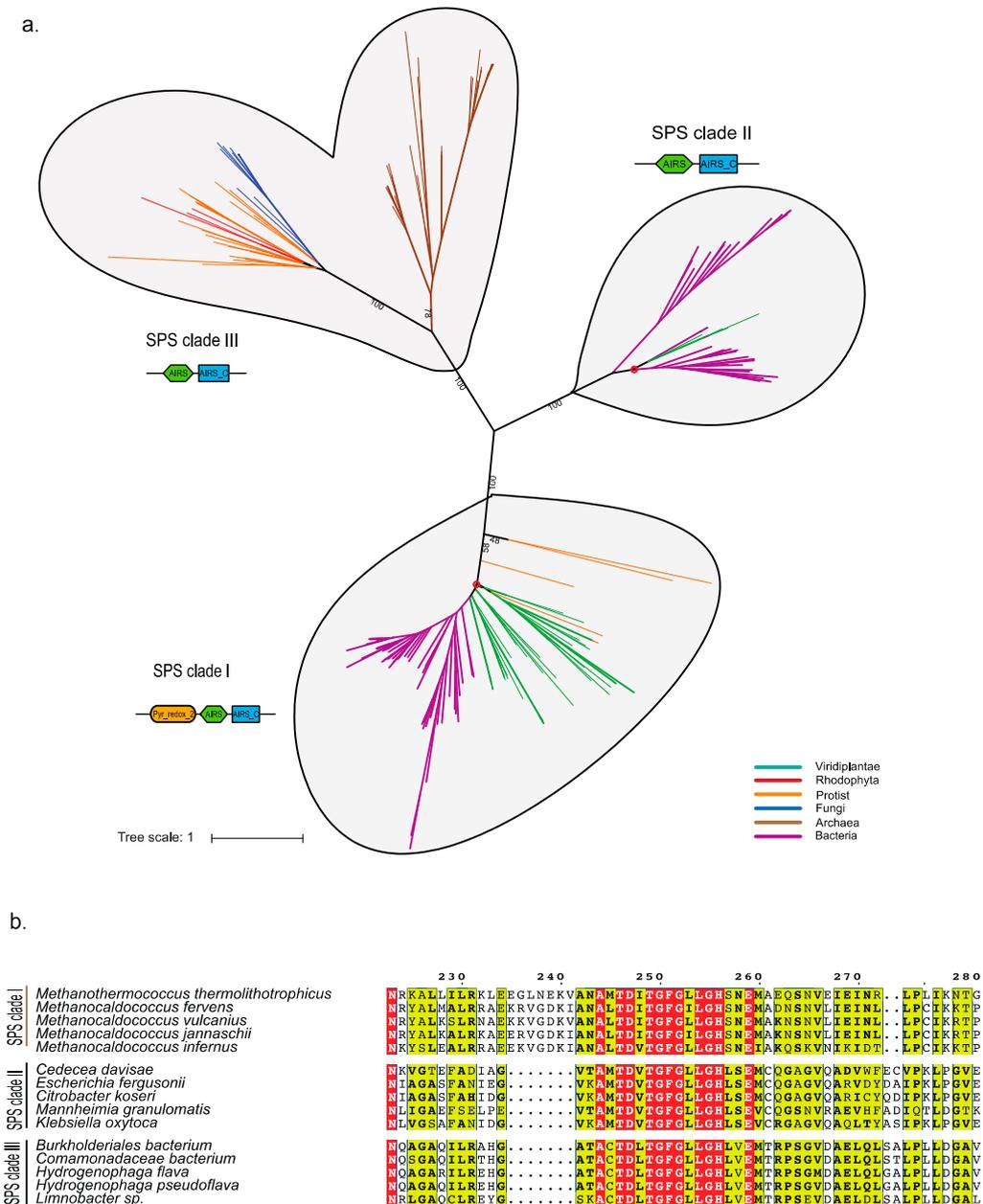
To further analyze the evolution of the Sec machinery, we conducted phylogenetic analyses of five genes encoding Sec-containing enzymes from the available Archaeplastida genome data set. The phylogenetic trees of PSTK, SBP2, and SPS showed either insufficient phylogenetic signal resulting in low support values for internal branches (PSTK) or very long branches in several taxa (SBP2, SPS) that led to spurious topologies due to long-branch attraction or indicated discordant gene histories (Supplementary Figure S3a–c).

