





OPEN

Calpains in cyanobacteria and the origin of calpains

Dominika Vešelényiová¹  , Lenka Hutárová¹, Alexandra Lukáčová², Mária Schneiderová¹, Matej Vesteg² & Juraj Krajčovič¹

Calpains are cysteine proteases involved in many cellular processes. They are an ancient and large superfamily of enzymes responsible for the cleavage and irreversible modification of a large variety of substrates. They have been intensively studied in humans and other mammals, but information about calpains in bacteria is scarce. Calpains have not been found among Archaea to date. In this study, we have investigated the presence of calpains in selected cyanobacterial species using in silico analyses. We show that calpains defined by possessing CysPC core domain are present in cyanobacterial genera *Anabaena*, *Aphanizomenon*, *Calothrix*, *Chamaesiphon*, *Fischerella*, *Microcystis*, *Scytonema* and *Trichormus*. Based on in silico protein interaction analysis, we have predicted putative interaction partners for identified cyanobacterial calpains. The phylogenetic analysis including cyanobacterial, other bacterial and eukaryotic calpains divided bacterial and eukaryotic calpains into two separate monophyletic clusters. We propose two possible evolutionary scenarios to explain this tree topology: (1) the eukaryotic ancestor or an archaeal ancestor of eukaryotes obtained calpain gene from an unknown bacterial donor, or alternatively (2) calpain gene had been already present in the last common universal ancestor and subsequently lost by the ancestor of Archaea, but retained by the ancestor of Bacteria and by the ancestor of Eukarya. Both scenarios would require multiple independent losses of calpain genes in various bacteria and eukaryotes.

Calpains (EC 3.4.22.17) are an ancient superfamily of cytosolic, non-lysosomal cysteine proteases. Proteases are often synthesised as inactive precursors possessing N-terminal inhibitory propeptides which prevent uncontrolled proteolysis, and they are activated by cleavage of the propeptide. Other functions of these propeptides include functional modulation, liganding, correct protease folding and compartmentalization¹. Such N-terminal propeptides are absent from calpains and instead, they are activated by binding of Ca²⁺ ions like many other cysteine proteases².

The first calpain was discovered by Guroff (1964) when he purified a calcium-activated enzyme present in a soluble fraction of the rat brain³. Other calpains have been later identified in various organisms including mammals, invertebrates, plants and fungi⁴ as well as in some bacteria, but they have not been found in Archaea⁵. Calpains are divided into two groups based on their structure: classical and non-classical ones (Fig. 1). Classical eukaryotic calpains are composed of large and small subunits that are assembled to a functional heterodimer after activation by calcium ions. The large subunit of classical calpains is composed of N-terminal domain, catalytic CysPC domain composed of protease core domains 1 and 2 (PC1 and PC2), C2-like domain and a penta-EF hand domain of the large subunit, PEF(L). Small subunit contains only two conserved domains: penta-EF hand domain of the small subunit, PEF(S), and a glycine-rich region (Fig. 1). Calpains found in bacteria (but also some eukaryotic ones) belong to the group of non-classical calpains, they are monomers lacking the small subunit, and they have in common with other calpains only the catalytic calpain domain CysPC⁵. The PEF domains are specialised for calcium binding, but the CysPC domain can bind Ca²⁺ as well⁶.

Although the number of calpain studies has dramatically increased in recent years, they have focused mainly on mammalian classical calpains, because of their biomedical and clinical importance⁷. Calpains are involved in the development of various diseases including limb-girdle muscular dystrophy^{8,9}, type II diabetes¹⁰, neurodegenerative disorders^{11,12} and cancer¹³. The information about calpains in bacteria and unicellular eukaryotes remains scarce, with the exception of calpains found in yeasts which are relatively well-characterised^{14,15}.

Cyanobacteria are one of the most ancient and major groups of Gram-negative bacteria. They obtain energy by photosynthesis. Cyanobacteria are the only diazotrophs producing oxygen as a by-product of the photosynthetic process¹⁶. They are an enormously diverse group with high adaptive capacity and many species have the

¹Department of Biology, Faculty of Natural Sciences, University of Ss. Cyril and Methodius in Trnava, Námestie J. Herdu 577/2, 917 01 Trnava, Slovakia. ²Department of Biology and Ecology, Faculty of Natural Sciences, Matej Bel University, 974 01 Banská Bystrica, Slovakia. ✉email: dominika.veselenyiova@ucm.sk

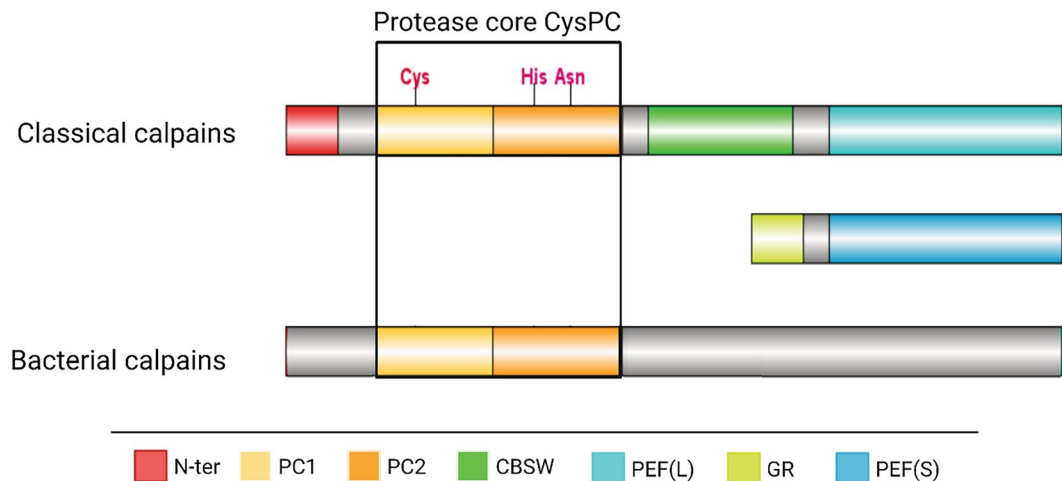


Figure 1. Structural comparison of classical and non-classical calpains. Classical eukaryotic calpains are heterodimers composed of large and small subunits. Each subunit is composed of conserved domains. Large subunit of classical calpains is composed of N-terminal domain, CysPC domain composed of protease core domains 1 and 2 (PC1 and PC2), calpain-type β -sandwich (CBSW) domain and a penta-EF hand domain of the large subunit, PEF(L). Small subunit contains two conserved domains: penta-EF hand domain PEF(S) and a glycine-rich region (GR). Classical calpains are absent from bacteria. Non-classical calpains present in some bacteria (but also in some eukaryotes) are monomers, typically with only a single conserved domain—calpain catalytic domain CysPc composed of PC1 and PC2.

