Early emergence of the FtsH proteases involved in photosystem II repair

S. SHAO, T. CARDONA, and P.J. NIXON+

Department of Life Sciences, Sir Ernst Chain Building – Wolfson Laboratories, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Abstract

Efficient degradation of damaged D1 during the repair of PSII is carried out by a set of dedicated FtsH proteases in the thylakoid membrane. Here we investigated whether the evolution of FtsH could hold clues to the origin of oxygenic photosynthesis. A phylogenetic analysis of over 6000 FtsH protease sequences revealed that there are three major groups of FtsH proteases originating from gene duplication events in the last common ancestor of bacteria, and that the FtsH proteases involved in PSII repair make a distinct clade branching out before the divergence of FtsH proteases found in all groups of anoxygenic phototrophic bacteria. Furthermore, we showed that the phylogenetic tree of FtsH proteases in phototrophic bacteria is similar to that for Type I and Type II reaction centre proteins. We conclude that the phylogeny of FtsH proteases is consistent with an early origin of water oxidation chemistry.

Additional key words: AAA+ protease; chloroplast; cyanobacteria; evolution; photoprotection; water oxidation.

Introduction

Oxygenic photosynthetic electron transport from water to $NADP^+$ requires the participation of two functionally distinct reaction centers (RCs) acting in series: photosystem II (PSII, a Type II RC containing quinone electron acceptors) and photosystem I (PSI, a Type I RC containing redox-active iron-sulphur clusters). Early ideas on for the emergence of oxygenic photosynthesis focused on the evolution of PSI and PSII from pre-existing RCs found in anoxygenic photosynthetic bacteria (Nitschke and Rutherford 1991) and the horizontal transfer of genes encoding chlorophyll biosynthetic enzymes and RC proteins (Hohmann-Marriott and Blankenship 2011, Fischer *et al.* 2016). Because existing anoxygenic photosynthetic bacteria contain just one type of RC, such 'gene acquisition' hypotheses require at least two bacterial ancestors: one providing a Type II RC and another a Type I RC.

In contrast, more recent phylogenetic analyses have led to the suggestion that the evolution of Type I and Type II RCs might have occurred in a single organism after a gene duplication event (Mulkidjanian *et al.* 2006, Sousa *et al.* 2013, Harel *et al.* 2015) and, possibly, that Type I and Type II RCs might have then been transferred at various stages of evolution to other types of nonphotosynthetic bacteria through horizontal gene transfer (HGT) (Mulkidjanian *et al.* 2006). Given the fact that the geochemical record of photosynthesis dates back to 3.5 to 3.8 billion years ago, such 'gene duplication' hypotheses raise the possibility that oxygenic photosynthesis might have evolved much earlier than previously thought, perhaps hundreds of millions of years before the Great Oxidation Event (Lyons *et al.* 2014).

A hallmark of oxygenic photosynthesis is the presence of photoprotective mechanisms to prevent or repair light-induced damage to PSII (Takahashi and Badger 2011). Given the known importance of thylakoid-embedded FtsH proteases for PSII repair in cyanobacteria (Silva *et al.* 2003, Komenda *et al.* 2006) and chloroplasts (Bailey *et al.* 2002, Zaltsman *et al.* 2005, Kato and Sakamoto 2009), we hypothesised that analysis of the evolution of FtsH could provide relevant information regarding the origin of oxygenic photosynthesis. Early phylogenetic attempts (Sakamoto *et al.* 2003, Yu *et al.* 2004, Yu 2005) using a limited sequence dataset suggested that the FtsH subunits required for PSII repair exist in two main forms, denoted Type A and Type B (Zaltsman *et al.* 2005). However, the evolutionary relationship between these FtsH subunits and the others found in nature remains poorly understood.

Received 8 June 2017, *accepted* 1 September 2017.
+Corresponding author; e-mail: p.nixon@imperial.ac.uk
Abbreviations: AAA+ – ATPase associated with diverse activities; HGT – horizontal gene transfer; RC – reaction center; ROS – reactive oxygen species.
Acknowledgements: SS was supported by a CSC Imperial Scholarship, and TC (grant BB/P00931X/1) and PJN (grants BB/L003260/1 and BB/N016807/1) by the Biotechnology and Biological Sciences Research Council.

Here we show that the FtsH subunits involved in PSII repair diverge early and seem to have branched out before those present in anoxygenic photosynthetic bacteria. Our results therefore support an early origin for oxygenic photosynthesis consistent with the 'gene duplication' hypothesis.

Materials and methods

Construction of FtsH sequence datasets: Protein sequences (6427) containing the M41 peptidase domain were retrieved from the *Pfam 30.0* database (Finn *et al.* 2016) under the entry PF01434 on 14 October 2016. Sequences lacking the AAA+ domain, under entry PF00004, were removed using the HMMER tool (Eddy 2011), yielding in total 6082 sequences belonging to 73 bacterial phyla and candidate phyla and 378 eukaryote species. It is noteworthy that the *Pfam* database is based on the manually and algorithmically curated *UniProt Reference Proteomes* database (www.uniprot.org/proteomes/), in which annotation errors, although rare, do exist. In addition, the *Pfam* database as well as the HMMER method is not sensitive to convergent evolution at the molecular level (Zakon 2002, Mistry *et al.* 2013), therefore compositional bias cannot be excluded. Other than the above limitations, this dataset covered over 3,100 genome-sequenced species spanning the domain Bacteria and Eukarya. Overall, this dataset was considered as a comprehensive representation of FtsH homologs in the context of the tree of life. The number of FtsH genes per genome was counted using a homemade python script depending on *Matplotlib 2.0.1*. The number of FtsH in Chlorobi and Chloroflexi was updated on 29 April 2017 by interrogating the *UniProt* database. For the comprehensive dataset of cyanobacterial FtsH, a copy of 103 cyanobacteria reference proteomes was interrogated and downloaded from the *UniProt Proteomes* database on 10 October 2016. In total 417 sequences were retrieved for cyanobacteria by searching for the AAA+ and M41 protease domains.

Sequence alignment and phylogenetic analysis: Sequences of the 6082 FtsH dataset were aligned using *MAFFT version 7* programme with the "L-INS-I" setting applied (Yamada *et al.* 2016). Gaps within the alignment were then removed by *TrimAl* (Capella-Gutiérrez *et al.* 2009) tool using the "gappyout" strategy, 570 characters were retained after the trimming process. Sequences of the 417 cyanobacterial FtsH dataset were similarly processed and 611 characters were retained in the final alignment. Phylogeny inference of the 6082 sequences dataset was carried out using the *CIPRES Science Gateway* supercomputer server (https://www.phylo.org/). *FastTree* programme, an approximate maximum likelihood method, was adopted using default setting for the inference. *FastTree* method computes the local support values with the Shimodaira-Hasegawa test (Shimodaira and Hasegawa 1999). Phylogenetic position of AtFtsHi3 and AtFtsHi4 were examined separately by the same methodology. Phylogeny of the 417 cyanobacterial FtsH dataset was inferred through ETE3 toolkit using the PhyML method (Guindon *et al.* 2010, Yamada *et al.* 2016); the applied amino acid substitution model was JTT/GTR (Jones-Taylor-Thornton/Generalised Time Reversible) and branch support adopted aLRT (approximate likelihood ratio test). The resulting unrooted trees were organised and beautified with iTOL (Letunic and Bork 2016). In addition, 312 sequences from phototrophic bacteria and from the Candidate Phyla Radiation (CPR) were aligned, trimmed and submitted to *PhyML 3.0* (Guindon *et al.* 2010) for phylogenetic inference.

Structural conservation analysis: FtsH sequences (270) were manually selected from the *UniProt* database covering 55 bacterial phyla or candidate phyla based on a recently inferred tree of life (Hug *et al.* 2016). Similar to the cyanobacterial FtsH dataset, this dataset was aligned, trimmed and inferred through ETE3 toolkit using *MAFFT*, *TrimAl* and *PhyML*. The resulting phylogenetic tree was submitted to *ConSurf* server for the structural conservation analysis. The ADP-bound *Thermus thermophilus* FtsH (G399L mutant; PDB code 2DHR) and the *apo*-state of *Thermotoga maritima* FtsH (K207A, K410L, and K415A triple mutant; PDB code 3KDS) were used as structural models for the *Consurf* analysis.

Results

Early diversification of FtsH proteases: Members of the membrane-embedded FtsH protease family play a role in the degradation of both soluble and membrane proteins and have been studied in bacteria, chloroplasts, and mitochondria (Smakowska *et al.* 2014, Nishimura *et al.* 2016, Bittner *et al.* 2017). Structural analyses have shown that FtsH forms hexameric complexes composed of either one or two types of FtsH protomer (Nishimura *et al.* 2016). The archetypal FtsH subunit contains an N-terminal membrane-spanning domain, consisting of one or two transmembrane regions, attached to a soluble fragment consisting of an AAA+ (ATPase associated with diverse activities) module and a highly characteristic C-terminal M41 protease domain containing a bound zinc ion (Tomoyasu *et al.* 1993, Leonhard *et al.* 1996).