The phylogenies of eEFsec and SecS were largely congruent with some support for internal branches (especially eEFsec) that roughly corresponded to the known phylogenetic relationships among higher order taxa, although relationships within some groups (e.g., streptophyte algae) remained unresolved (Figure 2b,c). The eEFsec phylogeny revealed four clades of sequences that were reasonably well supported: clade I comprised 3 sequences of Rhodophyta, clade II 6 sequences of picoplanktonic Mamiellophyceae, clade III 7 sequences of streptophyte algae, and clade IV 9 sequences from the TUC clade (3 sequences of Trebouxiophyceae and 6 sequences of Chlorophyceae).

#### Phylogenetic Analysis of Eukaryotic SPS Proteins

We built an SPS gene set comprising both prokaryotes and eukaryotes to reconstruct a global SPS phylogenetic tree (Figure 3, Supplementary Figure S4). SPS split into three well-separated clades: clade I including a diverse range of bacteria, most of the Viridiplantae, and protists with secondary plastids (Stramenopiles, cryptophytes, haptophytes and Apicomplexa), clade II containing bacteria and four species of green algae (*Chara braunii*; *Gonium pectorale*; *C. reinhardtii*; and *Volvox carteri*), and clade III including archaea, a diverse range of protists (photosynthetic and non-photosynthetic), fungi, and three rhodophytes but no other Archaeplastida (Supplementary Table S3). The sequence of SPS clade I contains three domains: Pyr\_redox\_2, AIRS and AIRS\_C. However, sequences of clade II and clade III only showed the presence of AIRS and AIRS\_C. The SPSs from clade II and clade III have different characteristics of domain arrangements (Supplementary Figure S4; as the phylogenetic tree suggested, potential horizontal gene transfer might have occurred in clades I and II.). The SPS of the three Volvocales (*C. reinhardtii*, *G. pectorale* and *V. carteri*) from clade II might have been acquired by horizontal gene transfer (HGT) from cyanobacteria, because they form a monophylum (92% bootstrap support) with two

terrestrial, filamentous cyanobacteria (*Tolypothrix bouiteillei*, *Scytonema hofmannii*) which are themselves nested within a larger radiation of bacteria (Supplementary Figure S4). For *C. braunii*, we suspect that this gene derived from either a (cyano) bacterial or volvoclean contamination.



**Figure 3.** Phylogenetic analysis of selenophosphate synthetase (SPS). (a) Reconstructed protein phylogeny of the reference set of SPS proteins. The red point denotes potential horizontal gene transfer events in SPS clades I and II. (b) Alignment of SPS domains of the three SPS clades.

### 2.5. Distribution of Types of Selenoproteins among Archaeplastida

A comprehensive analysis of the distribution of selenoproteins revealed that picoplanktonic Mamiellophyceae possess an expanded set of selenoproteins, whereas some selenoproteins had a scattered distribution among other Archaeplastida (Figure 2d, and Supplementary Figure S2). This may be related to the distinct types of eEFsec and SecS present in the Mamiellophyceae (Figure 2b,c). Functional annotation of the selenoproteins in the genomes of the Mamiellophyceae showed that they are mainly involved in oxidative stress response and adaptation. The MsrA selenoprotein, e.g.,

is a key Sec-containing enzyme for the repair of oxidatively damaged peptides. However, MsrA\_b, a bacterium-like MsrA selenoprotein, was identified only in the picoplanktonic Mamiellophyceae and in *M. viride* (Figure 2d), suggesting that early-diverging lineages of aquatic Viridiplantae might be subjected to stronger oxidative stress, and MsrA\_b but not MsrA (Supplementary Figure S2) is essential for these species to perform the repair of peptides. Another Sec-containing oxidoreductase (FrnE) is present in the Mamiellophyceae and in *M. viride* but not in any other Archaeplastida genome sequenced (Figure 2d). FrnE is a cadmium-inducible protein that is characterized as a disulfide isomerase having a role in oxidative stress tolerance. Therefore, it also supports the above hypothesis that Mamiellophyceae and *M. viride* (or perhaps scaly green algae, in general) need these enzymes to cope with stronger oxidative stress. In this context, it is interesting to note that in the bloom-forming pelagophyte alga *A. anophagefferens*, which has the second largest number of selenoproteins reported (50), a large number of redox active selenoproteins were overexpressed upon infection by a giant virus of the Mimiviridae clade [30], which suggests that viral infections, that are also prominent in the picoplanktonic Mamiellophyceae (prasinoviruses; [31]) and have also been described in *M. viride* [32], may elicit similar responses in their hosts. Viral infections are unknown in the three Volvocales studied (*C. reinhardtii*, *V. carteri*, and *G. pectorale*), however Volvocales are often subject to invasion by parasitic protists or fungi [33–35] and this could perhaps explain the presence of selenoproteins in these taxa.