ability to tolerate extreme conditions¹⁷. Therefore, they have colonised almost all habitats on the Earth with the access to sunlight and they play a significant role in biochemical processes in nature¹⁸. The chloroplasts of eukaryotic supergroup Archaeplastida comprising glaucophytes, and red and green algae including land plants¹⁹ have originated from cyanobacteria in the process termed primary endosymbiosis^{20–22}.

Due to the advances in genomics, transcriptomics and proteomics, the genetic makeup of cyanobacteria has been studied more intensively and their significance in biotechnological applications has increased. Cyanobacteria can be a source of bioactive compounds including pharmaceuticals and toxins²³. Several calcium binding proteins (CaBPs) have been discovered in cyanobacteria. These proteins play a significant role in bacterial cells, mainly in processes such as cell division and development, motility, homeostasis, stress response, secretion, molecular transport, cellular signalling and host–pathogen interactions²⁴. Nevertheless, the information about cysteine proteases from the calpain superfamily in cyanobacteria remains limited.

In this study, we have conducted bioinformatic search for calpain homologs in proteomes of various selected cyanobacterial species, mainly colonising extreme environments and species with biotechnological significance. The putative interacting partners of cyanobacterial calpains have also been identified *in silico*. We have also performed the phylogenetic analysis of calpain core CysPC domain to infer the phylogenetic position of the identified cyanobacterial calpains.

Results

Since information about calpains in bacteria is still limited, we decided to search for these cysteine proteases in 50 selected cyanobacterial species (Supplementary Table S1). Using HMM of calpain catalytic domain (CysPC) and subsequent conserved domain prediction by CDD and Pfam, we identified 13 putative calpain sequences (Table 1) in 10 of 50 cyanobacterial species (10/50; 20%), seven species (*Anabaena minutissima*, *Aphanizomenon flosaquae*, *Calothrix parasitica*, *Fischerella thermalis*, *Fischerella muscicola*, *Scytonema hofmannii* and *Trichormus variabilis*) belonging to order Nostocales, one species (*Microcystis aeruginosa*) belonging to Chroococcales and two species (*Chamaesiphon minutus* and *Chamaesiphon polymorphus*) to Synechococcales. Three putative calpains were identified in *S. hofmannii*, two in *F. thermalis*, while only one calpain was identified in other cyanobacterial species.

The Supplementary Table S2 shows the sequences of 13 cyanobacterial calpains identified in this study. Their sequence length ranges from 382 amino acid residues in *F. thermalis* to significantly longer sequences in *C. minutus* and *C. polymorphus* (1145 and 1160 amino acid residues, respectively). The domain structures of all 13 putative calpains analysed using CDD and Pfam are summarised in Table 1. All identified calpains contain conserved CysPC domain at the C-terminus and the most of them also contain single or multiple bacterial pre-peptidase C-terminal domains (PPCs) at the N-terminus (Table 1, Fig. 2).

The alignment of CysPC domains from 13 cyanobacterial calpains, two another bacterial calpains, human calpains CAPN1 and CAPN2, and from two plants DEK1 calpains, and the sequence logo generated from this alignment is presented in Fig. 3. By definition, all CysPC domains should share a catalytic triad of amino acids typical for calpains—Cys (C), His (H) and Asn (N). All these three residues were correctly aligned for all 13 cyanobacterial CysPC domains (Fig. 3). Figure 4 shows the alignment of PPC domains present in cyanobacterial calpains.

Organisms	Sequence length	CS	Conserved domain(s)	Pfam ID	Number of domains
<i>Anabaena minutissima</i>	608		CysPC	cl00051	1
<i>Aphanizomenon flosaquae</i>	964		CysPC	cl00051	1
			PPC	pfam04151	5
<i>Calothrix parasitica</i>	998		CysPC	cl00051	1
			PPC	pfam04151	1
<i>Chamaesiphon minutus</i>	1145		CysPC	cl00051	1
			PPC	pfam04151	1
<i>Chamaesiphon polymorphus</i>	1160		CysPC	cl00051	1
			PPC	pfam04151	1
<i>Fischerella muscicola</i>	585		CysPC	cl00051	1
			PPC	pfam04151	1
<i>Fischerella thermalis 1</i>	585		CysPC	cl00051	1
			PPC	pfam04151	1
<i>Fischerella thermalis 2</i>	392		CysPC	cl00051	1
<i>Microcystis aeruginosa</i>	677		CysPC	cl00051	1
			PPC	pfam04151	2
<i>Scytonema hofmannii 1</i>	578		CysPC	cl00051	1
<i>Scytonema hofmannii 2</i>	509		CysPC	cl00051	1
<i>Scytonema hofmannii 3</i>	442		CysPC	cl00051	1
<i>Trichormus variabilis</i>	827		CysPC	cl00051	1
			PPC	pfam04151	2

Table 1. Information about calpains found in cyanobacteria. All sequences possess a single catalytic core domain of calpains (CysPC). Most sequences also possess one, two or five bacterial pre-peptidase C-terminal domains (PPC). All identified CysPC domains contain three catalytic sites (CS) typical for calpains.