We constructed a comprehensive dataset comprising

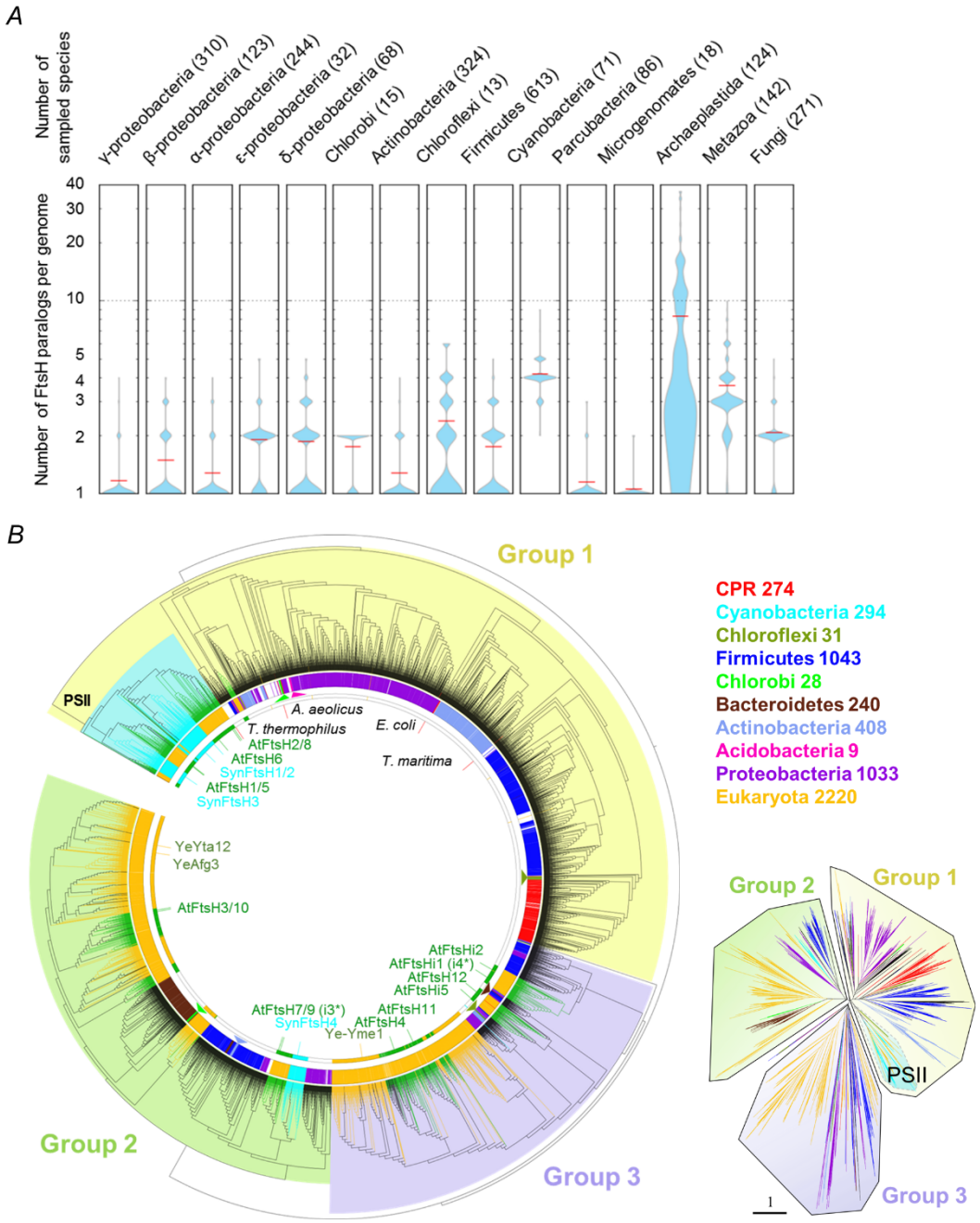

Fig. 1. Multiplicity and evolution of FtsH proteases. *A*: FtsH copy number in proteomes from major bacteria phyla or classes. Red lines indicate the mean value of the number of FtsH per genome and the light blue curve represents the frequency of a given number of FtsH per genome per clade. *B*: Phylogeny of the FtsH family. Left panel shows a circular cladogram and the right corner shows the unrooted tree. The three proposed groups of orthologs are coloured accordingly; the branch containing well-characterised FtsH proteases responsible for PSII repair in *Arabidopsis* and *Synechocystis* is also highlighted (PSII). Major taxonomic groups are colour-coded in the thick inner circle. Eukaryotic and cyanobacterial FtsH proteases are showed in the thin inner circle to emphasise photosynthetic eukaryotes and Cyanobacteria. Representative FtsH proteases from yeast (Ye), *Arabidopsis* (At), and *Synechocystis* (Syn) are colour-labelled, positions of AtFtsHi3 and AtFtsHi4 were confirmed in a separate analysis. Four FtsH proteases with structural information available are labelled in red line and species name.

6082 FtsH homologs from over 3100 genome-sequenced species by screening the *Pfam* database for proteins containing both the AAA+ and M41 protease domains. Amongst the prokaryotes, cyanobacteria showed the highest number of FtsH homologs with approximately 4 FtsH homologs per sequenced genome (Fig. 1*A*). Other

prokaryote groups also contained multiples copies of FtsH, but in most cases not greater than 3. In the case of eukaryotes, photosynthetic organisms display the highest multiplicity, with more than 8 FtsH subunits per genome, while fungi and animals contained on average about 2 and 4 FtsH copies, respectively (Fig. 1*A*).

As illustrated in Fig. 1*B*, a phylogenetic analysis of the FtsH sequences (*see* Materials and methods) resolved three potential orthologous groups denoted here as Group 1, Group 2, and Group 3. Each group reproduced a branch topology that was consistent with previous phylogenetic and phylogenomic studies on the diversification of bacteria (Ciccarelli *et al.* 2006, Jun *et al.* 2010, Segata *et al.* 2013) and roughly in line with the Hydrobacteria-type (*e.g.* Proteobacteraia, Acidobacteria, Chlorobi, *etc.*) and Terrabacteria-type divisions (*e.g.* Actinobacteria, Chloroflexi, Firmicutes, *etc.*) proposed by Battistuzzi and coworkers (Battistuzzi and Hedges 2009, Marin *et al.* 2017). These data suggest that FtsH proteases were likely to be present in the last common ancestor of the domain Bacteria and that the three groups may have originated from ancestral gene duplication events early during the evolution of bacteria. Group 1 FtsH showed the greatest diversity and was almost universally found across bacteria, with strains containing a single FtsH protease most likely to have retained a Group 1 FtsH.

Notably, within Group 1, we detected an early diverging, clearly resolved branch (Shimodaira–Hasegawa–like local support 0.983, *see* Materials and methods) containing FtsH sequences exclusively found in cyanobacteria and in photosynthetic eukaryotes. These included SynFtsH2/3 from *Synechocystis* sp. PCC 6803 and AtFtsH1/2/5/8 from *Arabidopsis thaliana,* all known to be involved in PSII repair (Komenda *et al.* 2012, Nishimura *et al.* 2016). The origin of this group seems to predate the radiation of FtsH orthologs from other bacteria phyla, including those of anoxygenic phototrophic bacteria within the same group (Fig. 2), for instance *Chlorobium tepidum* (Chlorobi), *Chloracidobacterium thermophilum* (Acidobacteria), and *Heliobacterium modesticaldum* (Firmicutes) containing Type I RCs and *Rhodobacter sphaeroides* (Proteobacteria) and *Chloroflexus aurantiacus* (Chloroflexi) containing

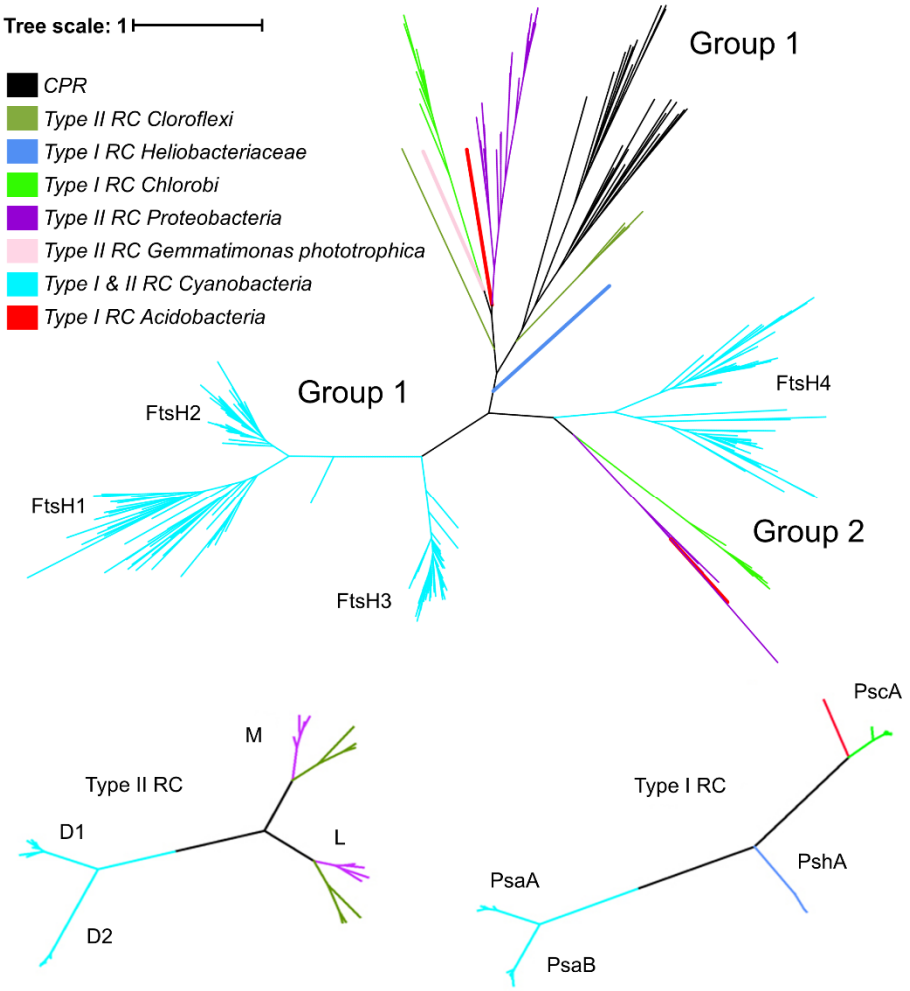

Fig. 2. Phylogenetic tree of FtsH proteases from phototrophic groups (top) compared to those from Type II (bottom left) and Type I (bottom right) reaction centre proteins, adapted from Cardona (2015). The bar indicates amino-acid substitutions per site.

Type II RCs. This early divergence was not found for the other FtsH proteases of cyanobacteria and photosynthetic eukaryotes required for PSII repair, such as AtFtsH7/9 and SynFtsH4 located in Group 2.

Most eukaryotic FtsH sequences were located in Group 2 and Group 3, with the exception of photosynthetic eukaryotes, which additionally contained Group 1 FtsH acquired from the primary cyanobacterial endosymbiont.

Fig. 2 shows a maximum likelihood phylogenetic tree calculated using the FtsH proteases of known phototrophic bacteria. On the left, the FtsH involved in PSII repair make a monophyletic group. The genome of *Heliobacterium modesticaldum*, encoding a single FtsH, branched basally among other Group 1 FtsH proteases of anoxygenic phototrophic bacteria. Proteobacteria, Acidobacteria, and Chlorobi had both Group 1 and Group 2 FtsH proteases. In both Group 1 and Group 2 the sequence from the phototrophic acidobacterium, *Chloracidobacterium thermophilum* branched out as a sister group to the Proteobacteria, and the Chlorobi sequences branched prior to the Acidobacteria and Proteobacteria split. This pattern is also observed in Fig. 1 which also contained non-phototrophic representatives of the same phyla and it is consistent with previous phylogenetic (Quaiser *et al.* 2003, Bryant *et al.* 2012, Sousa *et al.* 2013, Greening *et al.* 2015, Cardona 2016a) and phylogenomic (Ciccarelli 2006, Dutilh *et al.* 2008, Wu and Eisen 2008, Ward *et al.* 2009, Jun *et al.* 2010, David and Alm 2011, Rinke *et al.* 2013, Segata *et al.* 2013, Marin *et al.* 2017) studies that have repeatedly confirmed the Acidobacteria as a sister clade of the Proteobacteria, or branching within the Proteobacteria as a sister clade of the Deltaproteobacteria. At the same time, the nearness of the Chlorobi to Acidobacteria and Proteobacteria is also consistent with the Chlorobi–Bacteroidetes–Fibrobacterere supergroup bifurcating prior to the radiation of the Proteobacteria (Wu and Eisen 2008, Jun *et al.* 2010, David and Alm 2011, Segata *et al.* 2013, Marin *et al.* 2017). Similarly, the phylogenetic proximity of the Gemmatimonadetes to the Chlorobi has also been demonstrated (Segata *et al.* 2013, Zeng *et al.* 2014), which is consistent with this group obtaining Type II reaction centres *via* horizontal gene transfer from the Proteobacteria (Zeng *et al.* 2014).