### 3. Discussion

#### 3.1. The Distribution of the Sec Machinery and Selenoproteins in Algae

It has been hypothesized that the Sec machinery and selenoproteins were lost in Viridiplantae upon transfer from an aquatic to a terrestrial environment perhaps related to the paucity of a suitable chemical species of selenium (i.e., selenite) in most terrestrial environments [7,9,20,36–39]. The results presented here support this notion and further suggest that the Sec machinery was lost in the common ancestor of embryophytes as all extant embryophytes lack this machinery in their genomes (Figure 1). The few enzymes of this machinery that were detected in the transcriptomes of some liverworts and mosses (Supplementary Table S2) likely represent contaminations. Interestingly, although the complete Sec machinery is still present in all classes of streptophyte algae, the number of selenoproteins detected in the subaerial species (*C. atrophyticus*, *K. nitens*, “*Spirotaenia* sp.”, *M. endlicherianum*) was low (1–3 proteins), whereas in the aquatic species (*M. viride*, *C. braunii*, *C. scutata*) more selenoproteins (4–9 proteins) were found (Figure 1). Very low numbers of selenoproteins (i.e., one protein) were also encountered in subaerial/acidophilic species of Chlorophyceae (*C. eustigma*, *C. subellipsoidea*) and in the subaerial/acidophilic Rhodoplantae (*G. sulphuraria*). These results corroborate the hypothesis that adaptation to subaerial/terrestrial or acidophilic habitats supports the gradual loss of selenoproteins in diverse groups of algae. We suspect that once selenoproteins have been lost, selection on maintaining the Sec machinery is abolished. Intermediate stages in this process may be seen in the subaerial chlorophyte *M. neglectum* (now *M. braunii*) and in the acidophilic red alga *C. merolae* [36], which each lost one enzyme (PSTK or SBP respectively) of the Sec machinery. We hypothesize that once the Sec machinery is lost, transfer of algae to aquatic (marine) habitats (as in most species of Rhodoplantae) will not lead to reappearance of selenoproteins (some red algae exposed to strong oxidative stress such as *P. umbilicalis* have developed intimate associations with bacteria that express selenoproteins [40,41]). Similarly, transcriptomes of later-diverging Zygnematophyceae (i.e., Desmidiaceae), that are predominantly aquatic in mostly acidic environments (bogs), also either lack selenoproteins or have only 1 or 2 selenoprotein(s) (Supplementary Table S2). It will be interesting to learn, once their genome sequences will become available, whether they display a Sec machinery or not. Palenik et al. [19] proposed a trade-off between increased Se requirements but decreased nitrogen requirements for peptide synthesis in *Ostreococcus* spp., and it is worth noting that this genus encodes a surprisingly high number of selenocysteine-containing proteins relative to its genome size [19]. The core Chlorophyta showed a similar number of genes involved in nitrogen metabolism as the

picoplanktonic Mamiellophyceae (Supplementary Table S4). In Trebouxiophyceae and Ulvophyceae (represented by *Ulva mutabilis*), fewer selenoproteins were identified than in the Mamiellophyceae. Functional annotation of the selenoproteins in Trebouxiophyceae and Ulvophyceae showed that they mainly participated in some redox activities such as redox signaling (thioredoxin reductase, TR) and oxidative stress response (glutathione peroxidase, GPx) (Supplementary Figure S2). However, it is still unclear why Trebouxiophyceae and Ulvophyceae possess fewer selenoproteins, the first occur in freshwater or are often subaerial, the latter is mostly multicellular and may not require the diversity of highly reactive selenoenzymes characteristic for picoeukaryotes.