Although most calpains are cytosolic, a few of eukaryotic calpains can be also found in organelles such as mitochondria, e.g. human calpain 10^{4,25} or they are transmembrane proteins in plasma membrane as in the case of plant calpain DEK1^{26,27}. Thus, we also performed prediction of transmembrane regions in cyanobacterial calpains. TMHMM did not identify transmembrane regions in any of cyanobacterial calpains suggesting their cytosolic localization.

Smart BLAST was used to evaluate whether the identified sequences are considered to belong to the calpain superfamily. All 13 sequences show reasonable similarity with members of this superfamily. However, the level of sequence identity is relatively low (~30%). This might be due to the lack of annotated bacterial calpain sequences in public databases and only a limited number of well-studied calpains from unicellular eukaryotes and bacteria.

We performed also homological modelling of the 3D structure of each identified cyanobacterial CysPC domain. The results are summarised in Supplementary Table S3. The modelled structure was aligned with the appropriate Protein Data Bank (PDB) template, and the alignment was evaluated based on the number of aligned amino acid residues and Root mean square deviation (RMSD). All 3D structures show significant similarity with the template with RMSD > 1 for all modelled CysPCs (Supplementary Table S4). 3D structures are shown in Fig. 5.

Using String DB²⁸, we predicted the putative interactions of cyanobacterial calpains with other proteins. Almost 40% of interaction partners of cyanobacterial calpains predicted by String DB were putative cyanobacterial proteins currently missing annotation in public databases. Putative interaction partners of cyanobacterial calpains are shown in Supplementary Fig. S1.

To determine evolutionary relationships between cyanobacterial calpains and calpains present in other bacteria and eukaryotes, we performed phylogenetic analysis of the CysPC domain. In contrast to other parts of calpain sequences, CysPC domain is highly conserved in all calpains. It consists of approximately 350 amino acid residues. The results of phylogenetic analysis are shown in Fig. 6. Bacterial and eukaryotic CysPC domains are clearly separated into two monophyletic clusters. All cyanobacterial calpains, except for *S. hofmannii 2*, form a monophyletic cluster within bacteria.

Discussion

We have searched for the presence of calpains in proteomes of 50 cyanobacterial species and we have identified calpains in 10 of them based on HMM of the catalytic CysPC domain typical for calpains proteins. The number of identified cyanobacterial species possessing calpains is relatively low, but as it has been shown previously, cyanobacteria are a highly diverse group and their genome content varies significantly even at the species and

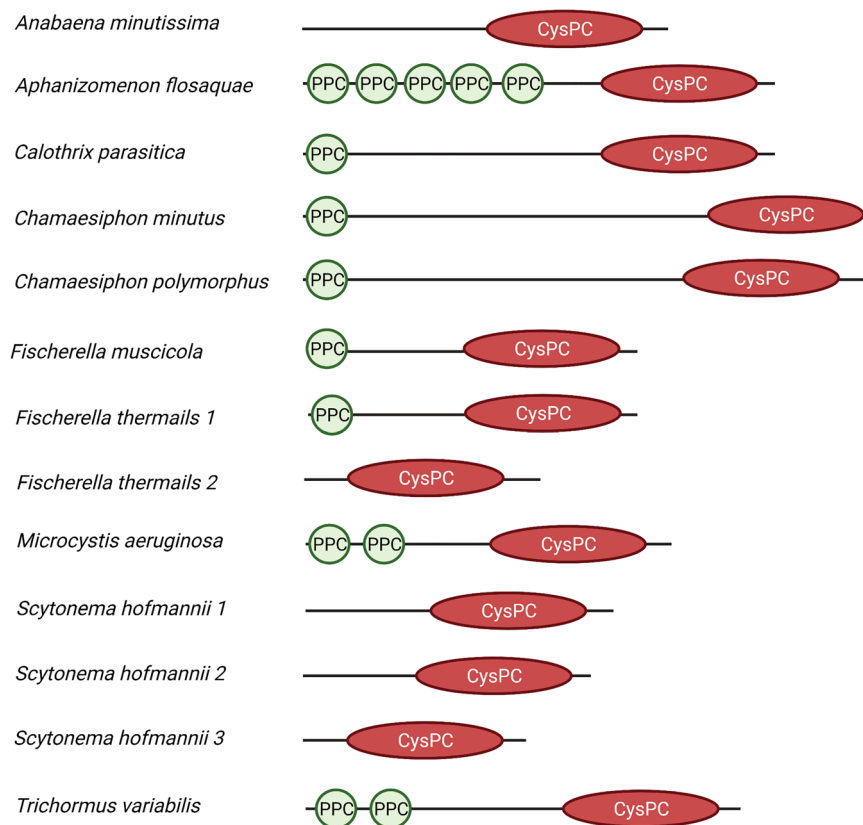


Figure 2. Domain structure of the identified cyanobacterial calpains. All calpains contain CysPc conserved domain (red ellipse) with a catalytic triad C, H, N typical for calpains. Some calpains possess the second conserved domain, PPC (green circle), which can be present in several copies.

strain levels²⁹. CysPC domain is in cyanobacteria often associated with PPC domain (Table 1, Fig. 2), which is typically present in bacterial secreted proteins at their C-terminus³⁰, while in cyanobacterial calpains, it is found at the N-terminus. The transmembrane helical regions are absent from all putative cyanobacterial calpains suggesting their cytosolic localisation. These findings are consistent with the study of calpains in other bacteria that also possess PPC at the N-terminus and do not possess any predictable transmembrane regions⁵.

Calpains are known to be involved in many cellular processes in multicellular eukaryotes such as aleurone bilayer development and positional cell division in plants³¹, and brain function, memory formation and the development of many pathological processes in mammals⁷. Calpains cleave a wide range of substrates, among which are e.g. protein kinases, receptor molecules and proteins involved in signal transduction. It has been proposed that calpains play main role in regulation of cell signalling rather than in protein digestion^{32,33}. However, their function in bacteria remains unknown.

