



**HAL**  
open science

## Golden magic: RSH enzymes for (p)ppGpp metabolism in the diatom *Phaeodactylum tricornutum*

Luisana Avilan, Carine Puppo, Adrien Villain, Emmanuelle Bouveret, Benoît Menand, Benjamin Field, Brigitte Gontero

### ► To cite this version:

Luisana Avilan, Carine Puppo, Adrien Villain, Emmanuelle Bouveret, Benoît Menand, et al.. Golden magic: RSH enzymes for (p)ppGpp metabolism in the diatom *Phaeodactylum tricornutum*. 2018. cea-01950162

HAL Id: cea-01950162

<https://cea.hal.science/cea-01950162>

Preprint submitted on 10 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Title:**

2 **Golden magic: RSH enzymes for (p)ppGpp metabolism in the diatom *Phaeodactylum***  
3 ***tricornutum*.**

4  
5 **Running title:**

6 **RSH enzymes for (p)ppGpp metabolism in in the diatom *Phaeodactylum tricornutum*.**

7  
8 L Avilan<sup>1</sup>, C Puppo<sup>1</sup>, A Villain<sup>2</sup>, E Bouveret<sup>3</sup>, B Menand<sup>4</sup>, B Field<sup>4\*</sup> and B Gontero<sup>1\*</sup>

9 <sup>1</sup>Aix Marseille Univ CNRS, BIP, UMR 7281, IMM FR 3479, 31 Chemin Joseph Aiguier,  
10 Marseille, France F-13009

11 <sup>2</sup>Aix Marseille Univ CNRS, IGS; UMR 7256, IMM FR 3479, Marseille, France F-13009

12 <sup>3</sup> Stress Adaptation and Metabolism in Enterobacteriae group, Department of Microbiology,  
13 Institut Pasteur, Paris 75015, France

14 <sup>4</sup> Aix Marseille Univ, CEA, CNRS, BIAM, Laboratoire de génétique et biophysique des plantes,  
15 Marseille, France F-13009

16

17 \*Corresponding authors :[ben.field@univ-amu.fr](mailto:ben.field@univ-amu.fr) and [bmeunier@imm.cnrs.fr](mailto:bmeunier@imm.cnrs.fr)

18

19 Total word count: 5240 words

20 Introduction (814 words), Materials and methods (740 words), Results (2251 words), Discussion  
21 (1216 words), Acknowledgements (13 words). SUM = 5037 words

22 4 Figures, and Supporting Information (3 Supplementary Figures, 4 Supplementary Tables and  
23 Two Supplementary Datasets).

24

## 1 **Highlight**

2 We discover RSH enzymes for (p)ppGpp metabolism in the diatom *Phaeodactylum tricornutum*  
3 and show that they have surprising functional and structural features, and belong to novel red-  
4 plastid lineage RSH clades.

## 6 **Abstract**

7 The nucleotides guanosine tetraphosphate and pentaphosphate (together known as (p)ppGpp or  
8 magic spot) are produced in plant plastids from GDP/GTP and ATP by RelA-SpoT homologue  
9 (RSH) enzymes. In the model plant Arabidopsis (p)ppGpp regulates chloroplast transcription and  
10 translation to affect growth, and is also implicated in acclimation to stress. However, little is  
11 known about (p)ppGpp metabolism or its evolution in other photosynthetic eukaryotes. Here we  
12 studied (p)ppGpp metabolism in the golden-coloured marine diatom *Phaeodactylum tricornutum*.  
13 We identified three expressed *RSH* genes in the *P. tricornutum* genome, and determined the  
14 enzymatic activity of the corresponding enzymes by heterologous expression in bacteria. We  
15 showed that two *P. tricornutum* RSH are (p)ppGpp synthetases, despite substitution of a residue  
16 within the active site believed critical for activity, and that the third RSH is a bifunctional  
17 (p)ppGpp synthetase and hydrolase, the first of its kind demonstrated in a photosynthetic  
18 eukaryote. A broad phylogenetic analysis then showed that diatom RSH belong to novel algal  
19 RSH clades. Together our work significantly expands the horizons of (p)ppGpp signalling in the  
20 photosynthetic eukaryotes by demonstrating an unexpected functional, structural and  
21 evolutionary diversity in RSH enzymes from organisms with plastids derived from red algae.

## 23 **Keywords**

24 Chloroplast, diatom, *Phaeodactylum tricornutum*, plastid, (p)ppGpp, RSH

## 26 **Abbreviations**

27 ACT Aspartate kinase, Chorismate mutase, TyrA domain  
28 CASH Cryptomonads, Alveolates, Stramenopiles and Haptophytes  
29 CRSH Ca<sup>2+</sup> activated RelA/SpoT homolog  
30 EF-hand Ca<sup>2+</sup> binding domain  
31 (p)ppGpp guanosine tetraphosphate and pentaphosphate

- 1 Rel<sub>seq</sub> *Streptococcus equisimilis* Rel
- 2 RSH RelA/SpoT homologue
- 3 TGS Threonyl-tRNA synthetase, GTPase, SpoT and RelA domain
- 4

## 1 **Introduction**

2 Plastids, the defining feature of photosynthetic eukaryotes, arose from the endosymbiosis of a  
3 cyanobacterium more than one billion years ago (Ponce-Toledo *et al.*, 2017). Massive  
4 cyanobacterial gene loss occurred following endosymbiosis, along with the transfer of genes to  
5 the nuclear genome of the eukaryotic host. Plastids now retain a small genome that encodes  
6 housekeeping and photosynthetic proteins, as well as a bacteria-like gene expression machinery.  
7 Bacteria-like regulatory systems that may be involved in acclimation to environmental  
8 perturbation are also present in plastids (Puthiyaveetil *et al.*, 2008; Field, 2018). One of these  
9 regulatory-systems is mediated by the nucleotides guanosine tetraphosphate and pentaphosphate  
10 (referred to as (p)ppGpp hereafter) whose levels are controlled by the antagonistic action of  
11 RelA-SpoT homologues (RSH). In bacteria, (p)ppGpp was originally identified as a magic spot  
12 on thin-layer chromatography plates in the 1960s (Cashel & Gallant, 1969). Now bacterial  
13 (p)ppGpp signalling is well characterised: (p)ppGpp accumulates in response to stress to reduce  
14 proliferation and activates acclimatory pathways by targeting enzymes involved in transcription,  
15 translation, and replication (Dalebroux & Swanson, 2012; Hauryliuk *et al.*, 2015).

16 In the photosynthetic eukaryotes, (p)ppGpp was discovered more recently, and (p)ppGpp  
17 signalling has principally been studied in flowering plants. Although (p)ppGpp accumulates in  
18 response to various different stresses (Takahashi *et al.*, 2004; Ihara *et al.*, 2015), the actual role of  
19 (p)ppGpp during stress acclimation is not yet clear. However, (p)ppGpp is known to act as a  
20 potent inhibitor of plastid gene expression *in vivo* (Maekawa *et al.*, 2015; Yamburenko *et al.*,  
21 2015; Sugliani *et al.*, 2016), and altering the capacity of a plant to make (p)ppGpp influences  
22 photosynthetic capacity, growth and development (Maekawa *et al.*, 2015; Sugliani *et al.*, 2016).  
23 Notably, (p)ppGpp appears to be important for regulating the equilibrium between the plastidic  
24 and nucleocytoplasmic compartments of the plant cell. Phylogenetic studies support the existence  
25 of at least three conserved families of plastid-targeted RSH enzymes in land plants named RSH1,  
26 RSH2/3 and RSH4 (Atkinson *et al.*, 2011). The model flowering plant *Arabidopsis thaliana*,  
27 where plant (p)ppGpp homeostasis is the most well understood, possesses representatives from  
28 each of these families: RSH1 that lacks (p)ppGpp synthetase activity and appears to function as  
29 the major (p)ppGpp hydrolase (Sugliani *et al.*, 2016), the closely related RSH2 and RSH3 that  
30 appear to act as the major (p)ppGpp synthetases (Mizusawa *et al.*, 2008; Maekawa *et al.*, 2015;  
31 Sugliani *et al.*, 2016), and a calcium- activated RSH (CRSH) from the RSH4 family that

1 possesses a C-terminal EF-hand domain implicated in calcium binding, and has calcium-  
2 dependent (p)ppGpp synthesis activity *in vitro* (Masuda *et al.*, 2008).

3 Plastids are the result of primary endosymbiosis, where a photosynthetic bacterium was  
4 engulfed by a eukaryote. Today several primary plastid lineages can be identified which are  
5 thought to share common ancestry. Plants and green algae (together known as the Viridiplantae)  
6 possess primary plastids of the green lineage. The Viridiplantae have two sister groups with  
7 primary plastids, the red algae (Rhodophyta) and glaucophytes (Glaucophyta). Primary plastids  
8 from the green and red lineages have also been transferred and mixed in new eukaryotic hosts to  
9 result in the astonishing range of plastid diversity that can be observed in nature (Archibald,  
10 2015). Plastid-targeted RSH enzymes from the RSH1, RSH2/3 and RSH4 families have been  
11 identified in many green and red algae, yet the metabolism and functions of (p)ppGpp in these  
12 organisms have barely been investigated (Atkinson *et al.*, 2011; Ito *et al.*, 2017; Field, 2018). One  
13 exception is a recent report on the red alga *Cyanidioschyzon merolae*, where CmRSH4b, a  
14 member of the RSH4 family, was shown to possess (p)ppGpp synthetase activity (Imamura *et al.*,  
15 2018). Interestingly, the inducible expression of CmRSH4b in *C. merolae* results in a reduction in  
16 plastid size and rRNA transcription in a similar manner to the expression of a (p)ppGpp  
17 synthetase in *Arabidopsis* (Sugliani *et al.*, 2016).

