

SCIENTIFIC REPORTS



OPEN

Phylogenetic analysis of the CDGSH iron-sulfur binding domain reveals its ancient origin

Soham Sengupta¹, Rachel Nechushtai², Patricia A. Jennings³, Jose' N. Onuchic⁴, Pamela A. Padilla¹, Rajeev K. Azad^{1,5} & Ron Mittler¹

The iron-sulfur (2Fe-2S) binding motif CDGSH appears in many important plant and animal proteins that regulate iron and reactive oxygen metabolism. In human it is found in C1SD1-3 proteins involved in diabetes, obesity, cancer, aging, cardiovascular disease and neurodegeneration. Despite the important biological role of the CDGSH domain, its origin, evolution and diversification, are largely unknown. Here, we report that: (1) the CDGSH domain appeared early in evolution, perhaps linked to the heavy use of iron-sulfur driven metabolism by early organisms; (2) a C1SD3-like protein with two CDGSH domains on the same polypeptide appears to represent the ancient archetype of CDGSH proteins; (3) the origin of the human C1SD3 protein is linked to the mitochondrial endosymbiotic event; (4) the C1SD1/2 type proteins that contain only one CDGSH domain, but function as homodimers, originated after the divergence of bacteria and archaea/eukaryotes from their common ancestor; and (5) the human C1SD1 and C1SD2 proteins diverged about 650–720 million years ago, and C1SD3 and C1SD1/2 share their descent from an ancestral C1SD about 1–1.1 billion years ago. Our findings reveal that the CDGSH domain is ancient in its origin and shed light on the complex evolutionary path of modern CDGSH proteins.

The CDGSH domain is part of an iron-sulfur (2Fe-2S) binding motif that appears in several important human proteins e.g., NEET proteins^{1–5}. This domain is characterized by the following consensus sequence, [C-X-C-X2-(S/T)-X3-P-X-C-D-G-(S/A/T)-H], in which the CDGSH sequence is underlined, and the 3Cys-1His 2Fe-2S coordinating amino acids (aa) are indicated in bold. It was initially annotated as a zinc finger binding domain, but was later shown to bind a 2Fe-2S iron-sulfur cluster^{4,6,7}. CDGSH proteins can be classified into Class I CDGSH proteins that contain only one copy of the Fe-S binding domain, and Class II CDGSH proteins that contain two copies of the Fe-S domain¹. In human, 3 different genes encode CDGSH proteins: C1SD1 encodes mitoNEET (mNT), a homodimer that is anchored to the outer mitochondrial membrane (OMM) and is involved in diabetes, obesity, cancer, cardiovascular disease and neurodegeneration^{6,8–20}. C1SD2 encodes NAF-1, also a membrane-anchored homodimer that is localized to the ER, OMM and the membranes that connect them, and is involved in cancer, neurodegeneration, skeletal muscle maintenance, aging and the regulation of autophagy and apoptosis^{21–33}. A NAF-1 dysfunctional variant was also found to be the causative agent of the human monogenic genetic disease Wolfram Syndrome 2 (WFS2) that is associated with juvenile diabetes, hearing deficiencies, neurodegeneration, blindness, and lower life expectancy^{34–38}. NAF-1 and mNT were also shown to regulate mitochondrial iron and reactive oxygen species (ROS) metabolism, a function that was proposed to be conserved among plant and animal NEET proteins^{4,17,39}. Both mNT and NAF-1 belong to the Class I family of CDGSH proteins¹. In contrast, C1SD3, the third CDGSH human protein, is different from C1SD1 and 2 because it is a monomer that contains 2 iron-sulfur (2Fe-2S) clusters, and is hence a Class II CDGSH protein. C1SD3 is not membrane anchored, and is localized to the matrix space of the mitochondria^{1,2,5}. Very little is known about the function of C1SD3 (also known as Miner2), but its expression level was found to be associated with tumorigenesis (<http://www.protein-atlas.org>). Furthermore, among the 3 human CDGSH proteins, it is the only one to be proposed as an essential protein⁴⁰ (<http://tubic.tju.edu.cn/deg/>).