The predicted interaction partners of identified cyanobacterial calpains differ significantly among studied cyanobacterial species. None of them has been predicted to interact with calpains in all cyanobacterial species and only few of them have been commonly predicted for two, three or four species. Methionine synthase is putatively interacting with calpains in four cyanobacterial species, while S8 peptidase and glycoside hydrolase family 3 proteins (such as beta-N-acetylhexosaminidase) with calpains in three cyanobacterial species. SecA involved in protein translocation across cytoplasmic and thylakoid membrane, TamB (a component of the translocation and assembly module autotransporter complex) and collagen triple helix repeat protein have been identified as putative calpain interacting partners only in two cyanobacterial species. Other annotated proteins putatively interacting with cyanobacterial calpains have been predicted only for a single cyanobacterial species and almost 40% of predicted interacting partners have been non-annotated proteins (Supplementary Fig. S1). Based on these results, it is currently difficult to draw any meaningful conclusion about a function of cyanobacterial calpains. The predicted interaction partners and the function of cyanobacterial calpains can be experimentally verified in the future.

We also conducted phylogenetic analysis of calpain core CysPC domain to infer the phylogenetic position of cyanobacterial calpains. The phylogenetic analysis revealed the monophyly of bacterial as well as of eukaryotic CysPCs with bootstrap support 97 and 98, respectively (Fig. 6). No horizontal gene transfers of CysPC domain from bacteria to eukaryotes or vice versa were detected using our taxon sampling. This is consistent with the results of Rawlings (2015) whose phylogenetic analysis identified only two recent horizontal gene transfers from eukaryotes to bacteria and no recent horizontal gene transfer from bacteria to eukaryotes⁵. The branching order within the domain Bacteria and within the domain Eukarya does not correspond to real evolutionary

relationships of bacterial and eukaryotic taxonomic groups, respectively. CysPC is thus unlikely to be a suitable marker for inferring the evolutionary relationships between organisms and it is also possible that several horizontal transfers of calpains have occurred within bacteria as well as within eukaryotes.

With the exception of *S. hofmannii* 2, all cyanobacterial CysPC domains are a monophyletic group within bacterial CysPC domains (Fig. 6). The alignment of cyanobacterial CysPC domains also confirms that CysPC domain 2 from *S. hofmannii* is the most divergent in comparison to other cyanobacterial CysPC domains (Fig. 3). The tree topology also disproves the hypothesis that cyanobacteria, from which chloroplasts of Archaeplastida evolved, were the endosymbiotic donors of archaeplastidial calpains.

The explanation of the origin of eukaryotic calpains depends on the opinion about the origin of eukaryotes themselves. The most popular hypothesis for the origin of eukaryotes suggests that eukaryotes evolved by the endosymbiosis of an alphaproteobacterial ancestor of mitochondria in an archaeal host³⁴, probably from the group Asgard archaea^{35,36}. Since archaea do not possess calpains, while some alphaproteobacteria do, under this scenario, the host archaeal cell could have obtained calpain gene from an alphaproteobacterial endosymbiont. This scenario would be supported if alphaproteobacterial CysPC domains would be placed at the base of eukaryotic CysPCs in the phylogenetic tree with high bootstrap support. Since this is not the case (Fig. 6), our tree does not support alphaproteobacterial origin of eukaryotic calpains. Nevertheless, the hypothesis, that an archaeal ancestor of eukaryotes or the last common ancestor of eukaryotes obtained the calpain gene from an unknown bacterial donor, e.g. via an ancient horizontal gene transfer, cannot be rejected. The scenario that eukaryotic calpains are derived from genes horizontally transferred from a bacterium has been also suggested by⁵.

Rawlings (2015) has also proposed that differential distribution of calpains in bacteria is the result of multiple ancient horizontal gene transfers among bacteria rather than multiple gene losses from various bacteria⁵. In our opinion, the alternative hypothesis that both bacterial ancestor as well as eukaryotic ancestor possessed calpain can be still considered. Currently less popular but still plausible hypotheses for the origin of eukaryotes suggest that Archaea and Eukarya are sister groups. The common ancestor of Archaea and Eukarya might have originated from a bacterium³⁷ or these two domains had a common undefined ancestor—a sister lineage of the domain Bacteria³⁸. An undefined archaeo-eukaryotic ancestor might have been even more complex than all contemporary archaea, Archaea domain might have arisen via reductive evolution of this archaeo-eukaryotic ancestor and the differences between genome contents of contemporary archaeal lineages could be explained by differential gene losses^{39–41}. Considering this scenario, the calpain gene could have been already present in the last universal common ancestor, lost in the ancestor of Archaea, while retained in the ancestor of Bacteria and in the ancestor of Eukarya. Since calpain genes are universally distributed in neither bacteria nor eukaryotes, all mentioned alternative scenarios would require multiple independent losses of calpain genes in various bacterial and eukaryotic lineages.

Methods

We have searched for calpains in silico in proteomes of 50 selected cyanobacterial species (Supplementary Table S1). Our selection was focused on cyanobacteria from extreme biotopes as well as those with biotechnological potency. The proteomic data are available online in NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>)⁴² and Uniprot Proteomes (https://www.uniprot.org/help/proteomes_manual) databases⁴³. Since calpain superfamily is relatively divergent and it has many members, to identify potential calpain sequences, we created Hidden Markov Model (HMM) of calpain catalytic core domain (CysPC). For HMM creation, annotated calpain sequences from various organisms were obtained from the UniProt database (<https://www.uniprot.org/>). HMM was built using HMMER 3.2.1⁴⁴. This model was then applied to cyanobacterial proteomes and putative calpain sequences were identified.

Putative calpains found by HMM were further analysed. Conserved domains were visualised using Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/cdd/>)⁴⁵ and Pfam (<https://pfam.xfam.org/>)⁴⁶, and catalytic sites were identified. The sequences, which did not contain full-length CysPC domain, were excluded from analyses. The identified cyanobacterial sequences were aligned using MAFFT (<https://www.ebi.ac.uk/Tools/msa/mafft/>)⁴⁷ and the sequence logo was generated using WebLogo (<https://weblogo.berkeley.edu/logo.cgi>)⁴⁸. TMHMM v. 2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>)⁴⁹ was used to identify putative transmembrane regions.