18 Diatoms (Bacillariophyceae) are a group of golden-coloured microalgae that contain  
19 complex plastids that originate from the secondary endosymbiosis of a red alga, with the addition  
20 of nuclear-encoded plastid-targeted green algal proteins left over from a previous endosymbiosis  
21 (Dorrell & Smith, 2011; Dorrell *et al.*, 2017). Diatoms are the predominant photosynthetic  
22 eukaryote in the oceans, where they account for around 40% of net primary production (Malviya  
23 *et al.*, 2016). Therefore, understanding (p)ppGpp synthesis in diatoms, where it is likely to  
24 regulate photosynthetic capacity (Maekawa *et al.*, 2015; Sugliani *et al.*, 2016), and play roles in  
25 diatom lifestyle, is of particular importance. To tackle this issue, we investigate (p)ppGpp  
26 metabolism in the model pennate marine diatom *Phaeodactylum tricornutum*. We first identified  
27 expressed *RSH* genes in the *P. tricornutum* genome and then determined if their gene products  
28 have (p)ppGpp synthetase / hydrolase activity by complementation of *Escherichia coli* (p)ppGpp  
29 biosynthesis mutants. Then, we place the structural and catalytic features of *P. tricornutum* RSH  
30 enzymes into an evolutionary context. Altogether our study advances our previously poor  
31 understanding of (p)ppGpp metabolism in the complex plastids of diatoms.

## 1 **Materials and Methods**

2

### 3 **Cloning of *P. tricornutum* RSH**

4 Genomic DNA from *P. tricornutum* Bohlin (strain number CCAP 1052/1A from Culture  
5 Collection of Algae and Protozoa, CCAP, Scottish Marine Institute) was isolated from cells  
6 grown in late-exponential phase, using an optimized method for diatoms (Puppo *et al.*, 2017;  
7 Kojadinovic-Sirinelli *et al.*, 2018). Sequences for the *P. tricornutum* RSH proteins: PtRSH1,  
8 PtRSH4a and PtRSH4b (accession codes: 11099, 7629 and 3397, respectively) were retrieved  
9 from Joint Genome Initiative (JGI) database  
10 (<https://genome.jgi.doe.gov/Phatr2/Phatr2.home.html>). The amino acid sequences of PtRSH1 and  
11 PtRSH4a reported in JGI were partial, so a manual search based on nucleotide sequence was  
12 performed to obtain the full-length amino acid sequences. The three PtRSH sequences are given  
13 in Fig. S1. The genes coding for the mature proteins without the bipartite chloroplast targeting  
14 peptide were amplified by PCR using the primers listed in Table S1. The theoretical transit  
15 peptide cleavage site was estimated from (Huesgen *et al.*, 2013) and by using alignments with  
16 plant RSH enzymes. Primers were designed to have 20-bp flanking regions homologous to the  
17 plasmid (pBAD24) each side of the *NcoI* and *HindIII* restriction sites. The plasmid construction  
18 was carried out using the Sequence and Ligation Independent Cloning (SLIC) method (Jeong *et*  
19 *al.*, 2012). PCR was performed using genomic DNA of *P. tricornutum*, the specified primers and  
20 Q5 DNA polymerase (New England Biolabs). Since the gene encoding PtRSH1 has an intron, the  
21 complete cloned gene was used as a template for the amplification of the two exons using the  
22 primers specified in Table S1. The final construct was assembled in the pBAD24 plasmid using  
23 SLIC cloning (Jeong *et al.*, 2012). All constructs were verified by sequencing (GATC Biotech).

24

### 25 **RSH activity assays**

26 To test (p)ppGpp synthetase activity, pBAD24 plasmids containing the *PtRSH* genes, SpoT or  
27 without a coding-sequence (empty) were transformed into *E. coli* strain EB425 (MG1655  $\Delta relA$   
28  $\Delta spoT$ ) (Wahl *et al.*, 2011) and grown at 37°C on M9 minimal medium agar plates without  
29 amino acids and in the presence or absence of 0.1% arabinose. To test (p)ppGpp hydrolase  
30 activity, the same plasmids were transformed into *E. coli* strain EB544 (MG1655  $\Delta relA spoT203$ )  
31 (My *et al.*, 2013). Pre-cultures from independent colonies for each replicate were diluted in 50 ml

1 Luria Bertani medium containing ampicillin (100  $\mu\text{g mL}^{-1}$ ) and arabinose (0.1% w/v). Growth  
2 was performed at 37°C under agitation at 175 rpm and optical density was measured at 600 nm  
3 every 30 min.

#### 4 5 **Prediction of plant and algal RSH enzyme structures**

6 Iterative Threading ASSEmbly Refinement (iTASSER)(Yang *et al.*, 2015) was used to predict  
7 the structure and ligands of the synthetase and hydrolase domain containing region of  
8 Arabidopsis and *P. tricornutum* RSH enzymes (default settings). The synthetase hydrolase  
9 regions were chosen by selection of the regions corresponding most closely to the synthetase  
10 region of *Bacillus subtilis* SAS1: AtRSH1, residues 347-542; AtRSH3, residues 404-559;  
11 PtRSH1, residues 403-584; PtRSH4a, residues 352-537; PtRSH4b, residues 432-631. The *B.*  
12 *subtilis* SAS1 crystal structure (9decA) was then manually selected in iTASSER for modelling.  
13 Results were visualized in PyMOL (Version 2.0 Schrödinger, LLC).

#### 14 15 **Phylogenetic analysis of RSH proteins from diatoms and other photosynthetic organisms**

16 Using *P. tricornutum* RSH proteins as queries, a set of homologous proteins was built using  
17 public data at the National Center for Biotechnology Information (NCBI)  
18 and JGI, as well as individual genome projects : *Synedra acus*  
19 (<http://www.lin.irk.ru/sacus/index.php?r=site/page&view=downloads&lang=en>), *Cyclotella*  
20 *cryptica* (<http://genomes.mcdb.ucla.edu/Cyclotella/download.html>), *Pseudo nitzschia*  
21 *multistriata* (<https://zenodo.org/record/495408#.W1sOJsIyXIU>) and *Asterionella formosa*  
22 (unpublished data, A. Villain, personal communication). RSH from similar plant and green algal  
23 species to those selected by Ito et al., (2017) were used. RSH proteins were selected only from  
24 fully sequenced genomes, and all predicted RSH per genome were used in the subsequent  
25 analysis. Two datasets were built: a 'whole dataset' containing all RSH sequences, and a 'RSH1  
26 specific set' including only RSH1 sequences. Multiple-sequence alignments of homologous  
27 proteins were performed separately on both datasets using MAFFT v7.402 (Kato *et al.*, 2017)  
28 with option –auto. Initial alignment of the 'whole dataset' was trimmed to include only the  
29 hydrolase and synthetase domains, by selecting columns 26 to 343 from *E. coli* SpoT, while full-  
30 length sequences were kept in the 'RSH1 specific' alignment. Alignment columns containing gaps  
31 in >30% sequences were then removed from both alignments (Supplementary dataset), and



1 maximum likelihood phylogenetic reconstructions were performed with FastTree version 2.1.10  
2 (Price *et al.*, 2010) with options -slow -mlacc 3 -wag -gamma. Protein domain architecture was  
3 analyzed by CD-search algorithm of NCBI with default parameters and an E-value threshold of  
4 0.01 (Marchler-Bauer *et al.*, 2015).

5

## 1 **Results**

### 2 3 **The nuclear genome of the model diatom *Phaeodactylum tricornutum* encodes three plastid-** 4 **targeted RSH enzymes**

5 We inspected the *P. tricornutum* genome for the presence of *RSH* genes using BLAST and  
6 identified three: *PtRSH1*, *PtRSH4a* and *PtRSH4b*. *PtRSH1* is the only *RSH* gene that contains an  
7 intron. We analysed the predicted protein sequences derived from the three genes and identified a  
8 number of conserved domains that are typical of RSH enzymes (Fig. 1A). The N-terminus of all  
9 the three RSH also possesses a domain with strong resemblance to a bipartite signal peptide,  
10 which is necessary for the transport of nuclear-encoded proteins to the plastid in diatoms  
11 according to Gruber et al. (2015). Analysis of amino acid frequency around the potential cleavage  
12 site in each sequence supported this assignment, and suggesting that all three RSH are plastid  
13 targeted (Fig 1B).