¹Department of Biological Sciences, University of North Texas, Denton, TX, 76203, USA. ²The Alexander Silberman Institute of Life Science, Hebrew University of Jerusalem, Edmond J. Safra Campus at Givat Ram, Jerusalem, 91904, Israel. ³Department of Chemistry & Biochemistry, University of California at San Diego, La Jolla, CA, 92093, USA. ⁴Center for Theoretical Biological Physics and Department of Physics, 239 Brockman Hall, 6100 Main Street- MS-61, Rice University, Houston, TX, 77005, USA. ⁵Department of Mathematics, University of North Texas, Denton, TX, 76203, USA. Correspondence and requests for materials should be addressed to R.M. (email: ronmittler@gmail.com)

The CDGSH domain appears in multiple proteins that belong to bacteria, archaea and many different unicellular and multicellular eukaryotic organisms, often in combination with other important domains such as *Cyt-b5*, *thioredoxin*, *Fer4_19*, *Rieske* and the *Ferritin-like* domain, indicating that it could be involved in various metabolic reactions in different organisms¹. Perhaps the most important feature of this 3Cys-1His, 2Fe-2S-binding domain, demonstrated for the human C1SD1 and C1SD2 NEET proteins, is that it is both a relatively stable iron-sulfur binding domain, but at the same time it can participate in different reactions that transfer electrons and/or its entire iron-sulfur cluster to different electron and/or cluster acceptor proteins, respectively^{41–51}. This feature may explain why the CDGSH domain is highly conserved from bacteria to human. In addition, it could serve as the basis for the participation of the CDGSH domain in many different important biological functions, as part of essential proteins. A recent study demonstrated, for example, that if the CDGSH domain of NAF-1 is mutated from a 3Cis-1His coordinating structure to a 4Cis coordinating one (a single aa mutation that stabilized the cluster 25-fold over), NAF-1 loses its key function in promoting cellular proliferation in cancer cells⁵². The important function of the CDGSH domain in human disease has also led to different attempts to target this domain with different drugs^{9,53,54}.

We recently used the three members of the human NEET protein family (C1SD1–3) as guides to conduct a phylogenetic analysis of eukaryotic NEET proteins¹. Our study suggested that the *Dictyostelium discoideum*'s CDGSH proteins might be the closest to the ancient archetype of eukaryotic NEET proteins. We further suggested that mNT and NAF-1 emerged via gene duplication around the origin of vertebrates¹. However, the evolutionary timings of these events were not determined. Furthermore, an in depth phylogenetic analysis of the CDGSH domain in bacteria and archaea was not performed.

Here we address the two ends of the CDGSH evolutionary path: The origination of the CDGSH domain in prokaryotes (archaea and bacteria) and eukaryotes; and its divergence times, particularly during the appearance of vertebrates. We show that the CDGSH domain appeared early in evolution, probably linked to the heavy use of Fe-S driven reactions by early organisms^{55–58}, and that its early appearance in archaea and bacteria is associated with the ancient 4Fe-4S binding domain *Fer4_19*. We further show that a C1SD3-like protein with two CDGSH domains on the same polypeptide (Class II), most likely represents the ancient archetype of CDGSH proteins. We also report that the human Class II (C1SD3) protein is more closely related to the bacterial CDGSH protein (of proteobacteria) than to the archaeal Class II CDGSH proteins, and that the human Class I (C1SD1/2) protein is more closely related to archaeal than bacterial CDGSH proteins. Using a molecular clock analysis we also show that the separation of the Class I and Class II eukaryotic CDGSH proteins could be traced to the origins of eukaryotic organisms, and that the human mNT and NAF-1 proteins diverged from their common ancestor ~650–720 million years ago (MYA).