SmartBLAST (<https://blast.ncbi.nlm.nih.gov/smartblast/>) was used to search for homologs of cyanobacterial calpains in other bacteria as well as in eukaryotes. 3D structure of putative calpains was predicted by Phyre2⁵⁰ to verify that the identified sequences are really calpains.

String DB was used for the prediction of putative interactions with other proteins²⁸ to elucidate possible function of cyanobacterial calpains. String DB is the database of experimentally determined as well as predicted protein interactions. The predictions of protein interactions are based on protein homology, gene neighbourhood, gene fusions, gene co-occurrence, gene co-expression and/or text mining²⁸.

The homological modelling of the 3D structure of cyanobacterial calpains was performed by Phyre2 server⁵⁰. Supplementary Table S3 contains the list of five best-fitting models for each calpain. The model with the highest confidence and the percentual identity was selected and aligned with an appropriate template structure downloaded from the Protein Data Bank (PDB) using MultiProt server⁵¹. All 3D structures were visualised by ChimeraX software version 1.4 (<https://www.cgl.ucsf.edu/chimerax>)⁵².

We gathered annotated calpain sequences from various organisms from the UniProt database (Supplementary Table S5) and we included them in phylogenetic analysis together with the identified cyanobacterial calpains (Supplementary Table S2). Only the regions corresponding to catalytic CysPC domain were used for the phylogenetic analysis. All CysPC sequences were aligned in MAFFT with automatic settings for amino acid sequences. IQ-Tree⁵³ was used for the construction of phylogenetic trees. Out of 168 models, WAG + F + I + G4⁵⁴ was selected

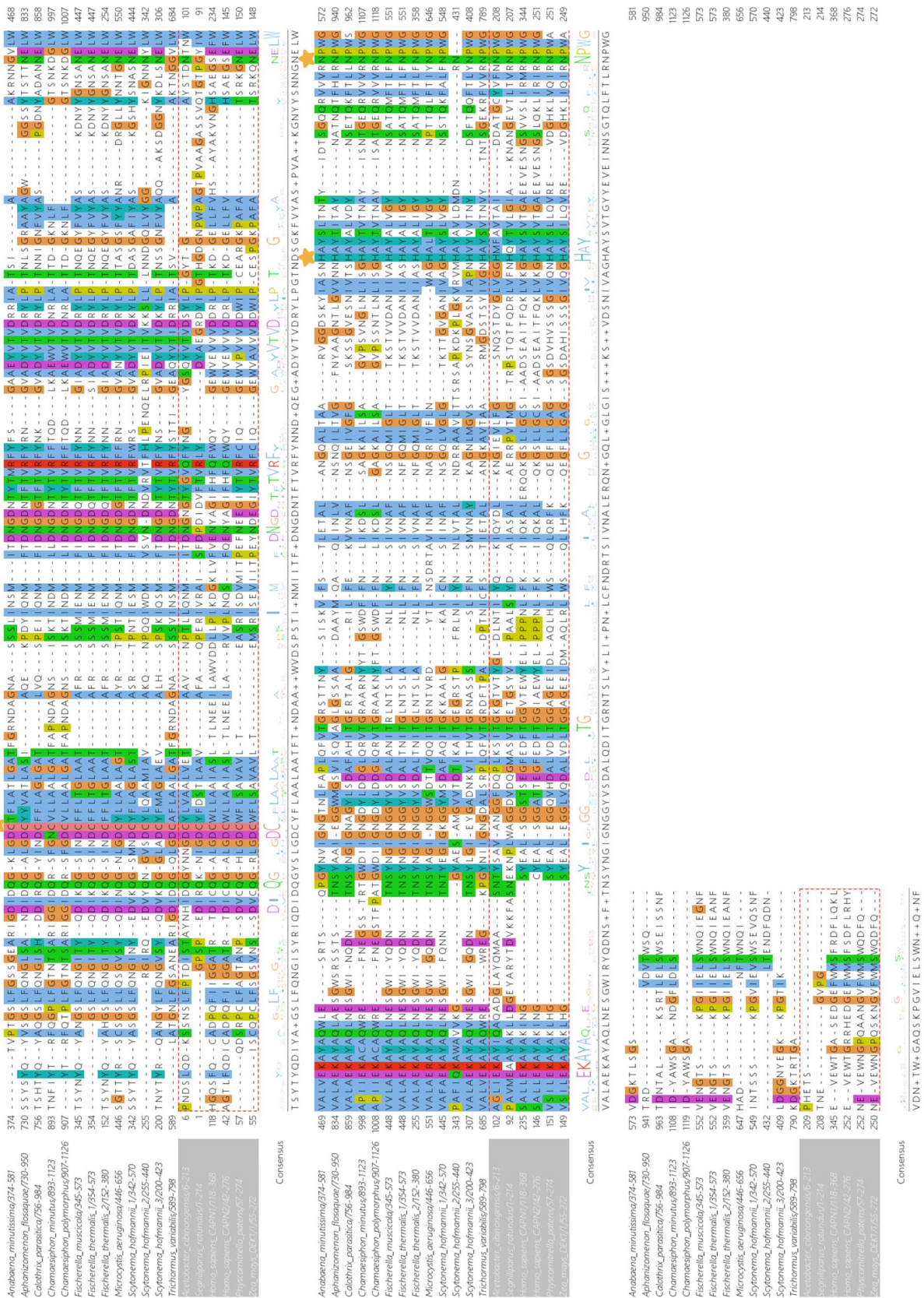


Figure 3. Multiple sequence alignment of CysPC domains present in the identified cyanobacterial calpains and the generated sequence logo. Two another bacterial calpains, CysPCs present in human calpains 1 and 2, and in DEK1 calpains and Physcomitrella patens were also included in the alignment. The presence of conserved C, H and N residues characteristic for CysPC domains is marked by yellow asterisks.