14 PtRSH4a and PtRSH4b bear (p)ppGpp hydrolase and synthetase domains that show signs  
15 of catalytic inactivation (Fig S2). The (p)ppGpp hydrolase domains are divergent compared to the  
16 hydrolase domains of known (p)ppGpp hydrolases, and lack many residues critical for hydrolase  
17 activity (Steinchen & Bange, 2016). While PtRSH4a and PtRSH4b contain domains with strong  
18 homology to bacterial (p)ppGpp synthetase domains, a glycine residue corresponding to G240 in  
19 *Streptococcus equisimilis* Rel (Relseq) and previously shown to be essential for synthetase  
20 activity in bacterial RSH (Wendrich & Marahiel, 1997; Hogg *et al.*, 2004) is substituted by an  
21 alanine (PtRSH4a) or a serine residue (PtRSH4b). Substitution of this glycine residue in the RSH  
22 enzymes of land plants has also been associated with the loss of (p)ppGpp synthetase activity  
23 (Masuda *et al.*, 2008; Sugliani *et al.*, 2016). In contrast, PtRSH1 bears well conserved (p)ppGpp  
24 hydrolase and synthetase domains. This analysis suggests that PtRSH1 may be a bifunctional  
25 enzyme, unlike the monofunctional RSH that have so far been identified in plants and green and  
26 red algae.

27 In addition to the N-terminal catalytic domains, all the three PtRSH also possess C-  
28 terminal regulatory domains, as observed in bacterial RSH enzymes. PtRSH1, PtRSH4a and  
29 PtRSH4b have a TGS domain, named after the proteins where it occurs: Threonyl-tRNA  
30 synthetase, GTPase, SpoT and RelA proteins. PtRSH1 has, in addition, an Aspartate kinase,  
31 Chorismate mutase, TyrA (ACT) domain, a conserved regulatory fold found in enzymes

1 regulated by small metabolites. The TGS and ACT domains have both been implicated in  
2 protein-protein interactions in bacterial RSH enzymes (Wout *et al.*, 2004; Battesti & Bouveret,  
3 2006; Ronneau *et al.*, 2018).

4 We then analyzed *P. tricornutum* expressed sequence tag (EST) databases (Scala *et al.*, 2002;  
5 Bowler *et al.*, 2008) to determine whether the genes encoding PtRSH are expressed. *PtRSH1*,  
6 *PtRSH4a* and *PtRSH4b* are indeed expressed in standard growth conditions (Table S2). *PtRSH4a*  
7 had the greatest number of ESTs indicating higher expression levels, and *PtRSH4a* EST  
8 abundance increased in response to a number of stresses including high levels of 2E, 4E  
9 decadienal, a volatile oxylipin associated with diatom stress signalling (Vardi *et al.*, 2006)(Table  
10 S2). This may suggest that diatom *RSH* genes are involved in defence responses, as previously  
11 observed for plant *RSH* (van der Biezen *et al.*, 2000; Givens *et al.*, 2004; Kim *et al.*, 2009;  
12 Abdelkefi *et al.*, 2018). Furthermore, recent transcriptome experiments show that *PtRSH1* and  
13 *PtRSH4b* are induced rapidly after the onset of nitrogen starvation and in response to incubation  
14 in darkness for eight hours (Matthijs *et al.*, 2016; Matthijs *et al.*, 2017).

15

### 16 **Catalytic activities of the *P. tricornutum* RSH enzymes**

17 Next, we investigated the potential catalytic activities of PtRSH1, PtRSH4a and PtRSH4b by  
18 expressing the predicted mature proteins in *E. coli* strains where the endogenous RSH genes, *relA*  
19 and *spoT*, are mutated.

20 To assay (p)ppGpp synthetase activity, we expressed the *P. tricornutum* RSH enzymes in  
21 a  $\Delta relA \Delta spoT$  mutant. This mutant is unable to grow on minimal medium, and growth can be  
22 restored by complementation using a plasmid expressing a functional (p)ppGpp synthetase. The  
23 expression of PtRSH1, PtRSH4a and PtRSH4b restored growth to the  $\Delta relA \Delta spoT$  mutant on  
24 minimal medium, as did expression of the positive control *spoT* gene from *E. coli* (Fig. 2A).  
25 Thus, all three PtRSH enzymes have (p)ppGpp synthetase activity.

26 We then used an assay that we recently developed to detect (p)ppGpp hydrolase activity  
27 using the  $\Delta relA spoT203$  mutant strain (Sugliani *et al.*, 2016). In this mutant strain SpoT is  
28 defective for (p)ppGpp hydrolase activity while retaining (p)ppGpp synthetase activity, which  
29 results in the constitutive over-accumulation of (p)ppGpp and a slow growth phenotype. Normal  
30 growth can be restored by expression of an enzyme with (p)ppGpp hydrolase activity such as  
31 SpoT. We found that the expression of the diatom PtRSH1 rescued the growth of this mutant in a

1 similar fashion to expression of the positive control *spoT* (Fig. 2B). In contrast, expression of  
2 PtRSH4a and PtRSH4b in the  $\Delta relA spoT203$  mutant strain did not restore normal growth, and  
3 we even observed slower growth, perhaps due to enhanced (p)ppGpp biosynthesis. The absence  
4 of detectable (p)ppGpp hydrolase activity in PtRSH4a and PtRSH4b is consistent with the  
5 divergent (p)ppGpp hydrolase domains in these enzymes that lack many of the residues essential  
6 for activity. All together, these results demonstrate that PtRSH4a and PtRSH4b are exclusive  
7 (p)ppGpp synthetase enzymes, and that PtRSH1 is a bifunctional RSH enzyme with (p)ppGpp  
8 synthetase and hydrolase activity.

9

### 10 **Active site modelling of *P. tricornutum* RSH enzymes**

11 All three RSH from *P. tricornutum* have unusual characteristics. As shown above, PtRSH1 is a  
12 bifunctional (p)ppGpp synthetase and hydrolase, a class of enzyme common in bacteria but not  
13 yet described in plants or algae. Furthermore, PtRSH4a and PtRSH4b are (p)ppGpp synthetases,  
14 despite the lack of a conserved glycine residue previously shown to be critical for (p)ppGpp  
15 synthesis as discussed above (Wendrich & Marahiel, 1997; Hogg *et al.*, 2004)(Fig. 3A).  
16 Therefore, to gain functional insights into these characteristics, we modelled the structures of the  
17 synthetase domains of the three *P. tricornutum* RSH enzymes using I-TASSER (Yang *et al.*,  
18 2015)(Fig. 3B). For the three PtRSH enzymes as well as for CRSH and RSH1 from *A. thaliana*  
19 we obtained models that matched the crystal structure of Small Alarmone Synthase 1 (SAS1)  
20 from *Bacillus subtilis* bound to the non-hydrolysable ATP analog  $\alpha,\beta$ -methyleneadenosine 5'-  
21 triphosphate (AMPCPP) (Steinchen *et al.*, 2015). Within the active site of SAS1 the conserved  
22 glycine residue G45 (corresponding to Relseq G240) is positioned on a beta strand neighbouring  
23 R46, a residue involved in binding AMPCPP in cooperation with a number of other residues (Fig.  
24 3A, 3B). In PtRSH4a and PtRSH4b the side-chains of the non-conserved residues corresponding  
25 to G45 did not greatly affect the active site structure, and their side chains projected away from  
26 the active site (Fig. 3B). In both cases, the neighbouring arginine residue corresponding to SAS1  
27 R46 was still predicted to bind AMPCPP. Interestingly, in the residues corresponding to R46 the  
28 arginine side-chain took several different orientations in AtRSH3, PtRSH4a, PtRSH4b and  
29 PtRSH1, all of which appear to aid in retaining the adenine and ribose moieties of AMPCPP.

30

1 In AtRSH1, which lacks (p)ppGpp synthetase activity (Sugliani *et al.*, 2016), the residue  
2 corresponding to R46 is substituted by a valine, whose side-chain does not retain AMCPP and  
3 due to its polarity cannot bond with the oxygens in the ribose ring (Fig. 3B). Many other  
4 conserved residues involved in AMCPP binding are also missing from AtRSH1. In contrast, the  
5 synthetase active site of PtRSH1 more closely resembles that of an active (p)ppGpp synthetase:  
6 there is a glycine residue at the position corresponding to SAS1 G45 and the neighbouring  
7 arginine involved in AMCPP binding takes a similar orientation to the corresponding arginine in  
8 AtRSH3, which acts as a ppGpp synthetase (Fig. 3B). These modelling results strongly support  
9 the results of our activity assays (Fig. 2), and show that the residue corresponding to SAS1 G45  
10 can be substituted by at least alanine (PtRSH4a) or serine (PtRSH4b) residues, without  
11 necessarily affecting the conformation of the active site or the orientation of the neighbouring  
12 arginine, R46.

13

#### 14 **Evolution and domain organization of *P. tricornutum* and other algal RSH enzymes**

15 Our work shows that the *P. tricornutum* RSH enzymes have significant differences with known  
16 plant and algal RSHs. However, the diatom RSH enzymes have not yet been placed within a  
17 phylogenetic context. This is necessary for determining whether they belong to new RSH clades,  
18 and for understanding when the RSH enzymes were acquired during the complex evolutionary  
19 history of the diatom plastid. Therefore, we constructed a phylogenetic tree based on the  
20 sequences of the hydrolase and synthetase domains of different RSH enzymes from bacteria,  
21 plants and algae (Fig. 4).