### 3.2. Probable Horizontal Gene Transfer of SPS and some Selenoproteins

SPS was detected in both prokaryotes and eukaryotes, although their sequence similarity is quite low (~30%; [4]). Our phylogenetic analyses resolved three clades of SPS genes with mixed species composition of prokaryotes and eukaryotes suggesting HGT among these unrelated organisms. For SPS clade II, we provided evidence that a single HGT event occurred from terrestrial cyanobacteria into the common ancestor of *C. reinhardtii*, *V. carteri*, and *G. pectorale*. Several selenoproteins of the picoplanktonic Mamiellophyceae may also have had their origin in the domain bacteria and been recruited from bacteria (perhaps via viruses) through HGT. Selenoproteins are relatively common in bacteria, about 34% of the sequenced bacteria utilize Sec, mostly different groups of proteobacteria (Figure 3b, Supplementary Figure S5 [38,40]). Phylogenetic analyses of selenoproteomes in bacteria have identified rampant losses of selenoproteins but also occasional HGT events, even between domains (bacteria and archaea) [42,43]. It is tempting to speculate that these HGTs supported bloom-forming, marine microalgae that often lack cell walls, their cells being covered only by mineralized or non-mineralized scales, to cope with viral invasions using their highly redox-reactive selenoproteins.

## 4. Materials and Methods

### 4.1. Data Information

A total of 38 genome sequences were used in this study, the genomes including 5 embryophytes, 7 streptophyte algae, 16 chlorophytes, 6 Rhodoplantae, 1 Glaucoplant and 3 photosynthetic protists (two stramenopiles and a haptophyte). The transcriptomes contained 121 green algae, 25 liverworts, 6 hornworts, 38 mosses, and 170 terrestrial plants (Supplementary Table S1A). The 33 whole genome assemblies were downloaded from the NCBI genome database. In addition, 5 newly assembled streptophyte algal genomes were used, including *Mesotaenium endlicherianum* (strain CCAC 1140), "*Spirotaenia* sp." (strain CCAC 0220), *Coleochaete scutata* (strain SAG 110.80), *Mesostigma viride* (strain CCAC 1140), *Chlorokybus atmophyticus* (strain CCAC 0220). The CCAC strains were obtained from the Culture Collection of Algae at the University of Cologne (<http://www.ccac.uni-koeln.de/>). All cultures were axenic, and during all steps of culture scale-up until nucleic acid extraction, axenicity was monitored by sterility tests as well as light microscopy. Total RNA was extracted from *M. viride* using the Tri Reagent Method, and from *C. atmophyticus* using the CTAB-PVP Method as described in Johnso [44]. Total DNA was extracted using a modified CTAB protocol [45,46]. The phylogenetic backbone of algae was retrieved from the NCBI taxonomy database (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). The completeness of genome assemblies was assessed by BUSCO 3.0.2 with eukaryote gene database [47]. The results were listed in the Supplementary Table S1B. We also counted the usage of stop codons for the single-copy genes. The results were shown in Supplementary Figure S1 (Supplementary Table S1B).

### 4.2. Sec Incorporation Machinery

The genome sequences were searched for the Sec incorporation machinery by the Selenoprofiles pipeline (version 3.0, <http://big.crg.cat/services/selenoprofiles>) with the parameter "-p machinery" [21,48]. Firstly, we ran the pipeline with profile-based Sec machinery. To reduce the

incomplete gene sequence mistakes, the blastp version 2.6.0+ (e-value  $< 10^{-5}$ ) was used against the predicted genes as in a special algae database to detect Sec machinery. In addition, transcriptome data were also searched using the same methods. First, the nucleic acid sequences were searched by Selenoprofiles, and then subjected to blastp (e-value  $< 10^{-5}$ ) with the predicted algae-specific Sec machinery database.

#### 4.3. Identification of the Selenocysteine tRNA ( $tRNA^{Sec}$ )

Secmarker version 0.4 (<http://secmarker.crg.es/index.html>) was used to identify the dedicated  $tRNA^{Sec}$  in the genome sequences [49]. The predicted secondary structure was drawn with the parameter “-plot”.

#### 4.4. Prediction of Selenoproteins and SECIS Elements

Selenoproteins were identified from the genome assemblies with Selenoprofiles with the parameter “-p metazoa, protist, prokarya”. The candidates were filtered with cutoff: e-value  $< 0.01$  and the sensible AWSic Z-score  $> -3$ . SECIS elements were searched in the 6-kb DNA sequences downstream of predicted selenoprotein genes at the SECISearch3 website (<http://seblastian.crg.es/>; with the parameter “-output\_three\_prime, -output\_secis”) [28].