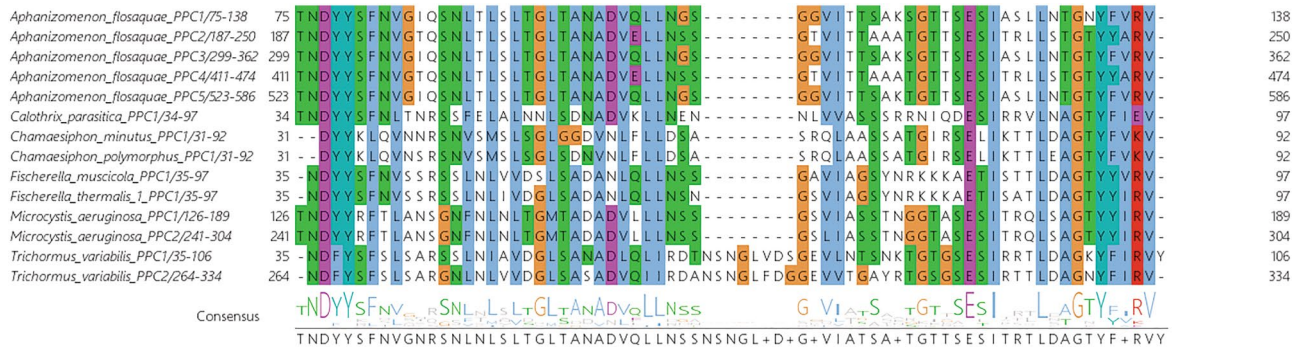


Figure 4. Multiple sequence alignment of PPC domains present in cyanobacterial calpains and the generated sequence logo.

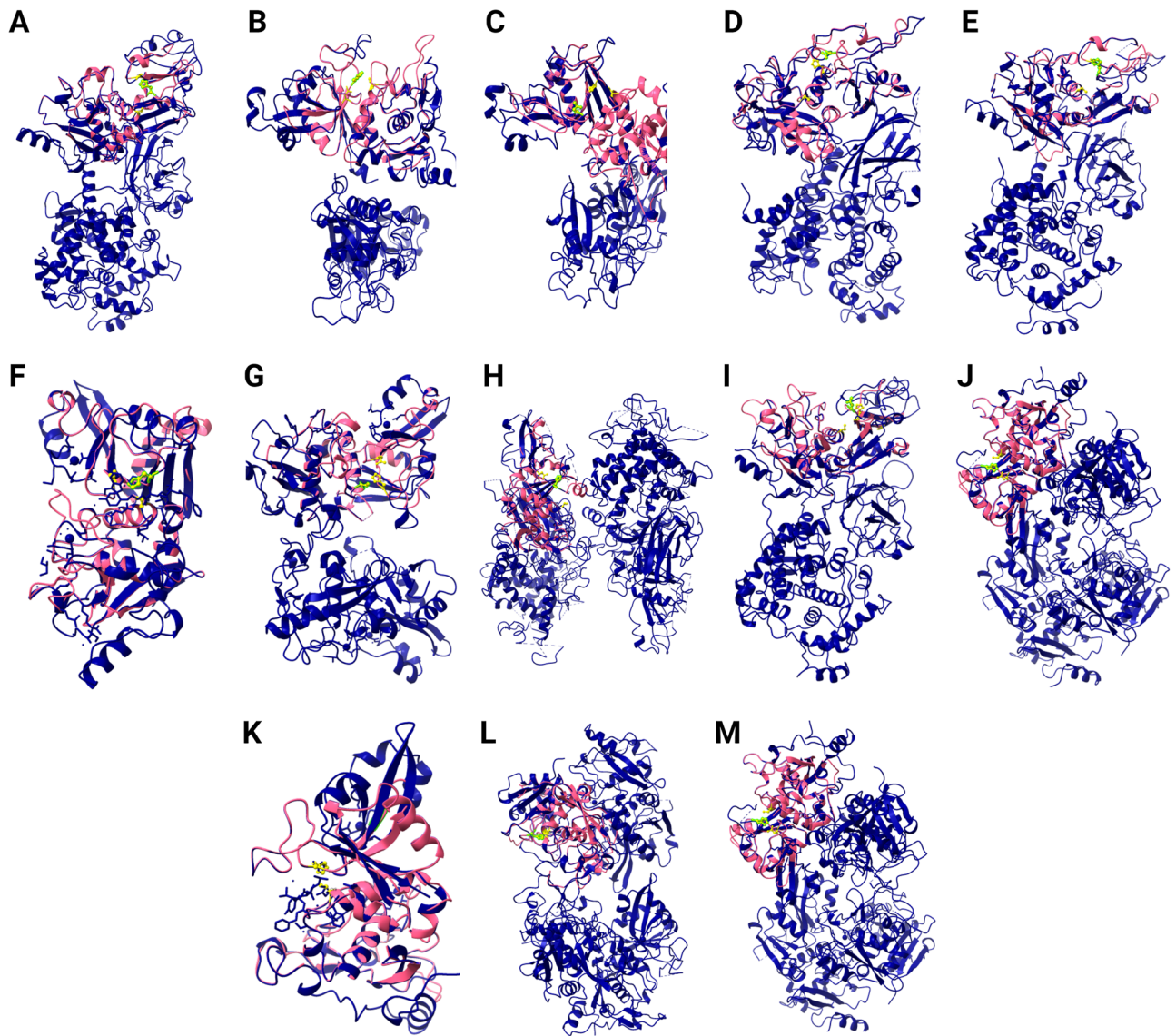


Figure 5. 3D structure modelling of cyanobacterial CysPC domains. The comparison of the template structure obtained from Protein Data Bank (dark blue) and the predicted 3D structure of the cyanobacterial CysPC domain (pink). Catalytic amino acid triad C, H, N is shown in yellow. (A) *Anabaena minutissima*, (B) *Aphanizomenon flosoaquae*, (C) *Calothrix parasitica*, (D) *Chamaesiphon minutus*, (E) *Chamaesiphon polymorphus*, (F) *Fischerella muscicola*, (G) *Fischerella thermalis 1*, (H) *Fischerella thermalis 2*, (I) *Microcystis aeruginosa*, (J) *Scytonema hofmannii 1*, (K) *Scytonema hofmannii 2*, (L) *Scytonema hofmannii 3*, (M) *Trichormus variabilis*.