22

23 PtRSH4a and PtRSH4b belong to a clade that is monophyletic with a sister clade containing  
24 the plant and green algae RSH4 enzymes, and which together can be considered as part of an  
25 extended RSH4 clade (Fig. 4). PtRSH4a and PtRSH4b are members of two well separated diatom  
26 RSH clades that form part of a larger clade composed of RSH from red algae and the CASH  
27 lineages (organisms descended from an endosymbiosis between a eukaryote and a red alga:  
28 Cryptomonads, Alveolates, Stramenopiles and Haptophytes)(Petersen *et al.*, 2014). We will refer  
29 to this clade as the red RSH4 clade, based on the principal source of chloroplast ancestry. Within  
30 the red RSH4 clade the substitution of the residue equivalent to G240 that we observed in  
31 PtRSH4a and PtRSH4b is widespread among the CASH lineages but not red algae. Interestingly,

1 the plant and algal sister clade also contains RSH4 enzymes from some red algae and CASH  
2 lineages, but not from diatoms. This clade can therefore be considered a red and green RSH4  
3 clade based on chloroplast ancestry. The existence of two monophyletic RSH4 clades indicates  
4 that the common ancestor of was likely to have had two RSH4 like enzymes that were  
5 subsequently retained, or selectively lost/duplicated in the different lineages.

6  
7 PtRSH1 likewise belongs to a red algal and CASH lineage clade (or red RSH1 clade) that is  
8 monophyletic with a sister clade containing the plant and green algal RSH1 enzymes (red and  
9 green RSH1 clade), and which together can be considered as part of an extended RSH1 clade  
10 (Fig. 4). The majority of members of the red RSH1 clade possess a synthetase domain where all  
11 residues known to be critical for (p)ppGpp synthetase activity are intact, suggesting that all these  
12 enzymes have the potential for bifunctionality. The red and green RSH1 clade contains RSH1  
13 enzymes from several CASH organisms, but there are no red algal RSH1 at its base. For this  
14 reason, it is more difficult to infer the evolutionary relationship between the two major RSH1  
15 clades. Phylogenetic analysis of full-length RSH1-like sequences complete with the C-terminal  
16 regulatory domain also supported the existence of a red RSH1 clade that is distinct from the red  
17 and green RSH1 clade (Fig. S3). However, the full-length RSH1 tree also did not clearly resolve  
18 the evolutionary history of the extended RSH1 clade. As observed in previous analyses of plant  
19 and algal RSH enzymes we found that the RSH1 enzymes group closely with RSH enzymes from  
20 the *Deinococcus*–*Thermus* bacterial phylum (Givens *et al.*, 2004; Atkinson *et al.*, 2011; Ito *et al.*,  
21 2017).

22  
23 Our analysis also reveals that other diatom species possess additional RSH enzymes to  
24 those found in *P. tricornutum*. Unlike the PtRSH, these additional RSH enzymes are not in clades  
25 that are monophyletic with RSH enzymes from plants or green algae, and form two distinct  
26 clades within the proteobacterial RSH (top of Fig. 4). These novel RSH clades are unlikely to be  
27 sequencing artefacts because there are multiple representatives from different species.  
28 Interestingly, we also found no evidence that CASH lineage organisms possess members of the  
29 RSH2/3 family suggesting that this family is unique to plants and green algae. Two CASH  
30 organisms, *Emiliania huxleyi* and *Vitrella brassicaformis*, also have orphan RSH that are not

1 associated with other algal RSH or with the major RSH families of plants, algae or bacteria (Fig.  
2 4).

3  
4 Recent studies strongly suggest that plastid arose from the endosymbiosis of a freshwater  
5 Gloeomargarita-like cyanobacterium (Ponce-Toledo *et al.*, 2017; Sánchez-Baracaldo *et al.*,  
6 2017). Strikingly, our phylogenetic analysis gives no indication that any of the major RSH groups  
7 (RSH1, RSH2/3 and RSH4) are more closely related to Gloeomargarita, or indeed any  
8 cyanobacterial species, than the RSH from other bacteria.

9

### 10 **Domain organization of *P. tricornutum* and other algal RSH enzymes**

11 In terms of domain structure, the majority of plant and green algal RSH4 have acquired a C-  
12 terminal EF-hand domain (Fig. 4). Two stramenopile RSH4 in the same clade as the plant and  
13 algal RSH4 also possess EF-hand domains, though at the N-terminus. No EF-hand domains could  
14 be detected in PtRSH4a and PtRSH4b or indeed any members of the sister RSH4 clade that  
15 contains only CASH and red algal RSH. However, a C-terminal MoaD-like protein domain was  
16 encountered in three separate members of this clade. We also observed isolated domain  
17 acquisitions in both clades of the extended RSH4 family: a major intrinsic protein domain in the  
18 alveolate *Symbiodinium microadriaticum*, a minor capsid protein VI domain in the alveolate  
19 *Vitrella brassicaformis*, a phosphopantetheinyl transferase domain in the haptophyte *Emiliania*  
20 *huxleyi*, and a forkhead domain in the green alga *Chlorella variabilis*. Interestingly, Ito et al.  
21 (2017) previously reported the acquisition of tetratricopeptide repeat domains at the C-terminus  
22 of RSH4 from some green algae species, acquisitions that were also detected here.

23

24 Less evidence of domain acquisition could be observed within the extended RSH1 family,  
25 where the majority of enzymes, including PtRSH1, contain both TGS and ACT domains.  
26 Nevertheless, an RSH1 from the brown alga *E. siliculosus* contains a Chlorophyll A-B binding  
27 protein domain, and an RSH1 from the haptophyte *Chrysochromulina* sp contains an endomucin  
28 domain.

29

## 1 **Discussion**

2 In most bacteria, nutrient limitation or stress provokes the rapid synthesis of the two nucleotide  
3 alarmones, pppGpp and ppGpp. In *E. coli*, (p)ppGpp synthesis is controlled by two enzymes,  
4 RelA and SpoT. Chloroplast localised RelA SpoT homologue (RSH) enzymes are also found in  
5 algae and plants, where they are implicated in the control of chloroplast function and plant  
6 development (Boniecka *et al.*, 2017; Field, 2018). Little is known about these RSH enzymes in  
7 diatoms, or indeed the entire red algal lineage. In this report, we showed that the genome of the  
8 diatom *P. tricornutum* encodes three RSH enzymes: PtRSH1, PtRSH4a and PtRSH4b. Analysis  
9 of their amino acids sequence revealed the presence of probable chloroplast targeting peptides,  
10 indicating that these enzymes are likely to be located within the chloroplast, the site of (p)ppGpp  
11 synthesis and its action in plants (Field, 2018) (Fig. 1A). Using heterologous expression in *E. coli*  
12 (p)ppGpp mutants we showed that all three *P. tricornutum* RSH are (p)ppGpp synthetases, and  
13 that PtRSH1 is also a (p)ppGpp hydrolase (Fig. 2). Gene expression data indicates that the *P.*  
14 *tricornutum* RSH are expressed, and that PtRSH4a is strongly induced in response to certain  
15 stresses (Table S2) (Scala *et al.*, 2002; Bowler *et al.*, 2008; Matthijs *et al.*, 2016; Matthijs *et al.*,  
16 2017). Finally, phylogenetic analysis indicated that PtRSH1, PtRSH4a and PtRSH4b are typical  
17 members of well-defined diatom RSH clades that belong to the previously identified plant and  
18 algal RSH1 and RSH4 families (Fig. 4)(Atkinson *et al.*, 2011; Ito *et al.*, 2017). These results  
19 indicate that the *P. tricornutum* RSH enzymes and their orthologues are sufficient for the  
20 establishment of a (p)ppGpp homeostasis mechanism in the chloroplasts of diatoms.

21  
22 PtRSH1 is unusual in that it functions as a bifunctional (p)ppGpp synthetase / hydrolase  
23 RSH. Bifunctional RSH such as SpoT are common in bacteria, but no bifunctional RSH enzymes  
24 have been demonstrated in the photosynthetic eukaryotes before now. We note however that  
25 members of the plant and green algal RSH2/3 family often possess almost intact hydrolase  
26 domains in addition to active synthetase domains (Fig. 4). However, hydrolase activity has not so  
27 far been demonstrated for one of these enzymes and their overexpression in plants results in  
28 (p)ppGpp accumulation (Maekawa *et al.*, 2015; Sugliani *et al.*, 2016). Interestingly, our  
29 phylogenetic analysis suggests that all members of the red RSH1 clade may be bifunctional (Fig.  
30 4). This is in contrast to the sister red and green RSH1 clade where the majority of RSH1 appear  
31 to have lost (p)ppGpp synthetase activity due to the loss of essential residues in the active site



1 (Sugliani *et al.*, 2016; Ito *et al.*, 2017). Our results are also supported by modelling of the active  
2 sites of PtRSH1 and Arabidopsis RSH1 (Fig. 3b). Arabidopsis RSH1, while preserving the global  
3 topology of the synthetase active site, has lost the arginine residue with a guanidinium group that  
4 is likely to retain ATP in its binding pocket. Structural data indicates that bifunctional RSH can  
5 switch between (p)ppGpp-hydrolase-OFF / (p)ppGpp-synthase-ON and hydrolase-ON / synthase-  
6 OFF configurations (Hogg *et al.*, 2004). Our heterologous expression results suggest that PtRSH1  
7 is likely to be capable of undergoing similar activity switches (Fig. 2). The regulation of PtRSH1  
8 and other members of the red RSH1 clade may therefore be significantly different to plant and  
9 green algal RSH1 and may reflect the different lifestyles of these organisms.