When the phylogeny of FtsH from phototrophic bacteria was compared to the phylogeny of Type I and Type II RC proteins (Beanland 1990, Bryant *et al.* 2007) (Fig. 2), we observed that the trees for the RC proteins follow an identical topology to that of the FtsH sequences. For example, cyanobacterial D1 and D2 subunits branch out before the divergence of L and M subunits of the anoxygenic Type II RCs of the Chloroflexi and Proteobacteria. Similarly, FtsH proteases involved in PSII repair (SynFtsH2/3) branch out before the divergence of Group 1 FtsH in Chloroflexi and Proteobacteria. Cyanobacterial PsaA and PsaB, the core subunits of PSI, also branch out before the divergence of PshA and PscA of anoxygenic homodimeric Type I RC proteins of Heliobacteria, Acidobacteria, and Chlorobi (Cardona 2015, 2016b), with heliobacterial PshA branching out before PscA. This is also mirrored in the tree of Group 1 FtsH proteases of phototrophs, with FtsH sequences involved in PSII repair branching out before those present in anoxygenic phototrophs containing Type I RCs. In Group 1 FtsH the heliobacterial sequence also diverged before the split of Acidobacteria and Chlorobi.

The significance of this is that, with the notable exception of the Gemmatimonadetes phototrophic bacteria, which seem to have acquired Type II RCs by HGT from Gammaproteobacteria, the dominant mode of evolution of FtsH proteases and RC proteins has been by vertical descent, with horizontal gene transfer being a secondary diversification force. It also suggests that the origin of oxygenic photosynthesis before the last common ancestor of cyanobacteria may have had a direct impact on the evolution of FtsH. Thus, the ancestral oxygenic Type II RC proteins might have co-evolved with ancestral Group 1 FtsH proteases from the origin of oxygenic photosynthesis and through the evolution of cyanobacteria and the diversification of photosynthetic eukaryotes. Our result is also in agreement with a birth-and-death model of protein evolution in which new proteins evolve by repeated gene duplication events, but some of the duplicated genes are maintained in the genome for long periods of time while others are eventually lost (Nei *et al.* 1997, Nei and Rooney 2005).

A peculiar case of horizontal gene transfer was noted for the FtsH within the phylum Bacteroidetes (Fig. 1). In this phylum, the dominant FtsH clustered within Group 2 FtsH proteases from early branching eukaryotes. This may indicate that it was gained from an early symbiotic association between ancestral population of eukaryotes and an ancestral strain of Bacteroidetes.

We also attempted to retrieve FtsH protease sequences from the Archaea. This was done by searching for FtsH homologs in 210 *Archaea* reference proteomes. We found a single FtsH sequence with accession number A0A0M0BK70 which was labelled “miscellaneous Crenarchaeota group archaeon SMTZ-80”. However, this FtsH has high similarity with an FtsH protease (H1XNZ9) from the bacterium *Caldithrix abyssi* DSM 13497 and therefore it is another likely case of HGT from bacteria to archaea.

Classification of cyanobacterial FtsH paralogs: Previous experimental data have shown that in *Synechocystis* 6803 degradation of D1 during PSII repair is mainly carried out by a thylakoid-embedded FtsH heterocomplex composed of FtsH2 and FtsH3 (Silva *et al.* 2003, Komenda *et al.* 2006, 2010; Cheregi *et al.* 2007, Boehm *et al.* 2012). The FtsH2/3 complex is not restricted to PSII repair and participates in the removal of unassembled membrane proteins (Komenda *et al.* 2006) as well as soluble proteins (Stirnberg *et al*. 2007). Also present in *Synechocystis* 6803 are FtsH1/FtsH3 heterocomplexes located in the cytoplasmic membrane (Krynická *et al.* 2014) and FtsH4 homo-complexes in the thylakoid and possibly

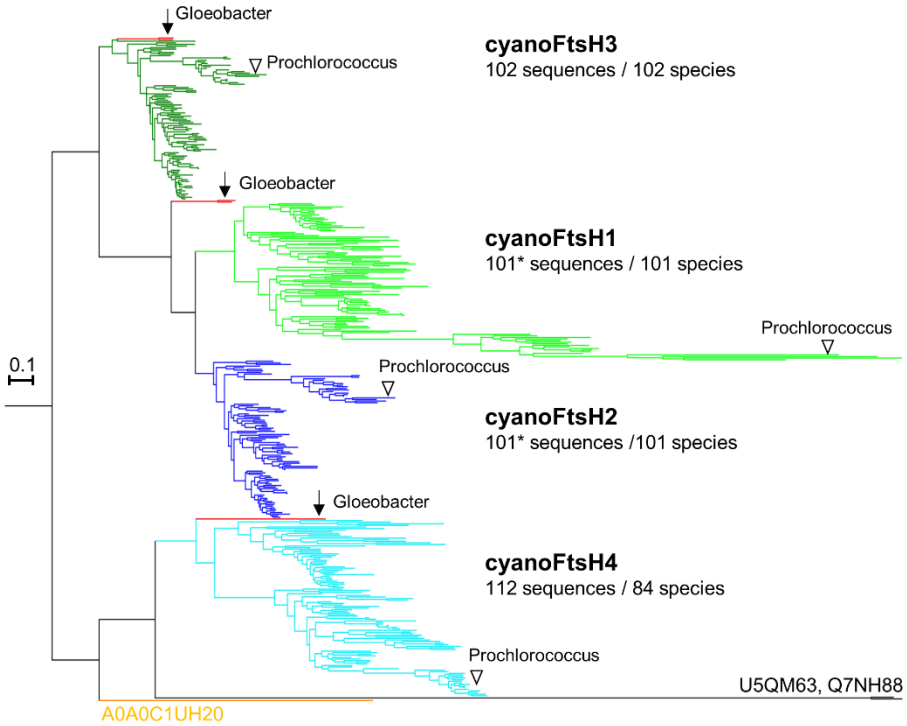

Fig. 3. Unrooted phylogenetic tree of cyanobacterial FtsH. *Black arrows* point to FtsH sequences from *Gloeobacter* spp. *Open triangles* point to FtsH from *Prochlorococcus* spp. Uncommon branches are labelled by the *UniProt Entry* of the protein (*see* the main text for details).

cytoplasmic membrane (Boehm *et al.* 2012, Sacharz *et al.* 2015). Both FtsH1 and FtsH3 are crucial for cell viability whereas FtsH2 and FtsH4 are dispensable (Mann *et al.* 2000), although growth of mutants lacking FtsH2 is extremely light-sensitive (Silva *et al.* 2003). In the case of *Arabidopsis*, at least one type A and one type B FtsH subunit are needed for growth (Zaltsman *et al.* 2005).

To assess the diversity of FtsH in cyanobacteria, we performed a phylogenetic analysis of all FtsH proteins found in 103 sequenced strains of cyanobacteria. Four distinctive groups were clearly resolved, which we designate cyanoFtsH1/2/3/4 based on the nomenclature used for *Synechocystis* 6803 (Fig. 3).

Fig. 3 shows that the last common ancestor of extant cyanobacteria had three FtsH paralogs: one ancestral to cyanoFtsH1 and cyanoFtsH2; a second one which was ancestral to cyanoFtsH3; and a third one which was ancestral to cyanoFtsH4. Fig. 3 also reveals that the duplication event that gave rise to cyanoFtsH1 and cyanoFtsH2 occurred after the divergence of the genus *Gloeobacter*, which lacks thylakoid membranes and contains the photosynthetic apparatus in the cytoplasmic membrane (Rexroth *et al.* 2011). This would suggest that the emergence of cyanoFtsH2, and the FtsH2/H3 complex involved in PSII repair is linked to the development of the thylakoid membrane system.

Each of the four groups is consistent with the known diversification of cyanobacteria, featuring the early branch of *Gloeobacter* (arrow) and long branches for the relatively more rapidly evolving clades of the marine *Synechococcus* and *Prochlorococcus* (open triangle) (Dvořák *et al.* 2011, Bombar *et al.* 2014, Komárek *et al.* 2014). The substitution rates (branch lengths) of different groups are noticeably different, with cyanoFtsH3 displaying the slowest rate and cyanoFtsH4 the highest, suggesting different evolutionary pressures upon them.

Species or strains missing one or more of the four types of FtsH are listed in Table 1; however, some assignments remain tentative until the genomes are fully sequenced. Overall, the majority (>99%) of cyanobacteria have a set of cyanoFtsH1/2/3 indicating the important role of these three types. The only species lacking cyanoFtsH3 is *Crocosphaera watsonii WH 8501*; however, we found in this organism a sequence (*UniProt* accession number: Q4BUC6) orthologous to cyanoFtsH3, but lacking the M41 peptidase. An unusual species is the recently described and sequenced *Neosynechococcus sphagnicola* (Dvořák *et al.* 2014), which apparently only contains cyanoFtsH3. Although it was isolated from an unusual environment, a peat bog (Dvořák *et al.* 2011), other peat bog cyanobacteria, such as *Synechococcus* sp. PCC 7502, *Pseudanabaena* sp. PCC 7429, and *Gloeocapsa* sp. PCC 73106 (Dvořák *et al.* 2014), all possess the expected complement of four FtsH paralogs. In addition, *Synechococcus sp.* JA-2-3B'a(2-13), isolated from hot spring microbial mat (Bhaya *et al.* 2007) was found to lack cyanoFtsH2.

Table 1. List of cyanobacteria that might have lost one or more FtsH during evolution. *UniProt* entries of proteins assigned to cyanoFtsH1/2/3 are indicated. Species with completely sequenced genomes are indicated *in bold*. *Gloeobacter kilaueensis* JS1 and *Gloeobacter violaceus* strain PCC 7421 are excluded due to their unique evolution (Fig. 3).