Results

Occurrence of the CDGSH domain in Archaea. The CDGSH domain is found in the genomes of extant archaea belonging to different taxa (Fig. 1). Similar to eukaryotic organisms¹, representatives of both Class I and Class II CDGSH proteins could be found in Archaea (Fig. 1). In addition, in several of the Class II CDGSH proteins of Archaea, the CDGSH domain was found in association with a member of the ancient *Fer4_19* 4Fe-4S cluster binding domain⁵⁷. As in eukaryotes¹, several classes of Archaea lack CDGSH proteins suggesting that some environmental adaptations or metabolic dependencies may not require CDGSH proteins. To conduct a more detailed analysis of CDGSH proteins in archaea, we constructed a phylogenetic tree using PhyML (see Methods) for several different representatives of archaeal CDGSH proteins using the human C1SD3 as an outlier to root the tree (Supplementary Fig. S1). As can be seen in Supplementary Fig. S1, the phylogenetic tree of archaeal CDGSH sequences suggests that Class II archaea CDGSH proteins with the *Fer4_19* domain appear to be the most derived (e.g. *Methanococci* and *Methanobacteria*), whereas the Class I CDGSH proteins of *Thermoprotei* appeared to be among the least derived (the term “derived” refers to branching events in a lineage following divergence from the last common ancestor in a phylogenetic tree). We also retrieved a time tree from TimeTree.org for these proteins (Supplementary Fig. S2). Interestingly, according to the time tree of Archaea (Supplementary Fig. S2), *Thermoprotei*, *Thermococci*, *Methanococci* and *Methanobacteria* appear to have diverged from common ancestors that could be traced back ~3.6–3.9 billion years ago (BYA). This finding could suggest that Class II with the *Fer4_19* and Class I CDGSH domain structure proteins were among the earliest to appear in Archaea. It is also possible that the *Thermoprotei* lineage initially had a Class II with the *Fer4_19* domain protein which was lost in the course of evolution, and only the single domain CDGSH protein was retained. Interestingly, very few examples of Class II CDGSH proteins without a *Fer4_19* domain could be found in archaea. The main forms of archaea CDGSH proteins included therefore the Class I and the Class II CDGSH proteins that contained the *Fer4_19* domain (either in the middle or in the N-terminal).

Occurrence of the CDGSH domain in bacteria. An analysis of CDGSH proteins in different bacterial phyla reveals that the dominant form of CDGSH proteins in bacteria is the Class II type CDGSH protein with two iron-sulfur binding domains present within the same polypeptide (Fig. 2). In contrast, Class I type CDGSH proteins that contain only one CDGSH domain appear in only a few phyla including Rhodothermaeota, Planctomycetes, and the radiation-resistant bacteria *Deinococcus-Thermus*. In many of the bacterial phyla the CDGSH domain appeared in association with other domains such as *Fer4_19*, *Glu_synthase*, *Ferritinlike*, and *Rieske_2*, suggesting that it is associated with many different pathways that are mostly linked to iron and iron-sulfur metabolism (Fig. 2). As with archaea (Fig. 1) and eukaryotes¹, several phyla of bacteria lack CDGSH proteins suggesting that some metabolic adaptations and/or energy pathways may not require CDGSH proteins. Alternatively, functions similar to those of CDGSH proteins in these bacterial phyla could have been performed by a different class of proteins. To conduct a more detailed analysis of CDGSH proteins in bacteria, we generated a phylogenetic tree of several different representatives of bacterial CDGSH proteins using PhyML with human