10  
11 PtRSH4a and PtRSH4b are also unusual among (p)ppGpp synthetases due to the  
12 substitution of a residue corresponding to SAS1 G45 / Rel<sub>seq</sub> G240. This conserved residue has  
13 previously been shown to be necessary for (p)ppGpp synthetase activity (Wendrich & Marahiel,  
14 1997; Hogg *et al.*, 2004) and substitution of this residue is widely used to infer the loss of  
15 (p)ppGpp synthetase activity (Masuda *et al.*, 2008; Mizusawa *et al.*, 2008; Sato *et al.*, 2015; Ito *et*  
16 *al.*, 2017; Imamura *et al.*, 2018). Modelling of the active site of PtRSH4a and PtRSH4b shows  
17 that the glycine substitution does not appear to affect ATP binding by the neighbouring arginine  
18 (Fig. 3b), and sequence analysis indicates that the glycine substitution is widespread in the red  
19 RSH4 clade. These data suggest that the previous reports of an association between the  
20 substitution of the residue corresponding to SAS1 G45 and the loss of synthetase activity may be  
21 conditional on the presence of other specific residues in the active site. PtRSH4a and PtRSH4b,  
22 and nearly all members of red RSH4 clade lack the EF-hand domain found at the C-terminal of  
23 the majority of members of the sister red and green RSH4 clade. A certain number of novel  
24 domain acquisitions were also observed within both the major RSH4 clades in addition to those  
25 previously identified in the red and green RSH4 clade (Ito *et al.*, 2017). Altogether, these findings  
26 suggest that extended RSH4 family enzymes act exclusively as (p)ppGpp synthetases, and are  
27 susceptible to domain acquisition, presumably for new regulatory functions. Indeed, altered  
28 regulation might be expected, because many enzymes from diatoms involved in processes such as  
29 CO<sub>2</sub> assimilation, sulphate assimilation have different regulatory properties than their orthologues  
30 in plants (Maberly *et al.*, 2010; Jensen *et al.*, 2017). Determining the functions of these domains  
31 in (p)ppGpp metabolism presents a fascinating challenge for future research.

1  
2 We show here that, when the RSH of genome-sequenced diatoms are considered, the  
3 evolutionary history of RSH enzymes in the photosynthetic eukaryotes is considerably more  
4 complex than previously thought (Givens *et al.*, 2004; Atkinson *et al.*, 2011; Ito *et al.*, 2017). We  
5 extend the described RSH1 and RSH4 families by showing the existence of sister clades specific  
6 to the red-lineage that show distinct functional and structural properties (Fig. 2-4). Interestingly,  
7 although Atkinson *et al.* (2011) included several CASH and red algal RSH1 enzymes in their  
8 phylogenetic analysis, they did not detect a clear separation of the RSH1 family into a red RSH1  
9 and a red and green RSH1 clade, or the existence of a red RSH4 clade. The difference between  
10 these studies is very likely to be due to our use of many more CASH RSH sequences, which are  
11 available today thanks to the ever growing list of sequenced CASH genomes. For the same  
12 reason we also detected two new CASH RSH families that are not monophyletic with any of the  
13 known RSH families in photosynthetic eukaryotes.

14  
15 Previous reports on the phylogenetic relationships of bacterial, plant and algal RSH  
16 enzymes have proposed that plant and algal RSH may have arisen through lateral gene transfer  
17 rather than vertical descent from the ancestral chloroplast (Givens *et al.*, 2004; Atkinson *et al.*,  
18 2011; Ito *et al.*, 2017). A major reason for this proposition is the grouping of the RSH1 family  
19 with RSH from the prokaryotic *Deinococcus–Thermus* phylum rather than cyanobacteria. Our  
20 analysis supports this idea by also showing that the extended RSH1 family groups with RSH  
21 from the *Deinococcus–Thermus* phylum (Fig. 4). Furthermore, our discovery of new CASH RSH  
22 families that group with bacteria are evidence that additional lateral gene transfers from bacteria  
23 may have occurred recently. This would suggest that lateral transfers of RSH genes can occur  
24 readily.

25  
26 Our report sheds new light on the RSH enzyme family and (p)ppGpp metabolism in the  
27 diatoms, and reveals many surprises. Further research is now required to elucidate the mechanism  
28 and role of (p)ppGpp signalling in the lifestyle of this important and diverse group of  
29 photosynthetic eukaryotes.

30  
31 **Acknowledgements**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

This work was supported by the Agence Nationale de la Recherche (ANR-15-CE05-0021-03).

#### **Author Contributions**

A.L, B.M, A.V, E.B, B.F and B.G conceived and planned the experiments. A.L, C.P, A.V and B.F performed the experiments. A.L, C.P, A.V, E.B, B.F and B.G contributed to the interpretation of the results. B.F, A.L and B.G wrote the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

#### **Supplementary data**

**Fig. S1** Full amino acid sequences of the three RSH proteins found in *P. tricornutum*.

**Fig. S2** Alignment of *P. tricornutum* RSH proteins.

**Fig. S3** Phylogenetic tree of RSH1 proteins from different diatoms and other photosynthetic eukaryotes.

**Table S1** List of the primers used in this study.

**Table S2** RSH EST data.

**Table S3** Accession numbers for sequences used in Fig. 4.

**Table S4** Domains found in RSH enzymes.

**Supporting Dataset 1** Alignment of RSH enzymes sequences (hydrolase and synthetase domains) in fasta format.

**Supporting Dataset 2** Alignment of RSH1 enzymes (whole sequence) in fasta format.

1 **Figure legends**

2 **Fig. 1 Primary structure of RSH from *P. tricornutum*.** (A) Schematic representation of  
3 domains found in *P. tricornutum* PtRSH1, PtRSH4a and PtRSH4b (JGI gene accession numbers  
4 11099, 7629 and 33947): (p)ppGpp hydrolase (HYD), (p)ppGpp synthetase (SYN), Threonyl-  
5 tRNA synthetase GTPase Spot (TGS) and Aspartate kinase, Chorismate mutase, TyrA (ACT).  
6 The bipartite peptide for targeting to the chloroplast is shown. Arrows indicate the cleavage site  
7 (AF) of the putative chloroplast signal peptide and the intron position (I) on the corresponding  
8 gene. (B) Sequence surrounding the cleavage site (AF) of the putative chloroplast signal peptides  
9 of *P. tricornutum* RSH. Bold residues indicate greater than 0.25 frequency of amino acid  
10 occurrence (Gruber *et al.*, 2015).

11  
12 **Fig. 2 Catalytic activity of *P. tricornutum* RSH.** (A) The coding sequences for mature PtRSH1,  
13 PtRH4a and PtRSH4b were introduced into the pBAD24 plasmid in MG1655  $\Delta$ relA  $\Delta$ spoT. Cells  
14 were incubated at 37°C on M9 minimal agar medium without amino acids in the absence or  
15 presence of the inducer arabinose. Empty vector and pBAD24-Spot from *E. coli* were included as  
16 controls. Complementation was also observed without induction, presumably due to leaky  
17 expression from pBAD24. (B) Growth curves of MG1655  $\Delta$ relA spoT203 containing coding  
18 sequences for the indicated enzymes in the plasmid pBAD24. The values are means  $\pm$  SD of three  
19 biological replicates.

20  
21 **Fig. 3 Comparative modelling of the synthetase domain of *P. tricornutum* RSH.** (A) Amino  
22 acid alignments for the (p)ppGpp synthetase domain of the *B. subtilis* SAS1 and *P. tricornutum*,  
23 *A. thaliana* (At), *E. coli* (Ec), *Streptococcus equisimilis* (Rseq) RSH. The alignment was  
24 performed using the Muscle algorithm (<https://www.ebi.ac.uk/Tools/msa/muscle/>) with default  
25 parameters. Pink horizontal bars indicate important catalytic motifs for (p)ppGpp synthesis  
26 (Steinchen & Bange, 2016). Black, dark grey and light grey backgrounds indicate percentage  
27 identity between protein sequences (100, 80, 60%, respectively). (B) The structure of the *B.*  
28 *subtilis* SAS1 active site bound to AMCPP compared to the corresponding domain in modelled  
29 structures of RSH from *P. tricornutum*, *A. thaliana* RSH1 and *A. thaliana* RSH3. The SAS1  
30 conserved glycine at position 45 is shown in red, and other residues implicated in AMCPP  
31 binding are shown in black. These residues are also indicated by an asterisk in 3A. Non-

1 conserved residues are indicated in single letter code in the modelled RSH active sites.  
2 Structures were modelled using the I-Tasser Modelling Suite with standard settings (Yang *et al.*,  
3 2015).