Species	cyanoFtsH1	cyanoFtsH2	cyanoFtsH3	cyanoFtsH4	Notes
Neosynechococcus sphagnicola	-	-	A0A098TRH5	-	peat bog (Dvořák *et al.* 2014)
Synechococcus sp. strain JA-2-3B'a(2-13)	Q2JHR8	-	Q2JNP0	-	hot spring microbial mat (Bhaya *et al.* 2007)
Aliterella atlantica CENA595	A0A0D8ZU81	A0A0D8ZT40	A0A0D8ZWS3	-	
Anabaena sp. 90	K7WSA3	K7WS23	K7VZY8	-	
Atelocyanobacterium thalassa (isolate ALOHA)	D3ENM5	D3EPJ8	D3EQB0	-	endosymbiont
Candidatus Synechococcus spongiarum 142	A0A0U1QKZ0	A0A0G8B1F9	A0A0G8AVK4	-	
Candidatus Synechococcus spongiarum SP3	-	A0A0G2HLH7	A0A0G2J4Y2	A0A0G2IVX6, K9UQL4	
Chrysosporum ovalisporum	A0A0P1C082	A0A0P1BUU7	A0A0P1BVD0		
Crocosphaera watsonii WH 8501	Q4C3U9	Q4BUM7	-	Q4BY73	
Cyanobacterium aponinum (strain PCC 10605)	K9Z414	K9Z622	K9Z6W1	-	
Cyanobacterium endosymbiont of Epithemia turgida isolate EtSB Lake Yunoko	A0A077JFW7	A0A077JK69	A0A077JIP6	-	endosymbiont
Cyanobacterium stanieri strain PCC 7202	K9YKE4	K9YIN6	K9YQT6	-	
Geitlerinema sp. PCC 7407	K9SC27	K9S5X2	K9SA34	-	
Geminocystis sp. NIES-3708	A0A0D6AE16	A0A0D6ABU5	A0A0D6AGY3	-	
Oscillatoria acuminata PCC 6304	K9TDN1	K9TGJ5	K9TBZ2	-	
Phormidium sp. OSCR	A0A0P8BT41	A0A0P8C7Y9	A0A0P8BW82	-	microbial mat (Cole *et al.* 2014)
Richelia intracellularis	X5JUF5	X5JFW6	X5JRC0	-	extracellular symbiont
Richelia intracellularis HH01	M1X0E5	M1X2X0	M1WZS3	-	Same as above
Richelia intracellularis HM01	M1WNU4	M1WNE7	M1WPH5	-	Same as above
Rubidibacter lacunae KORDI 51-2	U5DKP8	U5DJI8	U5DKU6	-	
Species	cyanoFtsH1	cyanoFtsH2	cyanoFtsH3	cyanoFtsH4	Notes
Neosynechococcus sphagnicola	-	-	A0A098TRH5	-	peat bog (Dvořák *et al.* 2014)
Synechococcus sp. strain JA-2-3B'a(2-13)	Q2JHR8	-	Q2JNP0	-	hot spring microbial mat (Bhaya *et al.* 2007)
Aliterella atlantica CENA595	A0A0D8ZU81	A0A0D8ZT40	A0A0D8ZWS3	-	
Anabaena sp. 90	K7WSA3	K7WS23	K7VZY8	-	
Atelocyanobacterium thalassa (isolate ALOHA)	D3ENM5	D3EPJ8	D3EQB0	-	endosymbiont
Candidatus Synechococcus spongiarum 142	A0A0U1QKZ0	A0A0G8B1F9	A0A0G8AVK4	-	
Candidatus Synechococcus spongiarum SP3	-	A0A0G2HLH7	A0A0G2J4Y2	A0A0G2IVX6, K9UQL4	
Chrysosporum ovalisporum	A0A0P1C082	A0A0P1BUU7	A0A0P1BVD0	-	
Crocosphaera watsonii WH 8501	Q4C3U9	Q4BUM7	-	Q4BY73	
Cyanobacterium aponinum (strain PCC 10605)	K9Z414	K9Z622	K9Z6W1	-	
Cyanobacterium endosymbiont of Epithemia turgida isolate EtSB Lake Yunoko	A0A077JFW7	A0A077JK69	A0A077JIP6	-	endosymbiont
Cyanobacterium stanieri strain PCC 7202	K9YKE4	K9YIN6	K9YQT6	-	
Geitlerinema sp. PCC 7407	K9SC27	K9S5X2	K9SA34	-	
Geminocystis sp. NIES-3708	A0A0D6AE16	A0A0D6ABU5	A0A0D6AGY3	-	
Oscillatoria acuminata PCC 6304	K9TDN1	K9TGJ5	K9TBZ2	-	
Phormidium sp. OSCR	A0A0P8BT41	A0A0P8C7Y9	A0A0P8BW82	-	microbial mat (Cole *et al.* 2014)
Richelia intracellularis	X5JUF5	X5JFW6	X5JRC0	-	extracellular symbiont
Richelia intracellularis HH01	M1X0E5	M1X2X0	M1WZS3	-	Same as above
Richelia intracellularis HM01	M1WNU4	M1WNE7	M1WPH5	-	Same as above
Rubidibacter lacunae KORDI 51-2	U5DKP8	U5DJI8	U5DKU6	-	

Among those strains lacking the full complement of four FtsH paralogs, the most common version, which was lost, was cyanoFtsH4. This is in line with previous mutagenesis studies in *Synechocystis* 6803 that have shown that a knock-out of SynFtsH4 did not result in any noticeable phenotypic difference (Mann *et al.* 2000), suggesting that the role of cyanoFtsH4 may be compensated for by other paralogs and proteases. A noticeable feature of some of the species lacking cyanoFtsH4 is the reduced complexity of metabolism and reduced genome

size due to symbiotic interactions, such as in the cases of *Athelocyanobacterium thalassa* (Zehr *et al.* 2008) and *Richelia intracellularis* (Hilton *et al.* 2013).

Three FtsH sequences found in cyanobacterial genomes showed very long branches (Fig. 3). Sequence A0A0C1UH20, retrieved from the genome of the heterocystous cyanobacterium *Hassalia byssoidea*, probably represents a case of HGT from an uncharacterized bacterium. Its closest relatives in the entire *NCBI RefSeq* database are found in strains of the phylum Bacteroidetes; however, the level of sequence identity was only 77%.

Sequences U5QM63 and Q7NH88, found in the genomes of *Gloeobacter kilaueensis* and *Gloeobacter violaceus,* respectively, also did not give a best hit with any other cyanobacterial sequence. A *BLAST* search showed only a relatively low sequence identity of about 45% with sequences from Proteobacteria and Firmicutes. These are therefore unique FtsH sequences with no particular phylogenetic affiliation. Either they are descended from ancestral cyanobacterial FtsH paralogs now lost in all other strains of cyanobacteria, or represent a very ancient event of HGT to the last common ancestor of *G. kilaueensis* and *G. violaceous* from a distantly related bacterium of an uncharacterized phyla of bacteria. Given the universality and relative high degree of conservation of FtsH across bacteria, plus the fact that the early branching genus *Gloeobacter* is not a particularly fast evolving clade of cyanobacteria, it seems unlikely that the divergent position of U5QM63 and Q7NH88 is due to unusually high rates of evolution in comparison to other cyanobacteria FtsH sequences.

Multiplicity of FtsH in photosynthetic eukaryotes: The last common ancestor of all eukaryotes likely inherited two bacterial FtsH paralogs, one from Group 2 and one from Group 3 (Fig. 1*B*). Photosynthetic eukaryotes retain in addition at least 3 FtsH paralogs from the cyanobacterial primary endosymbiont: one of these possibly branching prior to the divergence of cyanoFtsH1/2 or at a point in time when these two had not had enough time to diverge; a second one homologous to cyanoFtsH3; and a third one homologous to cyanoFtsH4. This pattern is consistent with recent phylogenetic evidence suggesting an early branching cyanobacterium as the primary endosymbiont (Ponce-Toledo *et al.* 2017). Furthermore, photosynthetic eukaryotes seem to have independently acquired another FtsH paralog from a third bacterial donor, closely related to the phylum Firmicutes. This is not surprising as previous studies have shown that early evolving eukaryotes acquired genes from a broad range of bacterial origins beyond that of the mitochondria (Pittis and Gabaldón 2016) and plastid ancestors (Dagan *et al.* 2013).

Among all clades in the tree of life, photosynthetic eukaryotes have the largest multiplicity of FtsH subunits (Fig. 1*A*), suggesting that after the establishment of the mitochondrion and chloroplast the ancestral FtsH genes underwent several additional duplication events. In the case of *Arabidopsis thaliana*, there are 12 nucleus-encoded FtsH homologs: three (AtFtsH3/4/10) targeted to the mitochondrion, eight (AtFtsH1/2/5/6/7/8/9/12) to the chloroplast, and one (AtFtsH11) possibly targeted to both the chloroplast and mitochondrion (Sakamoto *et al.* 2003, Heazlewood *et al.* 2004, Urantowka *et al.* 2005, Chen *et al.* 2006, Lu *et al.* 2014). *Arabidopsis* also contains four FtsH homologues (AtFtsHi1/2/4/5) containing a disrupted zinc-ion binding site and hence most likely protease inactive.

Fig. 1*B* shows that mitochondrial AtFtsH3/10 subunits of *A. thaliana*, found in Group 2, share a common origin with the mitochondrial YeYta12/YeAfg3(Yta10) subunits of *S. cerevisiae* which form a heterocomplex (Arlt *et al.* 1996); however, YeYta12 and YeAfg3 resulted from a distinct gene duplication event to the one that gave rise to the divergence of AtFtsH3 and AtFtsH10 which might explain why the *Arabidopsis* homologues can form homocomplexes not just heterocomplexes (Piechota *et al.* 2010). Mitochondrial AtFtsH4 found in Group 3, is closely related to the third mitochondrial FtsH homologue of *S. cerevisiae* termed YeYme1, which forms a homocomplex (Baker *et al.* 2011).

Interestingly, AtFtsH11, which is targeted to the chloroplast and possibly the mitochondrion (Urantowka *et al.* 2005), is also closely related to YeYme1. Thus, from a phylogenetic perspective it would seem that AtFtsH11 was of mitochondrial origin and was later co-opted to support chloroplast function.

Experimental evidence has shown that AtFtsH1/2/5/7/8/9 are all targeted to the chloroplast (Sakamoto *et al*. 2003) consistent with their cyanobacterial origins. AtFtsH1/5 appear to have originated from cyanoFtsH3, and AtFtsH2/6/8 from an ancestral cyanoFtsH1/2. These data are in line with experimental data showing a common role for SynFtsH2/3 and AtFtsH1/2/5/8 in PSII repair (Bailey *et al.* 2002, Zaltsman *et al.* 2005, Komenda *et al.* 2006, Boehm *et al.* 2012, Nishimura *et al.* 2016). Currently there is no evidence linking the AtFtsH6 subunit to PSII repair. Instead AtFtsH6 was recently reported to be involved in regulating acquired thermotolerance, or "thermomemory", by degrading the plastidial heat shock protein HSP21 (Sedaghatmehr *et al.* 2016). A previously suggested crucial role for AtFtsH6 in degrading LHCII, the light-harvesting complex of PSII (Zelisko *et al.* 2005), now seems less likely (Wagner *et al.* 2011).