#### 4.5. Phylogenetic Tree Construction

In phylogenetic analysis, each candidate was searched by Selenoprofiles and blastp version 2.6.0+ [44] to detect more candidates (e-value  $< 1 \times 10^{-5}$ ). Multiple sequence alignments were performed by MAFFT version 7.310 [50,51]. In eEFSec, SecS, PSTK, and SBP2, the maximum-likelihood tree was constructed for each protein family using the IQ-TREE software with 500 bootstrap replicates [52]. The SPS maximum-likelihood trees were constructed for each protein family using the RAxML version 8.2.4 with the GTR+I+G model [53,54]. For the phylogeny of SPS (SelD), the bacteria sequences were downloaded from the non-redundant (NR) database by submitting every alga SPS sequences to nr databases. All target bacterial sequences were retrieved but only several randomly chosen sequences in each bacterial phylum were used for the SPS phylogenetic analyses. Representative archaea and protist sequences were used in the analysis of SPS. In addition to this, the lately reported 9 fungi that utilize Sec were also added (192 sequences) [4,50].

#### 4.6. Identification of Conserved Motifs and Domains.

Pfam 32.0 (<http://pfam.xfam.org/>) was used to identify the domains in the Sec incorporation machinery [55]. Additional motifs were identified by Multiple Em for Motif Elicitation 5.0.5 (MEME, <http://meme-suite.org/>). The alignment of the SPS domain was visualized by ESPript 3.0.

## 5. Conclusions

A phylogenomic analysis of the selenocysteine (Sec) machinery and selenoproteins in genomes and transcriptomes of diverse Archaeplastida provided evidence for complete or partial loss of the Sec machinery in several, unrelated lineages accompanied by loss of selenoproteins. In streptophytes, the Sec machinery and selenoproteins were apparently lost in the common ancestor of embryophytes, as the Sec machinery was present in all lineages of streptophyte algae but absent in embryophytes. The number of selenoproteins identified in algae correlated with the type of their habitats, low numbers of selenoproteins were encountered in algae thriving in subaerial/terrestrial or acidic environments. The large number of selenoproteins found in some bloom-forming, marine microalgae may be related to their function in the defense against viral infections. Some components of the Sec machinery and selenoproteins may have been acquired by algae through horizontal gene transfer from bacteria.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/12/3020/s1>. The sequences of selenoprotein which we identified from the green algae (*Mesostigma viride*, *Chlorokybus*

*atmophyticus*, *Klebsormidium nitens*, *Chara braunii*, *Coleochaete scutata*, “*Spirotaenia* sp.”, *Mesotaenium endlicherianum*) are available in the CNGB Nucleotide Sequence Archive (CNSA: <http://db.cngb.org/cnsa>; accession number CNP0000452). The specific details regarding other genes which were used in this study are available in supplementary File S5.

**Author Contributions:** Data curation, Y.X. and L.L.; Formal analysis, H.L., H.W. and H.L.; Funding acquisition, X.L. and H.L.; Investigation, H.L.; Methodology, L.L., G.Z. and S.W.; Project administration, T.W. and H.L.; Resources, M.M.; Software, Y.X. and L.L.; Supervision, S.K.S., X.L., S.W. and H.L.; Visualization, H.L.; Writing—original draft, S.K.S. and S.W.; Writing—review & editing, S.K.S., X.F. and M.M.

**Funding:** Financial support was provided by National Key Research and Development Program of China (No.2017YFB0403904) and the Shenzhen Municipal Government of China (Grant numbers No. JCYJ20151015162041454 and No. JCYJ20160331150739027).

**Acknowledgments:** We thank Shifeng Cheng for kindly providing the gene sequences of *Spirotaenia* sp. and *Mesotaenium endlicherianum*.