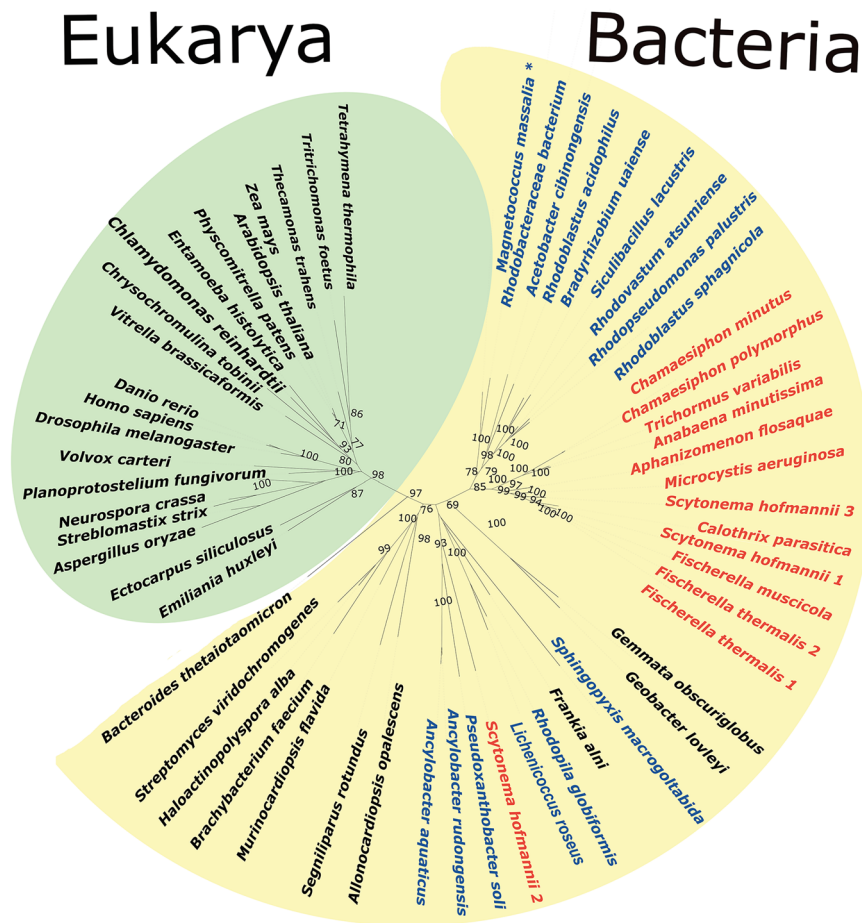


Figure 6. Unrooted phylogenetic tree of calpain catalytic CysPC domains. Cyanobacteria are in red and alphaproteobacteria in blue. *Recent phylogenetic studies of alphaproteobacteria suggested that magnetotactic bacteria including *Magnetococcus* spp. should be excluded from alphaproteobacteria and placed into the separate class Magnetococcia⁵⁶.

as the best suiting model for our dataset. Bootstrap was set to 1000. Phylogenetic tree was visualized using ITOL (<https://itol.embl.de/>)⁵⁵.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary files).

Received: 21 April 2022; Accepted: 8 August 2022

Published online: 16 August 2022

References

- Boon, L., Ugarte-Berzal, E., Vandooren, J. & Opendakker, G. Protease propeptide structures, mechanisms of activation, and functions. *Crit. Rev. Biochem. Mol. Biol.* **55**, 111–165 (2020).
- Wiederanders, B., Kaulmann, G. & Schilling, K. Functions of propeptide parts in cysteine proteases. *Curr. Protein Pept. Sci.* **4**, 309–326 (2003).
- Guroff, G. A neural, calcium-activated proteinase from the soluble fraction of rat brain. *J. Biol. Chem.* **239**, 149–155 (1964).
- Goll, D. E., Thompson, V. F., Li, H., Wei, W. & Gong, J. The calpain system. *Physiol. Rev.* **83**, 731–801 (2003).
- Rawlings, N. D. Bacterial calpains and the evolution of the calpain (C2) family of peptidases. *Biol. Direct.* **10**, 1–12 (2015).
- Liang, Z. *et al.* The catalytic domain CysPc of the DEK1 calpain is functionally conserved in land plants. *Plant J.* **75**, 742–754 (2013).
- Ono, Y., Saido, T. C. & Sorimachi, H. Calpain research for drug discovery: Challenges and potential. *Nat. Rev. Drug Discov.* **15**, 854–876 (2016).
- Richard, I. *et al.* Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* **81**, 27–40 (1995).
- Wang, C., Liang, W. C., Minami, N., Nishino, I. & Jong & Y. J., Limb-girdle muscular dystrophy type 2A with mutation in CAPN3: The first report in Taiwan. *Pediatr. Neonatol.* **56**, 62–65 (2015).
- Buraczynska, M., Wacinski, P., Stec, A. & Kuczmazewska, A. Calpain-10 gene polymorphisms in type 2 diabetes and its micro- and macrovascular complications. *J. Diabetes Complic.* **27**, 54–58 (2013).
- Nixon, R. A. The calpains in aging and aging-related diseases. *Ageing Res. Rev.* **2**, 407–418 (2003).
- Vosler, P. S., Brennan, C. S. & Chen, J. Calpain-mediated signaling mechanisms in neuronal injury and neurodegeneration. *Mol. Neurobiol.* **38**, 78–100 (2008).