AtFtsH7/9 found in the chloroplast envelope (Sakamoto *et al.* 2003, Wagner *et al.* 2012) are most closely related to cyanoFtsH4 whose function remains unclear. The FtsH-inactive subunit, AtFtsHi3, which only possesses the AAA+ domain, is phylogenetically close to AtFtsH7/9 and may have evolved through loss of the protease domain.

AtFtsH12 has no evolutionary counterpart in cyanobacteria, but instead shows clear proximity to Group 3 FtsH subunits found in the phylum Firmicutes. Found in the same clade as FtsH12 are the inactive FtsH proteases

AtFtsHi1, AtFtsHi2, AtFtsHi4, and AtFtsHi5, which are involved in plastid differentiation during embryogenesis (Kadirjan-Kalbach *et al.* 2012, Lu *et al.* 2014) and, like AtFtsH12, are located in the chloroplast envelope (Ferro *et al.* 2010, Lu *et al.* 2014). Whether AtFtsHi1/2/4/5 subunits are found in the same FtsH complex as AtFtsH12 awaits confirmation.

Both the red alga *Cyanidioschyzon merolae* and the green alga *Chlamydomonas reinhardtii* lack counterparts of AtFtsHi1/i2/i5/12 but do contain 5 FtsH homologues. Two are found in the PSII-specific clades containing AtFtsH1/5 and AtFtsH2/8/6, respectively; one corresponds to AtFtsH3/10; and one to AtFtsH4/11. Notably, *Chlamydomonas reinhardtii* has a counterpart of AtFtsH7/9, or SynFtsH4, while *Cyanidioschyzon merolae* lacks this and instead possesses an additional counterpart to AtFtsH3/10. These observations would suggest that the diversification of FtsH seen in some land plants postdates the divergence of higher plants and algae.

Structural conservation of FtsH proteases: *ConSurf* analysis was performed to assess the structural conservation of the cytosolic region of all types of bacterial FtsH using 270 sequences randomly selected from species covering 55 bacterial phyla or candidate phyla (Table 1S, *supplement available online*). In this analysis regions of strong structural conservation were identified by inserting trimmed FtsH sequences into the crystal structure of the soluble FtsH fragment of *T. thermophilus* (Suno *et al.* 2006). As shown in Fig. 4, the AAA+ domain of bacterial FtsH is well-conserved (purple) with only a small highly divergent area (green) on the surfaces towards the membrane or exposed to the cytoplasm. The surface that interacts with the adjacent AAA+ domain is strictly conserved, which suggested that a ring-shape structure and even that the hexameric complex is a universal feature of all bacterial FtsH. The protease domain, however, was more divergent with the exception of the strictly conserved protease tunnel containing the "HEXXH" motif involved in binding Zn^{2+}. Although sequence differences in the protease domain could reflect relaxation of structural constraints, it might also indicate species-specific interactions such as between FtsH complexes and prohibitins in mitochondria (Steglich *et al.* 1999) and

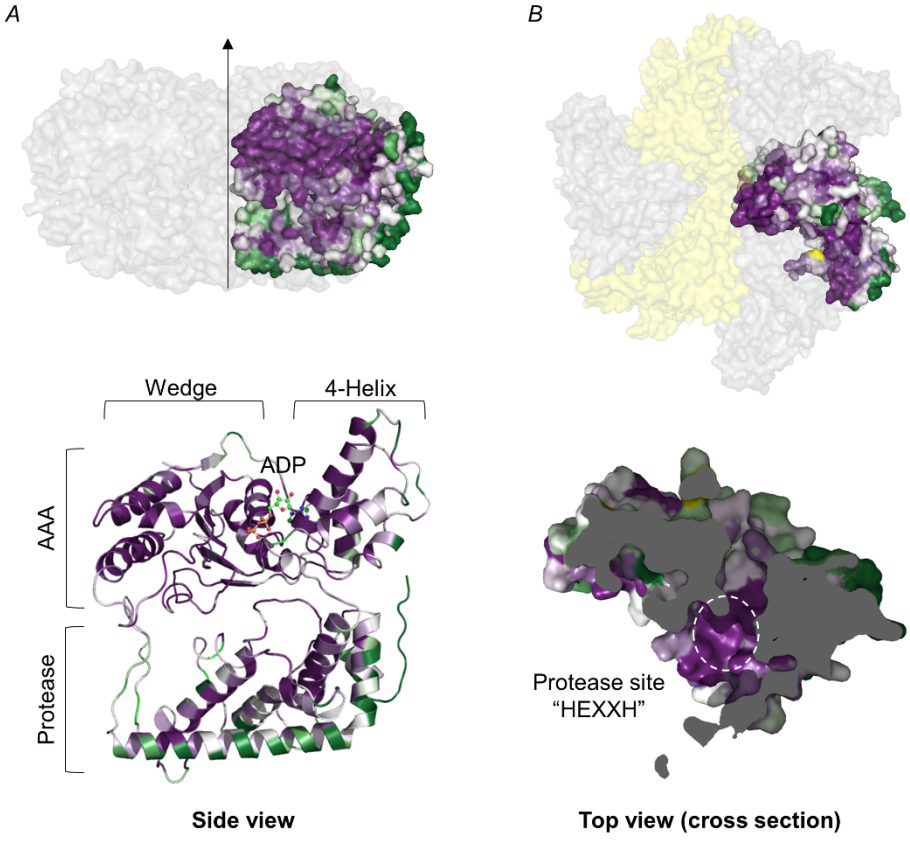

Fig. 4. Structural conservation of FtsH protease sampled from 55 phyla of bacteria. The crystal structure model submitted for *ConSurf* analysis was from *T. thermophiles* (PDB ID: 2DHR). Purple represents highly conserved regions, while green represents poorly conserved regions. *A* – a top view, *B* – a side view. A single monomer is highlighted while the other five as shown with transparency. The two yellow-coloured monomers highlight the three-fold-symmetric structure. *Dashed circle* highlighted the protease activity site "HEXXH".

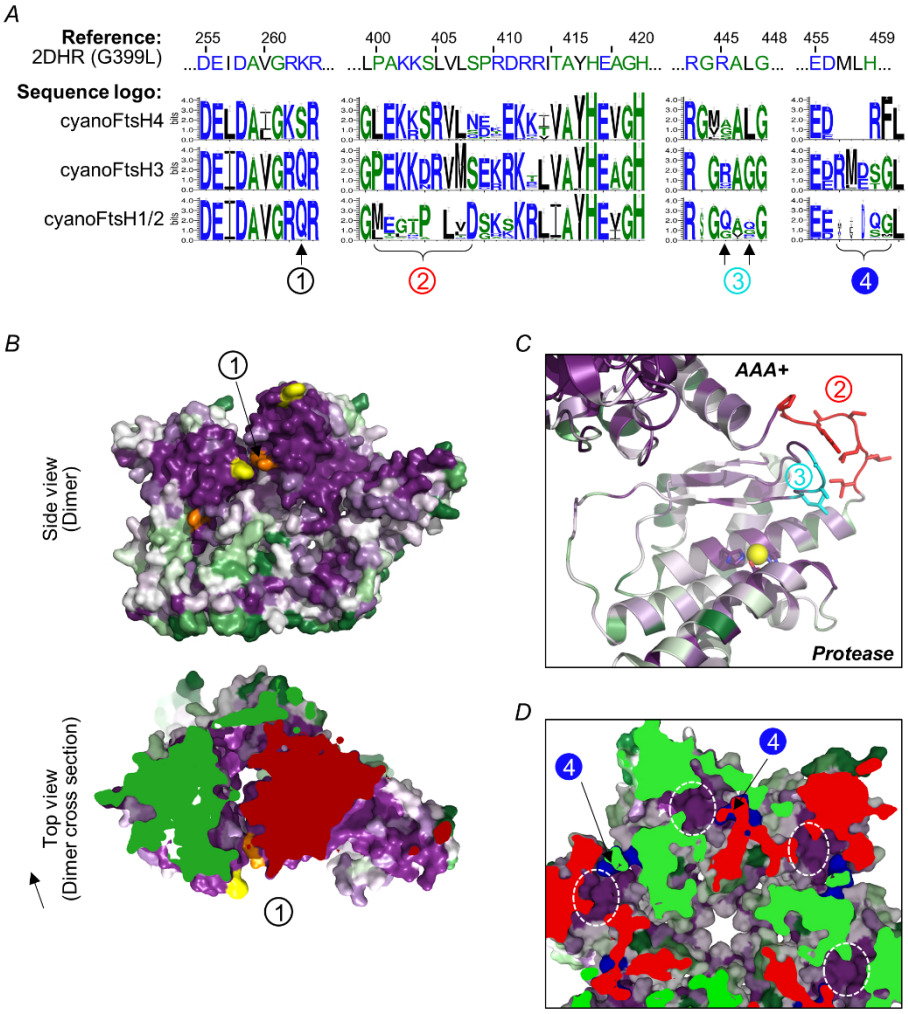

Fig. 5. *A*: Sequence logo profiles of four types of cyanoFtsH. Y axis is information content in bits of amino acid composition at the position indicated in the X axis; the size of a character, single-letter code of amino acid, represents how conserved that amino acid is at that position, the larger the letter the more conserved it is. Colour scheme: *green*, neutral residues; *black*, hydrophobic residues; and *blue*, hydrophilic residues. The amino acid numbering is the same as in the crystal structure from *T. thermophilus* (PDB ID: 2DHR). The four sites of difference between cyanoFtsH are indicated with *arrows* or *brackets*. *B*: Position 1 is shown in *orange* to highlight its relative position to the pore phenylalanine shown in *yellow*. The interior of the two monomers are coloured *red* and *green*, respectively. *C*: Positions 2 and 3 are emphasised in *red* and *cyan*, respectively, and selected residues are showed in stick format. The structure from *T. maritima* (PDB ID: 3KDS) was used for depiction because the proximity between position 2 and 3 is only observed in the *apo*-state. *D*. Top view (cross-section) of Position 4 showed using the structure from *T. thermophilus* (PDB ID: 2DHR). The conserved "HEXXH" motif is indicated by *dashed white circles*.

cyanobacteria (Boehm *et al.* 2012) and with Psb29/Thf1 in cyanobacteria and possibly chloroplasts (Becková *et al.* 2017). The cytosolic region of FtsH might also be involved in the recognition of the N-terminal region of substrates.