**Conflicts of Interest:** The authors declare no competing interests.

## Abbreviations

BV	Bangiophyceae Florideophyceae
ROS	Reactive Oxygen Species
Sec	Selenocysteine
Se	Selenium
SECIS	Selenocysteine Insertion Sequence
PSTK	O-phosphoseryl-transfer tRNA <sup>Sec</sup> kinase
SecS	Sec Synthase
SPS	Selenophosphate Synthetase 2
SBP2	SECIS-binding Protein 2
eEFSec	Sec-specific Elongation Factor
CTAB	Cetyl Trimethylammonium Bromide

## References

1. Rayman, M.P. Selenium and human health. *Lancet* **2012**, *379*, 1256–1268. [[CrossRef](#)]
2. Avery, J.C.; Hoffmann, P.R. Selenium, Selenoproteins, and Immunity. *Nutrients* **2018**, *10*, 1203. [[CrossRef](#)] [[PubMed](#)]
3. Novoselov, S.V.; Rao, M.; Onoshko, N.V.; Zhi, H.; Kryukov, G.V.; Xiang, Y.; Weeks, D.P.; Hatfield, D.L.; Gladyshev, V.N. Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*. *EMBO J.* **2002**, *21*, 3681–3693. [[CrossRef](#)] [[PubMed](#)]
4. Mariotti, M.; Salinas, G.; Gabaldon, T.; Gladyshev, V.N. Utilization of selenocysteine in early-branching fungal phyla. *Nat. Microbiol.* **2019**, *4*, 759–765. [[CrossRef](#)] [[PubMed](#)]
5. Araie, H.; Suzuki, I.; Shiraiwa, Y. Identification and characterization of a selenoprotein, thioredoxin reductase, in a unicellular marine haptophyte alga, *Emiliana huxleyi*. *J. Biol. Chem.* **2008**, *283*, 35329–35336. [[CrossRef](#)]
6. Papp, L.V.; Holmgren, A.; Khanna, K.K. Selenium and selenoproteins in health and disease. *Antioxid. Redox Signal* **2010**, *12*, 793–795. [[CrossRef](#)]
7. Lobanov, A.V.; Fomenko, D.E.; Zhang, Y.; Sengupta, A.; Hatfield, D.L.; Gladyshev, V.N. Evolutionary dynamics of eukaryotic selenoproteomes: Large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol.* **2007**, *8*, 1–16. [[CrossRef](#)]
8. Bulteau, A.L.; Chavatte, L. Update on selenoprotein biosynthesis. *Antioxid. Redox Signal.* **2015**, *23*, 775–794. [[CrossRef](#)]
9. Schiavon, M.; Pilon-Smits, E.A. The fascinating facets of plant selenium accumulation—Biochemistry, physiology, evolution and ecology. *New Phytol.* **2017**, *213*, 1582–1596. [[CrossRef](#)]
10. Böck, A.; Forchhammer, K.; Heider, J.; Barion, C. Selenoprotein synthesis: An expansion of the genetic code. *Trends Biochem. Sci.* **1991**, *16*, 463–467. [[CrossRef](#)]
11. Berry, M.J.; Banu, L.; Chen, Y.; Mandel, S.J.; Kieffer, J.D.; Harney, J.W.; Larsen, P.R. Recognition of UGA as a selenocysteine codon in Type I deiodinase requires sequences in the 3′ untranslated region. *Nature* **1991**, *353*, 273–276. [[CrossRef](#)]