4  
5 **Fig. 4 Phylogenetic tree of RSH proteins from diatoms and other photosynthetic organisms.**

6 The synthetase and hydrolase domains of different RSH proteins were used to infer phylogenetic  
7 relationships using maximum likelihood reconstructions. Diatom RSH are indicated by orange  
8 text and branches, CASH RSH are indicated by orange branches, red algal RSH are indicated by  
9 red branches, RSH from plants and green algae are condensed and indicated by green branches,  
10 cyanobacteria are condensed and indicated by cyan branches, and bacteria are condensed and  
11 indicated by black branches. The scale bar indicates the number of substitutions per sites.  
12 Statistical support for branches was estimated with the SH-like local support method. Only  
13 support values greater than 50% are shown. RSH groups are indicated to the right, and red or  
14 green circles at the nodes indicate the base of each major RSH clade Accession numbers and the  
15 alignment are provided in supplementary Table S3 and Supporting Dataset 1. Domain  
16 architecture is shown to the right of the phylogenetic tree. Domains were identified using the  
17 NCBI Conserved Domain (CD) algorithm. Domains: (p)ppGpp hydrolase (HYD), (p)ppGpp  
18 synthetase (SYN), Threonyl-tRNA synthetase GTPase Spot (TGS) and Aspartate kinase,  
19 Chorismate mutase, TyrA (ACT). Pale coloured HYD and SYN domains columns indicate a  
20 likely loss of activity due to the substitution of important residues. Other domains are indicated  
21 by numbers and listed in supplementary Table S4. A dot indicates that the domain is present in  
22 several members of a condensed clade.

23

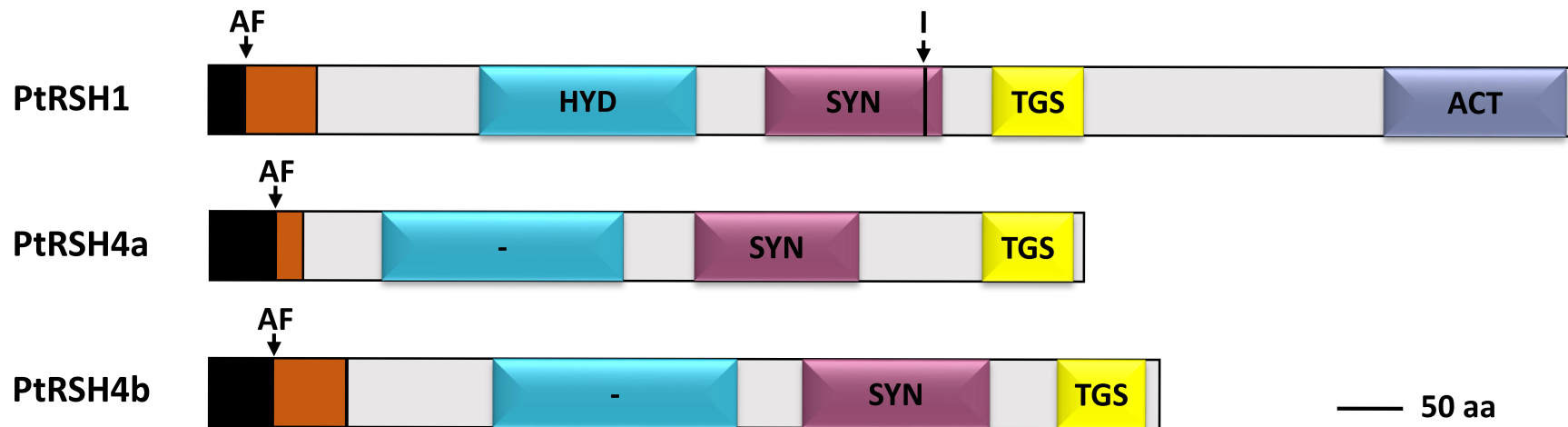
## References

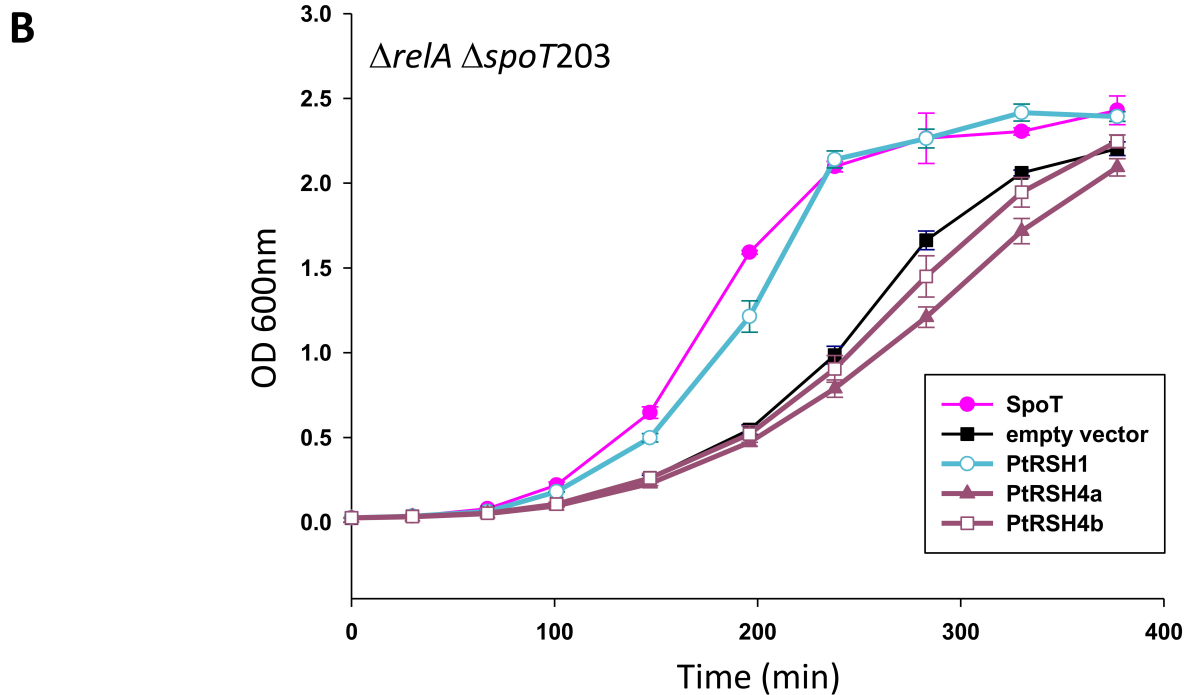
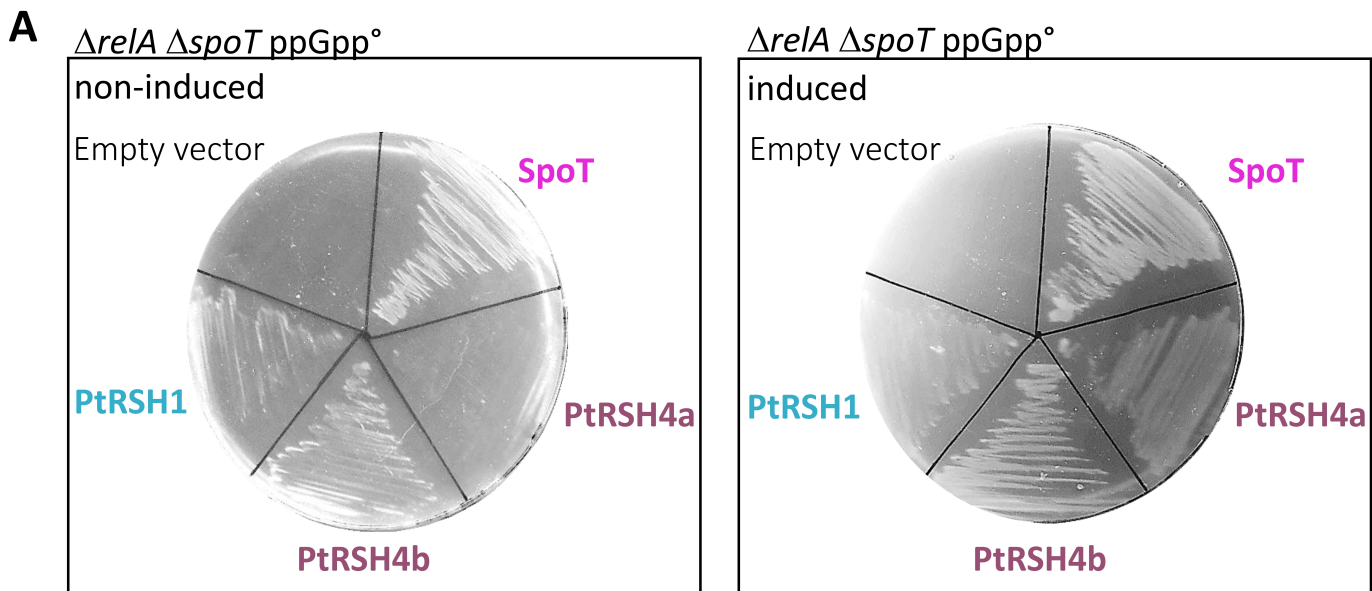
- Abdelkefi H, Sugliani M, Ke H, Harchouni S, Soubigou-Taconnat L, Citerne S, Mouille G, Fakhfakh H, Robaglia C, Field B. 2018.** Guanosine tetraphosphate modulates salicylic acid signalling and the resistance of *Arabidopsis thaliana* to Turnip mosaic virus. *Molecular Plant Pathology* **19**, 634-646.
- Archibald JM. 2015.** Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences* **112**, 10147-10153.
- Atkinson GC, Tenson T, Hauryliuk V. 2011.** The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS One* **6**, e23479.
- Battesti A, Bouveret E. 2006.** Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. *Mol Microbiol* **62**, 1048-1063.
- Boniecka J, Prusinska J, Dabrowska GB, Goc A. 2017.** Within and beyond the stringent response-RSH and (p)ppGpp in plants. *Planta* **246**, 817-842.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otiillar RP, et al. 2008.** The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239.
- Cashel M, Gallant J. 1969.** Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature* **221**, 838.
- Dalebroux ZD, Swanson MS. 2012.** ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol* **10**, 203-212.
- Dorrell RG, Gile G, McCallum G, Méheust R, Baptiste EP, Klinger CM, Brillet-Guéguen L, Freeman KD, Richter DJ, Bowler C. 2017.** Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife* **6**, e23717.
- Dorrell RG, Smith AG. 2011.** Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell* **10**, 856-868.
- Field B. 2018.** Green magic: regulation of the chloroplast stress response by (p)ppGpp in plants and algae. *J Exp Bot* **69**, 2797-2807.
- Givens RM, Lin MH, Taylor DJ, Mechold U, Berry JO, Hernandez VJ. 2004.** Inducible expression, enzymatic activity, and origin of higher plant homologues of bacterial RelA/SpoT stress proteins in *Nicotiana tabacum*. *J Biol Chem* **279**, 7495-7504.
- Gruber A, Rocop G, Kroth PG, Armbrust EV, Mock T. 2015.** Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J* **81**, 519-528.
- Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. 2015.** Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **13**, 298-309.
- Hogg T, Mechold U, Malke H, Cashel M, Hilgenfeld R. 2004.** Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response [corrected]. *Cell* **117**, 57-68.
- Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR. 2013.** Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* **8**, e74483.
- Ihara Y, Ohta H, Masuda S. 2015.** A highly sensitive quantification method for the accumulation of alarmones ppGpp in *Arabidopsis thaliana* using UPLC-ESI-qMS/MS. *J Plant Res* **128**, 511-518.
- Imamura S, Nomura Y, Takemura T, Pancha I, Taki K, Toguchi K, Tozawa Y, Tanaka K. 2018.** The checkpoint kinase TOR (target of rapamycin) regulates expression of a nuclear-encoded chloroplast RelA-SpoT homolog (RSH) and modulates chloroplast ribosomal RNA synthesis in a unicellular red alga. *The Plant Journal* **94**, 327-339.