The formation of heterocomplexes in the thylakoid membrane of cyanobacteria and photosynthetic eukaryotes and in the mitochondrion are probably the result of convergent evolution. The mitochondrial FtsH paralogs Yta12 and Afg3 (Yta10) adopt a heterohexameric structure, the so-called m-AAA protease complex, where "m" refers to matrix. It has been shown that *Synechocystis* 6803 forms SynFtsH2/FtsH3 and SynFtsH1/FtsH3 heterocomplexes (Boehm *et al.* 2012, Krynická *et al.* 2014). However, the cyanobacterial heterocomplexes are evolutionarily distant from the mitochondrial heterocomplex, as the Yta12/Afg3 proteases emerged from Group 3, while SynFtsH1/2/3 emerged from Group 1. This indicates convergent evolution from homocomplex to heterocomplex. In support of this, Lee *et al*. (2011) showed that the mutation of just two residues in the protease domain was enough to enable Yta12 to form a homocomplex.

A comparison of the cyanoFtsH4 and cyanoFtsH1/2/3: Following the phylogenetic typing of 417 cyanobacterial FtsH into four groups, the structural conservation of each individual group was examined *via ConSurf* analysis to

identify conserved and divergent regions of each group. Sequence logo profiles were generated to allow the comparison of conserved regions. The cyanoFtsH1 and cyanoFtsH2 sequences were grouped together for generating the sequence logo profiles due to the evolutionary proximity, as described above. We found four sites (Fig. 5*A*) that are very well conserved within each paralog group, but significantly divergent between them, implying that differences in function might be governed by the specific structural differences at these sites.

Site 1, at position 263 (*Thermus thermophilus* FtsH numbering), is a strictly conserved glutamine in cyanoFtsH1, cyanoFtsH2, and cyanoFtsH3, but a serine in cyanoFtsH4. As illustrated in Fig. 5*B*, in which a *ConSurf* analysis was performed on cyanobacterial FtsH sequences modelled into the crystal structure of a soluble fragment of FtsH from *Thermus thermophilus*, this residue lies on a highly conserved surface within a short distance [~9 Å in structure 2DHR from *T. thermophilus* (Suno *et al.* 2006)] to the phenylalanine in the adjacent subunit which contributes to the hydrophobic pore through which the target protein enters the FtsH complex (Yamada-Inagawa *et al.* 2003, Suno *et al.* 2006). A similar proximity of this residue to the pore residues was also observed in the structure of FtsH from *Thermotoga maritima* (Bieniossek *et al.* 2006). In the hexameric structure of *Synechocystis* FtsH2/3 heterocomplex, this conserved glutamine would be present in all six monomers, and is therefore likely to occupy a similar position close to the pore residues and might contribute to substrate specificity of the complex.

Sites 2 and 3, at positions 400–408 and 445–447, respectively, are close to the flexible glycine [G399 in *T. thermophilus* (Suno *et al.* 2006, Vostrukhina *et al.* 2015) and G404 in *T. maritima* (Bieniossek *et al.* 2009)] and lid helix (Bieniossek *et al.* 2009, Suno *et al.* 2012) regions R443-E455 in *T. thermophilus* (Suno *et al.* 2006, Vostrukhina *et al.* 2015) (Fig. 5*C*), whose structural flexibility is considered crucial for the intradomain movements needed for full functionality of the complex (Bieniossek *et al.* 2009, Suno *et al.* 2012, Vostrukhina *et al.* 2015). Mutations in the lid helix lead to a decrease of not only the protease activity but also the ATPase activity (Suno *et al.* 2012), strongly indicating the lid helix and flexible glycine interact with each other. CyanoFtsH4 possesses a conserved leucine at position 400 and 447, in contrast to the highly flexible proline and glycine found in cyanoFtsH3 at these respective positions.

Site 4, at position 457–459, shows great variation in sequence between FtsH, including amino acid deletions or insertions. This region is in proximity to the protease active site in all six monomers. Changes in length and composition might therefore control accessibility and/or specificity of substrate processing (Fig. 5*D*).

Discussion

ATPases associated with various cellular activities (AAA+) are an ancient and diverse group of proteins (Frickey and Lupas 2004, Snider *et al.* 2008). At least six distinct protein families containing an AAA+ domain have been traced back to the last universal common ancestor (Iyer *et al.* 2004). One of these six families, denoted the "classical AAA clade" by Iyer *et al*. (2004), was already involved in protein folding and degradation in the universal ancestor of all life. FtsH proteases are likely to have originated from that ancestral protease during the early divergence of the domains Archaea and Bacteria. Prior to the diversification of Bacteria into the major phyla, the ancestral FtsH protease had already duplicated several times. Overall, our data are largely consistent with a pattern of vertical descent within Bacteria, which is not uncommon for universally conserved proteins with essential functions. Bacteria with single FtsH subunits are usually associated with smaller genomes and limited metabolic versatility, the hallmark of symbiotic associations, such as those of Parcubacteria and Microgenometes (Hug *et al.* 2016), for example.

Early ideas suggested that oxygenic photosynthesis might have evolved after the acquisition of Type I and Type II RCs from separate bacterial sources. However, recent detailed analysis of the phylogeny of RC proteins indicates that HGT of RCs to an ancestral nonphotosynthetic cyanobacterium from anoxygenic phototrophic bacteria is unlikely (Cardona 2016b); furthermore, the evolution of RC proteins shows that the last common ancestor to all phototrophic bacteria had already evolved Type I and Type II RC proteins from an earlier gene duplication event (Sousa *et al.* 2013, Harel *et al.* 2015). In addition, it has been argued that the evolution of the structural complexity of PSII and the origin of the oxygen-evolving manganese cluster can only be explained if both types of RC had been evolving in cooperation since the dawn of photosynthesis and that water oxidation might therefore have occurred at a far earlier stage of evolution than previously thought (Cardona 2016b, 2017). This translates to the phylogeny of RC subunits in PSII and PSI showing a significant phylogenetic distance to those in anoxygenic phototrophs in the absence of fast rates of evolution (Cardona 2015, Cardona *et al.* 2017). Given that oxygen-evolving PSII complexes are susceptible to photodamage, the early evolution of PSII would imply an early evolution of mechanisms to protect PSII and to repair damaged PSII. Such a scenario is consistent with our detection of an early divergence of the specific set of FtsH proteases found in present day cyanobacteria and chloroplasts that are involved in PSII repair (Fig. 2).

A plausible intermediate in the evolution of Type II oxygen-evolving complexes is a Type II RC that was able to oxidise Mn^{2+} ions but not water (Cardona *et al.* 2015). Such a RC would still require the generation of highly

oxidising species within the RC, such as chlorophyll cations and tyrosine free radicals, and so would be susceptible to oxidative damage (Komenda *et al.* 2000). Consequently, the early diversification of FtsH proteases might also reflect the need to repair intermediate types of RC operating at higher redox potential.

It is a common misconception that the main driving force behind the diversification of photosynthesis in Bacteria and the origin of oxygenic photosynthesis in cyanobacteria has been the HGT of photosynthetic components. This misconception arises for two reasons: the first one is from the incorrect assumption that the phylogeny of reaction centre proteins is not informative. The second reason is from the incorrect assumption that gene losses are less probable than HGT. Or in the case of photosynthesis, that multiple independent losses of photosynthesis in Bacteria are less likely than the acquisition of photosynthesis *via* HGT. Our results show that the phylogeny of FtsH proteins matches remarkably well the phylogeny of Type I and Type II reaction centres, which strongly suggests that the origin of photosynthesis predates the diversification of most phyla of bacteria. Our results also highlight that photosynthesis has been passed down vertically in most phyla, with HGT being a secondary mechanism of diversification.

One conspicuous case of HGT detected in our phylogenetic analysis is the transfer event of an FtsH from an early branching eukaryote to an ancestral Bacteroidetes. It is well known that Bacteroidetes's closest living relatives are the Chlorobi (Gupta and Lorenzini 2007), so this HGT event suggests that the divergence of Bacteroidetes and the Chlorobi occurred after the evolution of eukaryotes. This is consistent with biomarker evidence for the origin of the phylum Chlorobi in the geochemical record 1.6 billion years ago (Brocks *et al.* 2005), which is coincidental with red algae fossils (Bengtson *et al.* 2017) and other well-documented fossils of early eukaryotes (Butterfield 2015). It is also consistent with molecular clock analysis of prokaryotes (David and Alm 2011, Marin *et al.* 2017). Phototrophic Chlorobi, the green sulfur bacteria, are considered to be a "primitive" and a very ancient phylum of anoxygenic photosynthetic organisms capable of using Fe^{2+} or H_2S as an electron donor to photosynthesis powered by a homodimeric Type I RC (Tice and Lowe 2004, Mix *et al.* 2005, Crowe *et al.* 2008). These are characteristics attributed to the earliest photosynthetic bacteria responsible for the most ancient geochemical and sedimentological traces of photosynthesis 3.5–3.8 billion years ago. From this perspective, our results suggest that the last common ancestor of Bacteroidetes and Chlorobi was phototrophic and that the phylum Bacteroidetes and other nonphototrophic Chlorobi evolved after losses of photosynthesis, as a mechanism of adaptation to heterotrophic or symbiotic lifestyles. Furthermore, the phylogeny of FtsH confirms the phylogenetic proximity of Acidobacteria and Proteobacteria, which is also replicated in evolutionary studies of the bacteriochlorophyll synthesis pathway (Sousa *et al.* 2013, Cardona 2016a), showing unequivocally that the last common ancestor of Acidobacteria and Proteobacteria was also capable of phototrophy (Cardona 2015). This implies that deltaproteobacteria and non-phototrophic gamma-, beta-, and alpha-proteobacteria diversified after losses of photochemical RCs.

Our analysis of the phylogeny of the FtsH protease family also offers insights into the potential role of relatively uncharacterised FtsH subunits in plants. For instance, the mitochondria-targeted Yme1 is involved in protein translocation into the intermembrane space (Rainey *et al.* 2006). A similar role might exist for its close relative in *Arabidopsis*, AtFtsH11. Wagner *et al*. (2011) argues that AtFtsH11 is exclusively located in the chloroplast envelope but not in mitochondrion, and importantly, that knockout of AtFtsH11 leads to a diminished level of several subunits of the protein transport machinery.

Why there has been such a multiplication of FtsH complexes in oxygenic photosynthetic organisms is intriguing. Possible evolutionary constraints include the need to maintain protein quality control in the multiple membrane compartments found in cyanobacteria, chloroplasts, and mitochondria and the fact that oxygenic photosynthesis is associated with the production of singlet oxygen and ROS (reactive oxygen species), leading to protein damage, particularly in the thylakoid membrane which houses the photosynthetic apparatus.