12. Low, S.C.; Berry, M.J. Knowing when not to stop: Selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.* **1996**, *21*, 203–208. [[CrossRef](#)]
13. Carlson, B.A.; Xu, X.M.; Kryukov, G.V.; Rao, M.; Berry, M.J.; Gladyshev, V.N.; Hatfield, D.L. Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12848–12853. [[CrossRef](#)] [[PubMed](#)]
14. Xu, X.M.; Carlson, B.A.; Irons, R.; Mix, H.; Zhong, N.; Gladyshev, V.N.; Hatfield, D.L. Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *Biochem. J.* **2007**, *404*, 115–120. [[CrossRef](#)] [[PubMed](#)]
15. Fletcher, J.E.; Copeland, P.R.; Driscoll, D.M.; Krol, A. The selenocysteine incorporation machinery: Interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* **2001**, *7*, 1442–1453. [[CrossRef](#)] [[PubMed](#)]
16. Labunskyy, V.M.; Hatfield, D.L.; Gladyshev, V.N. Selenoproteins: Molecular pathways and physiological roles. *Physiol. Rev.* **2014**, *94*, 739–777. [[CrossRef](#)]
17. Gobler, C.J.; Lobanov, A.V.; Tang, Y.Z.; Turanov, A.A.; Zhang, Y.; Doblin, M.; Taylor, G.T.; Sanudo-Wilhelmy, S.A.; Grigoriev, I.V.; Gladyshev, V.N. The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, *Aureococcus anophagefferens*. *ISME J.* **2013**, *7*, 1333–1343. [[CrossRef](#)]
18. Kryukov, G.V.; Castellano, S.; Novoselov, S.V.; Lobanov, A.V.; Zehtab, O.; Guigo, R.; Gladyshev, V.N. Characterization of mammalian selenoproteomes. *Science* **2003**, *300*, 1439–1443. [[CrossRef](#)]
19. Palenik, B.; Grimwood, J.; Aerts, A.; Salamov, A.; Putnam, N.H.; Dupont, C.L.; Jorgensen, R.A.; Rombauts, S.; Zhou, K.; Otilar, R.; et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7705–7710. [[CrossRef](#)]
20. Araie, H.; Shiraiwa, Y. Selenium utilization strategy by microalgae. *Molecules* **2009**, *14*, 4880–4891. [[CrossRef](#)]
21. Mariotti, M.; Guigo, R. Selenoprofiles: Profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *BMC Bioinform.* **2010**, *26*, 2656–2663. [[CrossRef](#)] [[PubMed](#)]
22. Yoon, H.S.; Nelson, W.; Linstrom, S.C.; Boo, S.M.; Poeschel, C.; Qiu, H.; Bhattacharya, D. Rhodophyta. In *Handbook of the Protists*; Archibald, J.M., Simpson, A.G.B., Slamovits, C.H., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 367–406.
23. Parte, S.; Sirisha, V.L.; D'Souza, J.S. Biotechnological applications of marine enzymes from algae, bacteria, fungi, and sponges. *Adv. Food Nutr. Res.* **2017**, *80*, 75–106. [[CrossRef](#)] [[PubMed](#)]
24. Qiu, H.; Yoon, H.S.; Bhattacharya, D. Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. *PLoS Curr.* **2016**, *8*. [[CrossRef](#)]
25. Price, N.M.; Harrison, P.J. Specific selenium-containing macromolecules in the marine diatom *Thalassiosira pseudonana*. *Plant Physiol.* **1988**, *86*, 192–199. [[CrossRef](#)] [[PubMed](#)]
26. Marin, B. Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. *Protist* **2012**, *163*, 778–805. [[CrossRef](#)] [[PubMed](#)]
27. Acton, E. *Coccomyxa subellipsoidea*, a new member of the palmellaceae. *Annals Bot.* **1909**, *23*, 573–577. [[CrossRef](#)]
28. Mariotti, M.; Lobanov, A.V.; Guigo, R.; Gladyshev, V.N. SECISearch3 and Seblastian: New tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res.* **2013**, *41*, e149. [[CrossRef](#)]
29. Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Carpenter, E.; Matasci, N.; Ayyampalayam, S.; Barker, M.S.; Burleigh, J.G.; Gitzendanner, M.A.; et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4859–E4868. [[CrossRef](#)]
30. Moniruzzaman, M.; Gann, E.R.; Wilhelm, S.W. Infection by a giant virus (AaV) induces widespread physiological reprogramming in *Aureococcus anophagefferens* CCMP1984—A harmful bloom algae. *Front Microbiol.* **2018**, *9*, 752. [[CrossRef](#)]
31. Weynberg, K.D.; Allen, M.J.; Wilson, W.H. Marine prasinoviruses and their tiny plankton hosts. *Viruses* **2017**, *9*, 43. [[CrossRef](#)]
32. Melkonian, M. Virus-like particles in the scaly green flagellate *Mesostigma viride*. *Br. Phycol. J.* **1982**, *17*, 63–68. [[CrossRef](#)]
33. Surek, B.; Melkonian, M. The filose amoeba *Vampyrellidium perforans* nov. sp. (Vampyrellidae, Aconchulinida): Axenic culture, feeding behaviour and host range specificity. *Arch. Protistenkd.* **1980**, *123*, 166–191. [[CrossRef](#)]
34. Hess, S. Hunting for agile prey: Trophic specialisation in leptophryid amoebae (Vampyrellida, Rhizaria) revealed by two novel predators of planktonic algae. *FEMS Microbiol. Ecol.* **2017**, *93*. [[CrossRef](#)] [[PubMed](#)]