13. Storr, S. J., Carragher, N. O., Frame, M. C., Parr, T. & Martin, S. G. The calpain system and cancer. *Nat. Rev. Cancer* **11**, 364–374 (2011).
14. Futai, E. *et al.* The protease activity of a calpain-like cysteine protease in *Saccharomyces cerevisiae* is required for alkaline adaptation and sporulation. *Mol. Gen. Genet.* **260**, 559–568 (1999).
15. Li, M., Martin, S. J., Bruno, V. M., Mitchell, A. P. & Davis, D. A. *Candida albicans* Rim13p, a protease required for Rim101p processing at acidic and alkaline pHs. *Eukaryot. Cell* **3**, 741–751 (2004).
16. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol.* **154**, 157–164 (2003).
17. Gaysina, L. A., Saraf, A. & Singh, P. Cyanobacteria in diverse habitats. In *Cyanobacteria From Basic Science to Applications* (eds Mishra, A. K. *et al.*) 1–28 (Academic Press, Cambridge, 2018).
18. Kulasooriya, S. Cyanobacteria: Pioneers of planet Earth. *Ceylon J. Sci.* **40**, 71–88 (2012).
19. Adl, S. M. *et al.* Revisions to the classification, nomenclature, and diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).
20. Hadariová, L., Vesteg, M., Hampl, V. & Krajčovič, J. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr. Genet.* **64**, 365–387 (2018).
21. Kleiner, F. H., Vesteg, M. & Steiner, J. M. An ancient glaucophyte c_6 -like cytochrome related to higher plant cytochrome c_{6A} is imported into muroplasts. *J. Cell Sci.* **134**, jcs255901 (2021).
22. Vesteg, M., Vacula, R. & Krajčovič, J. On the origin of chloroplasts, import mechanisms of chloroplast-targeted proteins, and loss of photosynthetic ability. *Folia Microbiol.* **54**, 303–321 (2009).
23. Maurya, S. K. & Mishra, R. Importance of bioinformatics in genome mining of cyanobacteria for production of bioactive compounds. In *Cyanobacteria from basic science to applications* (eds Mishra, A. K. *et al.*) 477–506 (Academic Press, Cambridge, 2018).
24. Domínguez, D. C., Guragain, M. & Patrauchan, M. Calcium binding proteins and calcium signaling in prokaryotes. *Cell Calcium* **10**, 1–13 (2015).
25. Ni, R. *et al.* Mitochondrial calpain-1 disrupts ATP synthase and induces superoxide generation in type 1 diabetic hearts: A novel mechanism contributing to diabetic cardiomyopathy. *Diabetes* **65**, 255–268 (2016).
26. Lid, S. E. *et al.* The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. *Proc. Natl. Acad. Sci. USA* **99**, 5460–5465 (2002).
27. Galletti, R. *et al.* Defective Kernel 1 promotes and maintains plant epidermal differentiation. *Development* **142**, 1978–1983 (2015).
28. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
29. Mohanta, T. K., Pudake, R. N. & Bae, H. Genome-wide identification of major protein families of cyanobacteria and genomic insight into the circadian rhythm. *Eur. J. Phycol.* **52**, 149–165 (2017).
30. Yeats, C., Bentley, S. & Bateman, A. New knowledge from old: *In silico* discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.* **3**, 3 (2003).
31. Olsen, O. A., Perroud, P. F., Johansen, W. & Demko, V. DEK1; missing piece in puzzle of plant development. *Trends Plant Sci.* **20**, 70–71 (2015).
32. Wang, K. K., Villalobo, A. & Roufogalis, B. D. Activation of the Ca²⁺-ATPase of human erythrocyte membrane by an endogenous Ca²⁺-dependent neutral protease. *Arch. Biochem. Biophys.* **260**, 696–704 (1988).
33. Moriyasu, Y. & Wayne, R. A novel calcium-activated protease in *Chara corallina*. *Eur. J. Phycol.* **39**, 57–66 (2004).
34. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
35. Spang, A. *et al.* Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* **4**, 1138–1148 (2019).
36. Liu, Y. *et al.* Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).
37. Cavalier-Smith, T. The Neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial mega-classification. *Int. J. Syst. Evol. Microbiol.* **52**, 7–76 (2002).
38. Woese, C. R., Kandler, O. & Wheelis, M. L. Toward a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579 (1990).
39. Forterre, P. The universal tree of life: An update. *Front. Microbiol.* **6**, 717 (2015).
40. Vesteg, M. & Krajčovič, J. The falsifiability of the models for the origin of eukaryotes. *Curr. Genet.* **57**, 367–390 (2011).
41. Vesteg, M., Šándorová, Z. & Krajčovič, J. Selective forces for the origin of spliceosomes. *J. Mol. Evol.* **74**, 226–231 (2012).
42. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
43. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
44. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R. Finn, R. D. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204. <https://www.ebi.ac.uk/Tools/hmmer/>. Accessed from 30 Jan 2021. (2018).
45. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
46. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
47. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
48. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
49. Moller, S., Croning, M. D. R. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653 (2001).
50. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Prot.* **10**, 845–858 (2015).
51. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003).
52. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
53. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
54. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
55. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
56. Muñoz-Gómez, S. A. *et al.* An updated phylogeny of the alphaproteobacteria reveals that the parasitic rickettsiales and holosporales have independent origins. *Elife* **8**, e42535 (2019).

Acknowledgements

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, and the Academy of Sciences (Grant VEGA 1/0694/2021), and by Operational Programme Integrated Infrastructure, project name: "Addressing the societal threats posed by the COVID-19 pandemic", Project code: 313011ASN4, co-financed by the European Regional Development Fund (ERDF)".

Author contributions

D.V. performed most bioinformatic analyses and prepared the first draft of the manuscript; L.H. contributed to the choice of cyanobacteria for analyses and interpretation of results; M.S. was involved in the identification of cyanobacterial calpains; M.V. contributed to the design of phylogenetic analysis and its interpretation; A.L. re-analysed the data and revised the manuscript; J.K. conceived the study. All authors edited the manuscript and approved its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18228-2>.

Correspondence and requests for materials should be addressed to D.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022