Recognition of damaged D1 by cyanobacterial FtsH complexes is thought to be mediated by partial disassembly of damaged PSII (Krynická *et al.* 2015) and binding of the N-terminal tail of D1 (Komenda *et al.* 2007). Why the FtsH4 homocomplex does not seem to play a major role in PSII repair even though it is found in the thylakoid membrane is unclear. One possible reason might be that specific interactions are required for binding of SynFtsH2/3 to damaged PSII, possibly involving the transmembrane regions of FtsH plus sequences interconnecting the two transmembrane regions on the lumenal side of the membrane (Bailey *et al.* 2001). In addition, substrate recognition might also be mediated *via* adaptor proteins (Kirstein *et al.* 2009). Our modelling has also identified a number of differences between FtsH1/2/3 and FtsH4 that might explain differences in their substrate specificity (Fig. 5) which can be tested through mutagenesis.

References

Arlt H., Tauer R., Feldmann H. *et al.*: The YTA10-12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. – Cell **85**: 875-885, 1996.

Bailey S., Silva P., Nixon P. *et al.*: Auxiliary functions in photosynthesis: the role of the FtsH protease. – Biochem. Soc. Trans. **29**: 455-459, 2001.

Bailey S., Thompson E., Nixon P.J. *et al.*: A critical role for the Var2 FtsH homologue of *Arabidopsis thaliana* in the photosystem II repair cycle *in vivo*. – J. Biol. Chem. **277**: 2006-2011, 2002.

Baker M.J., Tatsuta T., Langer T.: Quality control of mitochondrial proteostasis. – CSH Perspect. Biol. **3**: 1-19, 2011.

Battistuzzi F.U., Hedges S.B.: A major clade of prokaryotes with ancient adaptations to life on land. – Mol. Biol. Evol. **26**: 335-343, 2009.

Beanland T.J.: Evolutionary relationships between "Q-type" photosynthetic reaction centres: Hypothesis-testing using parsimony. – J. Theor. Biol. **145**: 535-545, 1990.

Becková M., Yu J., Krynická V. *et al.*: Structure of Psb29/Thf1 and its association with the FtsH protease complex involved in photosystem II repair in cyanobacteria. – Philos. T. R. Soc. B **372**: 1730, 2017.

Bengtson S., Sallstedt T., Belivanova V. *et al.*: Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. – PLoS Biol. **15**: e2000735, 2017.

Bhaya D., Grossman A.R., Steunou A.-S. *et al.*: Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. – ISME J. **1**: 703-713, 2007.

Bieniossek C., Niederhauser B., Baumann U.M.: The crystal structure of *apo*-FtsH reveals domain movements necessary for substrate unfolding and translocation. – P. Natl. Acad. Sci. USA **106**: 21579-21584, 2009.

Bieniossek C., Schalch T., Bumann M. *et al.*: The molecular architecture of the metalloprotease FtsH. – P. Natl. Acad. Sci. USA **103**: 3066-3071, 2006.

Bittner L.M., Arends J., Narberhaus F.: When, how and why? Regulated proteolysis by the essential FtsH protease in *Escherichia coli*. – Biol. Chem. **398**: 625-635, 2017.

Boehm M., Yu J., Krynicka V. *et al.*: Subunit organization of a *Synechocystis* hetero-oligomeric thylakoid FtsH complex involved in Photosystem II repair. – Plant Cell **24**: 3669-3683, 2012.

Bombar D., Heller P., Sanchez-Baracaldo P. *et al.*: Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. – ISME J. **8**: 2530-2542, 2014.

Brocks J.J., Love G.D., Summons R.E. *et al.*: Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. – Nature **437**: 866-870, 2005.

Bryant D.A., Costas A.M.G., Maresca J.A. *et al.*: *Candidatus Chloracidobacterium thermophilum*: An aerobic phototrophic acidobacterium. – Science **317**: 523-526, 2007.

Bryant D.A., Liu Z., Li T. *et al.*: Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, *Acidobacteria*. – In: Burnap R.L., Vermaas W. (ed.): Functional Genomics and Evolution of Photosynthetic Systems. Pp. 47-102. Springer, Dordrecht 2012.

Butterfield N.J.: Early evolution of the Eukaryota. – Palaeontology **58**: 5-17, 2015.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T.: trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. – Bioinformatics **25**: 1972-1973, 2009.

Cardona T.: A fresh look at the evolution and diversification of photochemical reaction centers. – Photosynth. Res. **126**: 111-134, 2015.

Cardona T.: Origin of bacteriochlorophyll a and the early diversification of photosynthesis. – PLoS ONE **11**: e0151250, 2016a.

Cardona T.: Photosystem II is a chimera of reaction centers. – J. Mol. Evol. **84**: 149-151, 2017.

Cardona T.: Reconstructing the origin of oxygenic photosynthesis: Do assembly and photoactivation recapitulate evolution? – Front. Plant Sci. **7**: 257, 2016b.

Cardona T., Sánchez-Baracaldo P., Rutherford A.W. *et al.*: Molecular evidence for the early evolution of photosynthetic water oxidation. – bioRxiv: https://doi.org/10.1101/109447, 2017.

Chen J., Burke J.J., Velten J. *et al.*: FtsH11 protease plays a critical role in *Arabidopsis* thermotolerance. – Plant J. **48**: 73-84, 2006.

Cheregi O., Sicora C., Kós P.B. *et al.*: The role of the FtsH and Deg proteases in the repair of UV-B radiation-damaged Photosystem II in the cyanobacterium *Synechocystis* PCC 6803. – BBA-Bioenergetics **1767**: 820-828, 2007.

Ciccarelli F.D., Doerks T., von Mering C. *et al.*: Toward automatic reconstruction of a highly resolved tree of life. – Science **311**: 1283-1287, 2006.

Cole J.K., Hutchison J.R., Renslow R.S. *et al.*: Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. – Front. Microbiol. **5**: 109, 2014.

Crowe S.A., Jones C., Katsev S. *et al.*: Photoferrotrophs thrive in an Archean Ocean analogue. – P. Natl. Acad. Sci. USA **105**: 15938-15943, 2008.

Dagan T., Roettger M., Stucken K. *et al.*: Genomes of stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. – Genome Biol. Evol. **5**: 31-44, 2013.

David L.A., Alm E.J.: Rapid evolutionary innovation during an Archaean genetic expansion. – Nature **469**: 93-96, 2011.

Dutilh B.E., Snel B., Ettema T.J.G. *et al.*: Signature genes as a phylogenomic tool. – Mol. Biol. Evol. **25**: 1659-1667, 2008.

Dvořák P., Casamatta D.A., Poulíčková A. *et al.*: *Synechococcus*: 3 billion years of global dominance. – Mol. Ecol. **23**: 5538-5551, 2014.

Dvořák P., Hindák F., Hašler P. *et al.*: Morphological and molecular studies of *Neosynechococcus sphagnicola*, gen. et sp. nov. (Cyanobacteria , Chroococcales). – Phytotaxa **170**: 24-34, 2011.

Eddy S.R.: Accelerated profile HMM searches. – PLoS Comput. Biol. **7**: e1002195, 2011.

Ferro M., Brugière S., Salvi D. *et al.*: AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. – Mol. Cell Proteomics **9**: 1063-1084, 2010.

Finn R.D., Coggill P., Eberhardt R.Y. *et al.*: The Pfam protein families database: towards a more sustainable future. – Nucleic Acids Res. **44**: D279-D285, 2016.

Fischer W.W., Hemp J., Johnson J.E.: Evolution of oxygenic photosynthesis. – Annu. Rev. Earth Pl. Sc. **44**: 647-683, 2016.

Frickey T., Lupas A.N.: Phylogenetic analysis of AAA proteins. – J. Struct. Biol. **146**: 2-10, 2004.

Greening C., Carere C.R., Rushton-Green R. *et al.*: Persistence of the dominant soil phylum *Acidobacteria* by trace gas scavenging. – P. Natl. Acad. Sci. USA **112**: 10497-10502, 2015.

Guindon S., Dufayard J.F., Lefort V. *et al.*: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. – Syst. Biol. **59**: 307-321, 2010.

Gupta R.S., Lorenzini E.: Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the *Bacteroidetes* and *Chlorobi* species. – BMC Evol. Biol. **7**: 71, 2007.

Harel A., Karkar S., Cheng S. *et al.*: Deciphering primordial cyanobacterial genome functions from protein network analysis. – Curr. Biol. **25**: 628-634, 2015.

Heazlewood J.L., Tonti-Filippini J.S., Gout A.M. *et al.*: Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, indicates plant-specific mitochondrial proteins. – Plant Cell **16**: 241-256, 2004.

Hilton J.A., Foster R.A., Tripp H.J. *et al.*: Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. – Nat. Commun. **4**: 1767, 2013.

Hohmann-Marriott M.F., Blankenship R.E.: Evolution of photosynthesis. – Annu. Rev. Plant Biol. **62**: 515-548, 2011.

Hug L.A., Baker B.J., Anantharaman K. *et al.*: A new view of the tree of life. – Nat. Microbiol. **1**: 16048, 2016.

Iyer L.M., Leipe D.D., Koonin E. V. *et al.*: Evolutionary history and higher order classification of AAA+ ATPases. – J. Struct. Biol. **146**: 11-31, 2004.

Jun S.-R., Sims G.E., Wu G.A. *et al.*: Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution. – P. Natl. Acad. Sci. USA **107**: 133-138, 2010.

Kadirjan-Kalbach D.K., Yoder D.W., Ruckle M.E. *et al.*: *FtsHi1/ARC1* is an essential gene in *Arabidopsis* that links chloroplast biogenesis and division. – Plant J. **72**: 856-867, 2012.

Kato Y., Sakamoto W.: Protein quality control in chloroplasts: A current model of D1 protein degradation in the photosystem II repair cycle. – J. Biochem. **146**: 463-469, 2009.

Kirstein J., Molière N., Dougan D.A. *et al.*: Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. – Nat. Rev. Microbiol. **7**: 589-599, 2009.

Komárek J., Kaštovský J., Mareš J. *et al.*: Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. – Preslia **86**: 295-335, 2014.

Komenda J., Barker M., Kuviková S. *et al.*: The FtsH protease *slr0228* is important for quality control of photosystem II in the thylakoid membrane of *Synechocystis* sp. PCC 6803. – J. Biol. Chem. **281**: 1145-1151, 2006.

Komenda J., Hassan H.A.G., Diner B.A. *et al*.: Degradation of the photosystem II D1 and D2 proteins in different strains of the cyanobacterium *Synechocytis* PCC 6803 varying with respect to the type and level of *psbA* transcript. – Plant Mol. Biol. **42**: 635-645, 2000.

Komenda J., Knoppová J., Krynická V. *et al.*: Role of FtsH2 in the repair of Photosystem II in mutants of the cyanobacterium *Synechocystis* PCC 6803 with impaired assembly or stability of the $CaMn_4$ cluster. – BBA-Bioenergetics **1797**: 566-575, 2010.

Komenda J., Sobotka R., Nixon P.J.: Assembling and maintaining the Photosystem II complex in chloroplasts and cyanobacteria. – Curr. Opin. Plant Biol. **15**: 245-251, 2012.