35. Seto, K.; Degawa, Y. *Collimyces mutans* gen. et sp. nov. (Rhizophydiales, Collimycetaceae fam. nov.), a new chytrid parasite of *Microglena* (Volvocales, clade Monadinia). *Protist* **2018**, *169*, 507–520. [[CrossRef](#)]
36. Matsuzaki, M.; Misumi, O.; Shin-I, T.; Ma ruyama, S.; Takahara, M.; Miyagishima, S.Y.; Mori, T.; Nishida, K.; Yagisawa, F.; Nishida, K.; et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae*. *Nature* **2004**, *428*, 653–657. [[CrossRef](#)]
37. Schiavon, M.; Ertani, A.; Parrasia, S.; Vecchia, F.D. Selenium accumulation and metabolism in algae. *Aquat. Toxicol.* **2017**, *189*, 1–8. [[CrossRef](#)] [[PubMed](#)]
38. Gojkovic, Ž.; Garbayo, I.; Ariza, J.L.G.; Márová, I.; Vílchez, C. Selenium bioaccumulation and toxicity in cultures of green microalgae. *Algal Res.* **2015**, *7*, 106–116. [[CrossRef](#)]
39. Kim, J.W.; Brawley, S.H.; Prochnik, S.; Chovatia, M.; Grimwood, J.; Jenkins, J.; LaButti, K.; Mavromatis, K.; Nolan, M.; Zane, M.; et al. Genome analysis of Planctomycetes inhabiting blades of the red alga *Porphyra umbilicalis*. *PLoS ONE* **2016**, *11*, e0151883. [[CrossRef](#)]
40. Maruyama, S.; Misumi, O.; Ishii, Y.; Asakawa, S.; Shimizu, A.; Sasaki, T.; Matsuzaki, M.; Shin-i, T.; Nozaki, H.; Kohara, Y.; et al. The minimal eukaryotic ribosomal DNA units in the primitive red alga *Cyanidioschyzon merolae*. *DNA Res.* **2004**, *11*, 83–91. [[CrossRef](#)]
41. Raven, J.A.; Giordano, M. Algae. *Curr. Biol.* **2014**, *24*, R590–R595. [[CrossRef](#)]
42. Zhang, Y.; Romero, H.; Salinas, G.; Gladyshev, V.N. Dynamic evolution of selenocysteine utilization in bacteria: A balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.* **2006**, *7*, R94. [[CrossRef](#)] [[PubMed](#)]
43. Peng, T.; Lin, J.; Xu, Y.Z.; Zhang, Y. Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria. *ISME J.* **2016**, *10*, 2048–2059. [[CrossRef](#)] [[PubMed](#)]
44. Johnson, M.T.; Carpenter, E.J.; Tian, Z.; Bruskiwich, R.; Burris, J.N.; Carrigan, C.T.; Chase, M.W.; Clarke, N.D.; Covshoff, S.; Depamphilis, C.W.; et al. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE* **2012**, *7*, e50226. [[CrossRef](#)] [[PubMed](#)]
45. Rogers SO, B.A. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant. *Plant Mol. Biol.* **1985**, *5*, 59–76. [[CrossRef](#)] [[PubMed](#)]
46. Sahu, S.K.; Thangaraj, M.; Kathiresan, K. DNA Extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *ISRN Mol. Biol.* **2012**, *2012*, 205049. [[CrossRef](#)]
47. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *BMC Bioinform.* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
48. Santesmasses, D.; Mariotti, M.; Guigo, R. Selenoprofiles: A computational pipeline for annotation of selenoproteins. *Methods Mol. Biol.* **2018**, *1661*, 17–28. [[CrossRef](#)]
49. Santesmasses, D.; Mariotti, M.; Guigó, R. Computational identification of the selenocysteine tRNA (tRNA<sup>Sec</sup>) in genomes. *PLoS Comput. Biol.* **2017**, *13*, e1005383. [[CrossRef](#)]
50. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
51. Nakamura, T.; Yamada, K.D.; Tomii, K.; Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *BMC Bioinform.* **2018**, *34*, 2490–2492. [[CrossRef](#)]
52. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
53. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *BMC Bioinform.* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]

54. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)] [[PubMed](#)]
55. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).