- Ito D, Ihara Y, Nishihara H, Masuda S. 2017.** Phylogenetic analysis of proteins involved in the stringent response in plant cells. *J Plant Res* **130**, 625-634.
- Jensen E, Clément R, Maberly SC, Gontero B. 2017.** Regulation of the Calvin–Benson–Bassham cycle in the enigmatic diatoms: biochemical and evolutionary variations on an original theme. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**.
- Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, Seen DS, Kang SG. 2012.** One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl Environ Microbiol* **78**, 5440-5443.
- Katoh K, Rozewicki J, Yamada KD. 2017.** MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. DOI: 10.1093/bib/bbx108
- Kim T-H, Ok SH, Kim D, Suh S-C, Byun MO, Shin JS. 2009.** Molecular characterization of a biotic and abiotic stress resistance-related gene RelA/SpoT homologue (PepRSH) from pepper. *Plant Science* **176**, 635-642.
- Kojadinovic-Sirinelli M, Villain A, Puppo C, Fon Sing S, Prioretti L, Hubert P, Grégori G, Zhang Y, Jean-François S, Jean-Michel C, et al. 2018.** Exploring the microbiome of the “star” freshwater diatom *Asterionella formosa* in a laboratory context *Environmental Microbiology* **in press**.
- Maberly SC, Courcelle C, Groben R, Gontero B. 2010.** Phylogenetically-based variation in the regulation of the Calvin cycle enzymes, phosphoribulokinase and glyceraldehyde-3-phosphate dehydrogenase, in algae. *J Exp Bot* **61**, 735-745.
- Maekawa M, Honoki R, Ihara Y, Sato R, Oikawa A, Kanno Y, Ohta H, Seo M, Saito K, Masuda S. 2015.** Impact of the plastidial stringent response in plant growth and stress responses. *Nature Plants* **1**, 15167.
- Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, Wincker P, Iudicone D, de Vargas C, Bittner L, et al. 2016.** Insights into global diatom distribution and diversity in the world’s ocean. *Proc Natl Acad Sci USA* **113**, E1516-E1525.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2015.** CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222-226.
- Masuda S, Mizusawa K, Narisawa T, Tozawa Y, Ohta H, Takamiya K. 2008.** The bacterial stringent response, conserved in chloroplasts, controls plant fertilization. *Plant Cell Physiol* **49**, 135-141.
- Matthijs M, Fabris M, Broos S, Vyverman W, Goossens A. 2016.** Profiling of the early nitrogen stress response in the diatom *Phaeodactylum tricorutum* reveals a novel family of RING-domain transcription factors. *Plant Physiology* **170**, 489-498.
- Matthijs M, Fabris M, Obata T, Foubert I, Franco-Zorrilla JM, Solano R, Fernie AR, Vyverman W, Goossens A. 2017.** The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricorutum*. *EMBO J* **36**, 1559-1576
- Mizusawa K, Masuda S, Ohta H. 2008.** Expression profiling of four RelA/SpoT-like proteins, homologues of bacterial stringent factors, in *Arabidopsis thaliana*. *Planta* **228**, 553-562.
- My L, Rekoske B, Lemke JJ, Viala JP, Gourse RL, Bouveret E. 2013.** Transcription of the *Escherichia coli* fatty acid synthesis operon *fabH* is directly activated by FadR and inhibited by ppGpp. *J Bacteriol* **195**, 3784-3795.
- Petersen J, Ludwig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014.** Chromera velia, Endosymbioses and the Rhodoplex Hypothesis—Plastid Evolution in Cryptophytes, Alveolates, Stramenopiles, and Haptophytes (CASH Lineages). *Genome Biol Evol* **6**, 666-684.
- Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017.** An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology* **27**, 386-391.
- Price MN, Dehal PS, Arkin AP. 2010.** FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.