Komenda J., Tichý M., Prášil O. *et al.*: The exposed N-terminal tail of the D1 subunit is required for rapid D1 degradation during photosystem II repair in *Synechocystis* sp PCC 6803. – Plant Cell **19**: 2839-2854, 2007.

Krynická V., Shao S., Nixon P.J. *et al.*: Accessibility controls selective degradation of photosystem II subunits by FtsH protease. – Nat. Plants **1**: 15168, 2015.

Krynická V., Tichý M., Krafl J. *et al.*: Two essential FtsH proteases control the level of the Fur repressor during iron deficiency in the cyanobacterium *Synechocystis* sp. PCC 6803. – Mol. Microbiol. **94**: 609-624, 2014.

Lee S., Augustin S., Tatsuta T. *et al.*: Electron cryomicroscopy structure of a membrane-anchored mitochondrial AAA protease. – J. Biol. Chem. **286**: 4404-4411, 2011.

Leonhard K., Herrmann J.M., Stuart R.A. *et al.*: AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria. – EMBO J. **15**: 4218-4229, 1996.

Letunic I., Bork P.: Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. – Nucleic Acids Res. **44**: W242-W245, 2016.

Lu X., Zhang D., Li S. *et al.*: FtsHi4 is essential for embryogenesis due to its influence on chloroplast development in *Arabidopsis*. – PLoS ONE **9**: e99741, 2014.

Lyons T.W., Reinhard C.T., Planavsky N.J.: The rise of oxygen in Earth's early ocean and atmosphere. – Nature **506**: 307-315, 2014.

Mann N.H., Novac N., Mullineaux C.W. *et al.*: Involvement of an FtsH homologue in the assembly of functional photosystem I in the cyanobacterium *Synechocystis* sp. PCC 6803. – FEBS Lett. **479**: 72-77, 2000.

Marin J., Battistuzzi F.U., Brown A.C. *et al.*: The timetree of prokaryotes: new insights into their evolution and speciation. – Mol. Biol. Evol. **34**: 437-446, 2017.

Mistry J., Finn R.D., Eddy S.R. *et al.*: Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. – Nucleic Acids Res. **41**: e121, 2013.

Mix L.J., Haig D., Cavanaugh C.M.: Phylogenetic analyses of the core antenna domain: Investigating the origin of photosystem I. – J. Mol. Evol. **60**: 153-163, 2005.

Mulkidjanian A.Y., Koonin E. V., Makarova K.S. *et al.*: The cyanobacterial genome core and the origin of photosynthesis. – P. Natl. Acad. Sci. USA **103**: 13126-13131, 2006.

Nei M., Gu X., Sitnikova T.: Evolution by the birth-and-death process in multigene families of the vertebrate immune system. – P. Natl. Acad. Sci. USA **94**: 7799-7806, 1997.

Nei M., Rooney A.P.: Concerted and birth-and-death evolution of multigene families. – Annu. Rev. Genet. **39**: 121-152, 2005.

Nishimura K., Kato Y., Sakamoto W.: Chloroplast proteases: updates on proteolysis within and across suborganellar compartments. – Plant Physiol. **171**: 2280-2293, 2016.

Nitschke W., William Rutherford A.: Photosynthetic reaction centres: variations on a common structural theme? – Trends Biochem. Sci. **16**: 241-245, 1991.

Piechota J., Kolodziejczak M., Juszczak I. *et al.*: Identification and characterization of high molecular weight complexes formed by matrix AAA proteases and prohibitins in mitochondria of *Arabidopsis thaliana*. – J. Biol. Chem. **285**: 12512-12521, 2010.

Pittis A.A., Gabaldón T.: Late acquisition of mitochondria by a

host with chimaeric prokaryotic ancestry. – Nature **531**: 101-104, 2016.

Ponce-Toledo R.I., Deschamps P., López-García P. *et al.*: An early-branching freshwater cyanobacterium at the origin of plastids. – Curr. Biol. **27**: 386-391, 2017.

Quaiser A., Ochsenreiter T., Lanz C. *et al.*: Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. – Mol. Microbiol. **50**: 563-575, 2003.

Rainey R.N., Glavin J.D., Chen H.-W. *et al.*: A new function in translocation for the mitochondrial *i*-AAA protease Yme1: import of polynucleotide phosphorylase into the inter-membrane space. – Mol. Cell. Biol. **26**: 8488-8497, 2006.

Rexroth S., Mullineaux C.W., Ellinger D. *et al.*: The plasma membrane of the cyanobacterium *Gloeobacter violaceus* contains segregated bioenergetic domains. – Plant Cell **23**: 2379-2390, 2011.

Rinke C., Schwientek P., Sczyrba A. *et al.*: Insights into the phylogeny and coding potential of microbial dark matter. – Nature **499**: 431-437, 2013.

Sacharz J., Bryan S.J., Yu J. *et al.*: Sub-cellular location of FtsH proteases in the cyanobacterium *Synechocystis* sp. PCC 6803 suggests localised PSII repair zones in the thylakoid membranes. – Mol. Microbiol. **96**: 448-462, 2015.

Sakamoto W., Zaltsman A., Adam Z. *et al.*: Coordinated regulation and complex formation of YELLOW VARIEGATED1 and YELLOW VARIEGATED2, chloroplastic FtsH metalloproteases involved in the repair cycle of photosystem II in *Arabidopsis* thylakoid membranes. – Plant Cell **15**: 2843-2855, 2003.

Sedaghatmehr M., Mueller-Roeber B., Balazadeh S.: The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in *Arabidopsis*. – Nat. Commun. **7**: 12439, 2016.

Segata N., Börnigen D., Morgan X.C. *et al.*: PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. – Nat. Commun. **4**: 2304, 2013.

Smakowska E., Czarna M., Janska H.: Mitochondrial ATP-dependent proteases in protection against accumulation of carbonylated proteins. – Mitochondrion **19**: 245-251, 2014.

Shimodaira H., Hasegawa M.: Multiple comparisons of log-likelihoods with applications to phylogenetic inference. – Mol. Biol. Evol. **16**: 1114-1116, 1999.

Silva P., Thompson E., Bailey S. *et al.*: FtsH is involved in the early stages of repair of photosystem II in *Synechocystis* sp PCC 6803. – Plant Cell **15**: 2152-2164, 2003.

Snider J., Thibault G., Houry W.A.: The AAA+ superfamily of functionally diverse proteins. – Genome Biol. **9**: 216, 2008.

Sousa F.L., Shavit-Grievink L., Allen J.F. *et al.*: Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. – Genome Biol. Evol. **5**: 200-216, 2013.

Steglich G., Neupert W., Langer T.: Prohibitins regulate membrane protein degradation by the *m*-AAA protease in mitochondria. – Mol. Cell. Biol. **19**: 3435-3442, 1999.

Stirnberg M., Fulda S., Huckauf J. *et al*.: Osmoregulation in *Synechocystis* sp. PCC 6803: the compatible solute synthesizing enzyme GgpS is one of the targets for proteolysis. – Mol. Microbiol. **63**: 86-102, 2007.

Suno R., Niwa H., Tsuchiya D. *et al.*: Structure of the whole cytosolic region of ATP-dependent protease FtsH. – Mol. Cell **22**: 575-585, 2006.

Suno R., Shimoyama M., Abe A. *et al.*: Conformational transition of the lid helix covering the protease active site is essential for the ATP-dependent protease activity of FtsH. – FEBS Lett. **586**: 3117-3121, 2012.

Takahashi S., Badger M.R.: Photoprotection in plants: A new light on photosystem II damage. – Trends Plant Sci. **16**: 53-60, 2011.

Tice M.M., Lowe D.R.: Photosynthetic microbial mats in the 3,416-Myr-old ocean. – Nature **431**: 549-552, 2004.

Tomoyasu T., Yamanaka K., Murata K. *et al.*: Topology and subcellular localization of FtsH protein in *Escherichia coli*. – J. Bacteriol. **175**: 1352-1357, 1993.

Urantowka A., Knorpp C., Olczak T. *et al.*: Plant mitochondria contain at least two *i*-AAA-like complexes. – Plant Mol. Biol. **59**: 239-252, 2005.

Vostrukhina M., Popov A., Brunstein E. *et al.*: The structure of *Aquifex aeolicus* FtsH in the ADP-bound state reveals a C_2-symmetric hexamer. – Acta Crystallogr. D **71**: 1307-1318, 2015.

Wagner R., Aigner H., Funk C.: FtsH proteases located in the plant chloroplast. – Physiol. Plantarum **145**: 203-214, 2012.

Wagner R., Aigner H., Pružinská A. *et al.*: Fitness analyses of *Arabidopsis thaliana* mutants depleted of FtsH metalloproteases and characterization of three FtsH6 deletion mutants exposed to high light stress, senescence and chilling. – New Phytol. **191**: 449-458, 2011.

Ward N.L., Challacombe J.F., Janssen P.H. *et al.*: Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. – Appl. Environ. Microbiol. **75**: 2046-2056, 2009.

Wu M., Eisen J.A.: A simple, fast, accurate method of phylogenomic inference. – Genome Biol. **9**: R151, 2008.

Yamada-Inagawa T., Okuno T., Karata K. *et al.*: Conserved pore residues in the AAA protease FtsH are important for proteolysis and its coupling to ATP hydrolysis. – J. Biol. Chem. **278**: 50182-50187, 2003.

Yamada K.D., Tomii K., Katoh K.: Application of the MAFFT sequence alignment program to large data - Reexamination of the usefulness of chained guide trees. – Bioinformatics **32**: 3246-3251, 2016.

Yu F.: Functional redundancy of AtFtsH metalloproteases in thylakoid membrane complexes. – Plant Physiol. **138**: 1957-1966, 2005.

Yu F., Park S., Rodermel S.R.: The *Arabidopsis* FtsH metalloprotease gene family: Interchangeability of subunits in chloroplast oligomeric complexes. – Plant J. **37**: 864-876, 2004.

Zakon H.H.: Convergent evolution on the molecular level. – Brain. Behav. Evol. **59**: 250-261, 2002.

Zaltsman A., Ori N., Adam Z.: Two types of FtsH protease subunits are required for chloroplast biogenesis and Photosystem II repair in *Arabidopsis*. – Plant Cell **17**: 2782-2790, 2005.

Zehr J.P., Bench S.R., Carter B.J. *et al.*: Globally distributed uncultivated oceanic N_2-fixing cyanobacteria lack oxygenic photosystem II. – Science **322**: 1110-1112, 2008.

Zelisko A., García-Lorenzo M., Jackowski G. *et al.*: AtFtsH6 is involved in the degradation of the light-harvesting complex II during high-light acclimation and senescence. – P. Natl. Acad. Sci. USA **102**: 13699-13704, 2005.

Zeng Y., Feng F., Medová H. *et al.*: Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. – P. Natl. Acad. Sci. USA **111**: 7795-7800, 2014.