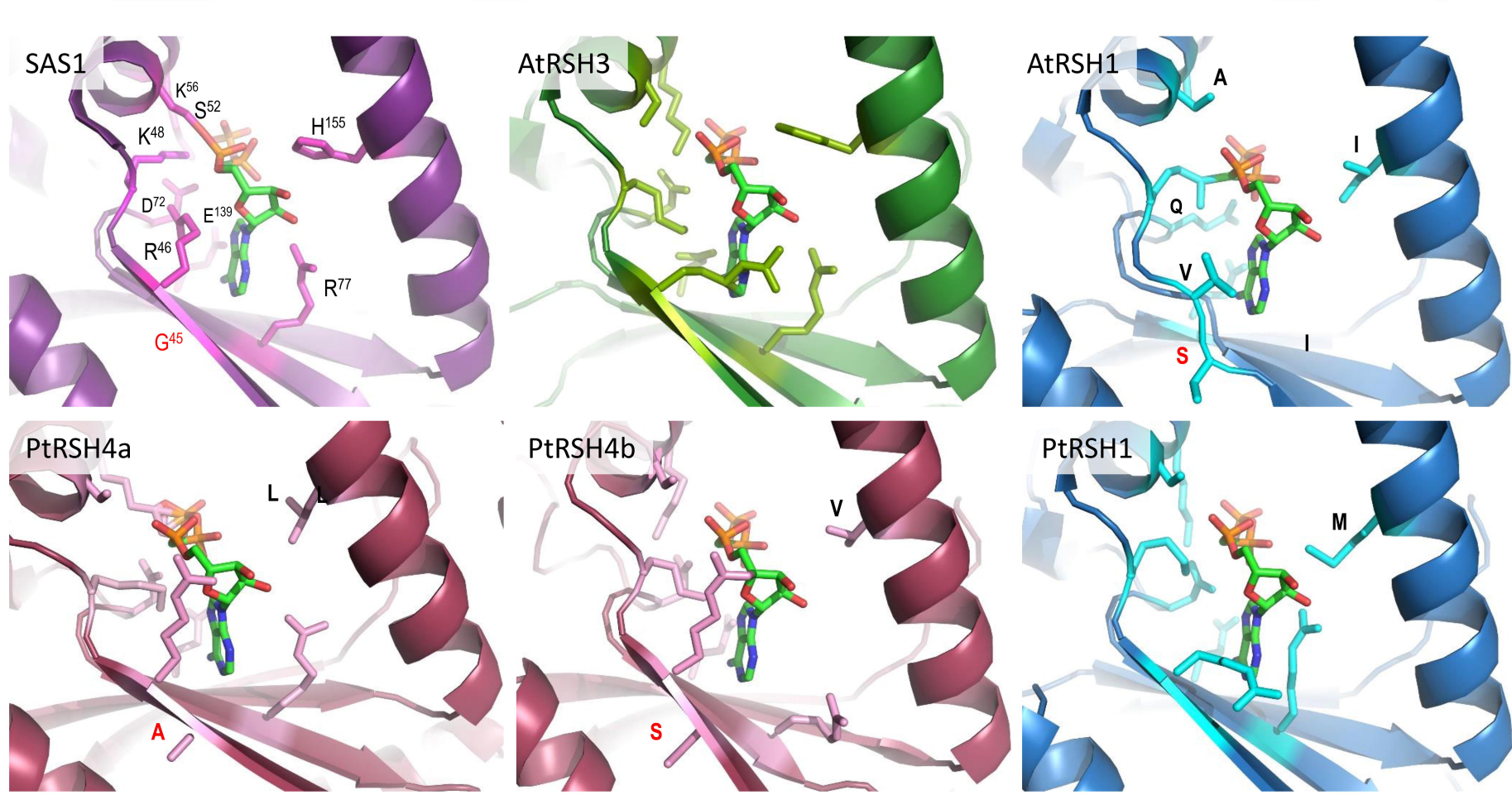
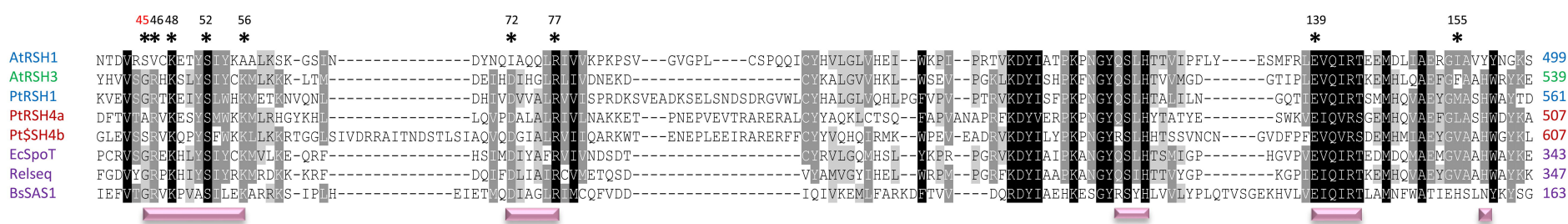
- Puppo C, Voisin T, Gontero B. 2017.** Genomic DNA extraction from the pennate diatom *Asterionella formosa* optimised for next generation sequencing. protocols.io. DOI: 10.17504/protocols.io.jytcpxwn
- Puthiyaveetil S, Kavanagh TA, Cain P, Sullivan JA, Newell CA, Gray JC, Robinson C, van der Giezen M, Rogers MB, Allen JF. 2008.** The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. Proc. Natl. Acad. Sci. USA **105**, 10061-10066.
- Ronneau S, Caballero-Montes J, Mayard A, Garcia-Pino A, Hallez R. 2018.** Regulation of (p)ppGpp hydrolysis by a conserved archetypal regulatory domain. bioRxiv. DOI: 10.1101/257592
- Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017.** Early photosynthetic eukaryotes inhabited low-salinity habitats. Proceedings of the National Academy of Sciences **114**, E7737-E7745.
- Sato M, Takahashi T, Ochi K, Matsuura H, Nabeta K, Takahashi K. 2015.** Overexpression of RelA/SpoT homologs, PpRSH2a and PpRSH2b, induces the growth suppression of the moss *Physcomitrella patens*. Biosci Biotechnol Biochem **79**, 36-44.
- Scala S, Carels N, Falciatore A, Chiusano ML, Bowler C. 2002.** Genome Properties of the Diatom *Phaeodactylum tricornutum*. Plant Physiology **129**, 993-1002.
- Steinchen W, Bange G. 2016.** The magic dance of the alarmones (p)ppGpp. Mol Microbiol **101**, 531-544.
- Steinchen W, Schuhmacher JS, Altegoer F, Fage CD, Srinivasan V, Linne U, Marahiel MA, Bange G. 2015.** Catalytic mechanism and allosteric regulation of an oligomeric (p)ppGpp synthetase by an alarmone. Proc Natl Acad Sci U S A **112**, 13348-13353.
- Sugliani M, Abdelkefi H, Ke H, Bouveret E, Robaglia C, Caffarri S, Field B. 2016.** An Ancient Bacterial Signaling Pathway Regulates Chloroplast Function to Influence Growth and Development in Arabidopsis. Plant Cell **28**, 661-679.
- Takahashi K, Kasai K, Ochi K. 2004.** Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants. Proc. Natl. Acad. Sci. USA **101**, 4320-4324.
- van der Biezen EA, Sun J, Coleman MJ, Bibb MJ, Jones JD. 2000.** Arabidopsis RelA/SpoT homologs implicate (p)ppGpp in plant signaling. Proc. Natl. Acad. Sci. USA **97**, 3747-3752.
- Vardi A, Formiggini F, Casotti R, De Martino A, Ribalet F, Miralto A, Bowler C. 2006.** A stress surveillance system based on calcium and nitric oxide in marine diatoms. PLoS Biol **4**, e60.
- Wahl A, My L, Dumoulin R, Sturgis JN, Bouveret E. 2011.** Antagonistic regulation of dgkA and plsB genes of phospholipid synthesis by multiple stress responses in *Escherichia coli*. Mol Microbiol **80**, 1260-1275.
- Wendrich TM, Marahiel MA. 1997.** Cloning and characterization of a relA/spoT homologue from *Bacillus subtilis*. Mol Microbiol **26**, 65-79.
- Wout P, Pu K, Sullivan SM, Reese V, Zhou S, Lin B, Maddock JR. 2004.** The *Escherichia coli* GTPase CgtAE cofractionates with the 50S ribosomal subunit and interacts with SpoT, a ppGpp synthetase/hydrolase. J Bacteriol **186**, 5249-5257.
- Yamburenko MV, Zubo YO, Borner T. 2015.** Abscisic acid affects transcription of chloroplast genes via protein phosphatase 2C-dependent activation of nuclear genes: repression by guanosine-3'-5'-bis(diphosphate) and activation by sigma factor 5. Plant J **82**, 1030-1041.
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015.** The I-TASSER Suite: protein structure and function prediction. Nat Methods **12**, 7-8.



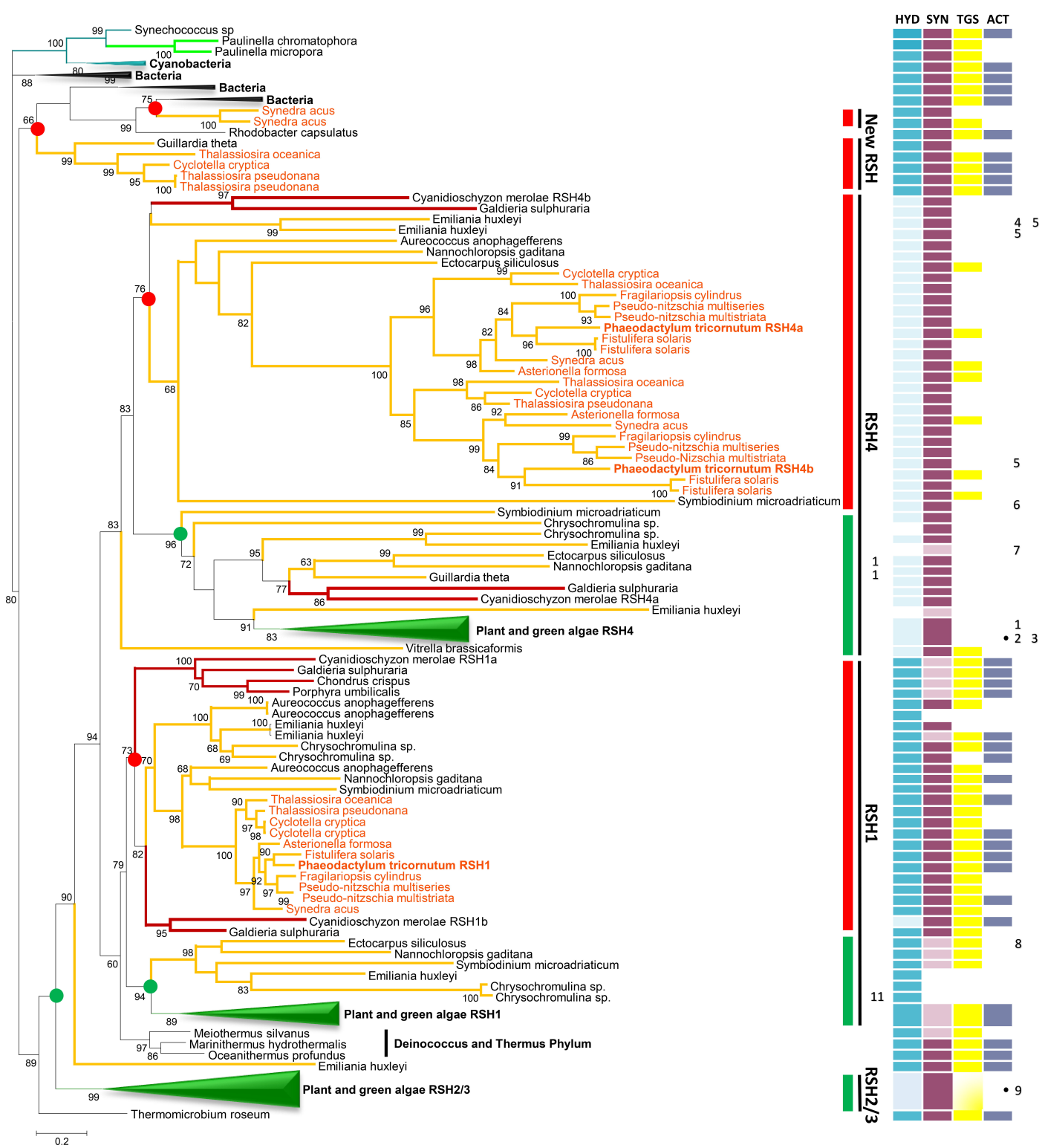
**A****B****Figure 1**



**Figure 2**



**Figure 3**



**Figure 4**