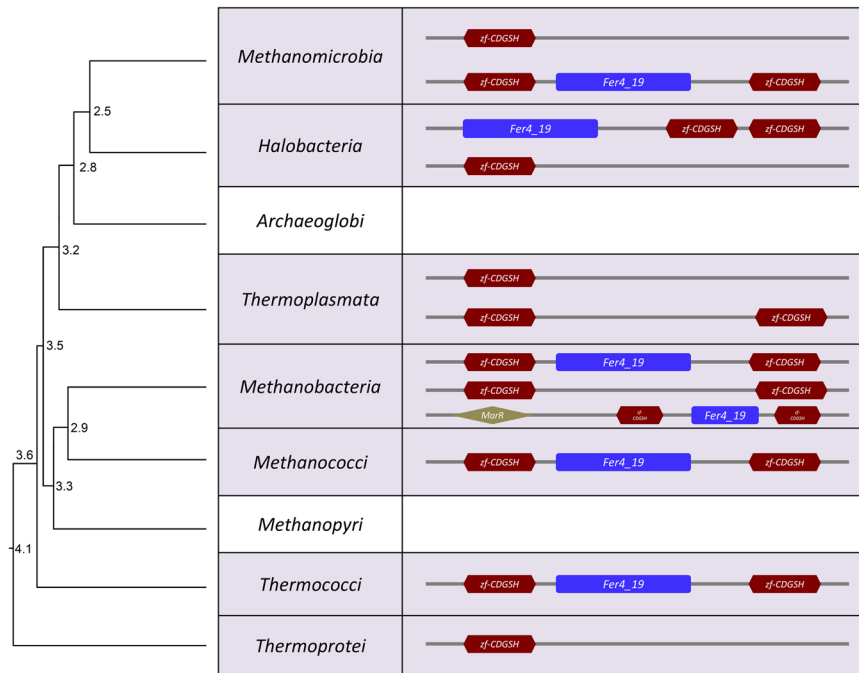


Figure 1. Occurrence and organization of the CDGSH domain in Archaea. A time tree for all archaea classes was obtained from TimeTree.org. The presence or absence of the CDGSH domain within each class was determined using the *Dictyostelium* Cisd3 (XP_647247.1) sequence as the query sequence to perform PSI-BLAST. Archaeal Cisd3 homologs in each class were subjected to PFAM domain analysis to identify the domain organization (see methods). Node ages are represented in billion years ago (BYA).

Cisd3 as an outlier to root the tree (Supplementary Fig. S3). The phylogenetic pattern indicated the Class II CDGSH proteins with two CDGSH domains and no other known domains, as in the Proteobacteria *Candidatus Pelagibacter ubique*, are the most derived ones, in contrast to those harboring other domains in addition to the two CDGSH domains, e.g. *Fer4_19* and CDGSH domains protein as in Proteobacteria *Octadecabacter arcticus*, which appears least derived and could potentially be the ancient archetype of bacterial CDGSH protein. The phylogenetic tree also highlights the gain and loss of domains in the evolution of bacterial CDGSH proteins, which even involved the loss of one of the CDGSH domains (Supplementary Fig. S3). Additionally, we observed that the presence of Class II CDGSH domains is restricted to only a few cyanobacteria (by performing PSI-BLAST as mentioned in Methods section), which have grouped largely with proteobacteria. Two cladistic patterns were observed, one group of cyanobacterial CDGSH protein sequences grouping with representatives of proteobacterial CDGSH protein sequences, both with only two CDGSH domains, while another group of cyanobacterial sequences grouped with representatives of proteobacterial sequences, both with an integrated *Glu_synthase* and *FMN_dh* domain. High bootstrap confidence on these clades suggests that these cyanobacterial CDGSH genes might have been acquired from proteobacteria via horizontal gene transfer. We also retrieved a time tree from TimeTree.org for these proteins (Supplementary Fig. S4). As opposed to archaea (Supplementary Fig. S1), in bacteria the most derived (and probably the most diverged) form of CDGSH proteins as discerned in the phylogeny is a Class II CDGSH protein that also has a glutamate synthase (*Glu_synthase*) domain (Supplementary Fig. S3). In contrast, the least derived (and probably the least diverged) CDGSH protein, as identified by this analysis, is the Class II CDGSH protein with the *Fer4_19* domain (Supplementary Fig. S3). This exact form of CDGSH proteins was also identified in bacterial genomes that have their shared ancestors tracing back to 3.9 BYA (Supplementary Fig. S4). These findings suggest that the appearance of the CDGSH domain in bacteria was likely a very ancient event and resulted in the emergence and evolution of a large variety of different CDGSH proteins that are observed in the genomes of many different extant bacteria (Fig. 2).

Because proteobacteria is thought to represent the bacterial taxa that originally gave rise to the mitochondria of eukaryotic cells via an endosymbiotic event^{59–61}, and because the human Cisd3 protein is localized to the mitochondria⁵, we constructed a phylogenetic tree of bacteria and archaea CDGSH proteins with human Cisd3 (Fig. 3A, Supplementary Fig. S5). Interestingly, although archaea and eukaryotes have been reported to have diverged later than bacteria and eukaryotes^{59,62}, the human mitochondrial Cisd3 protein appears as a sister taxon to the proteobacterial Class II CDGSH protein and away from the archaea Class I or Class II CDGSH proteins in the phylogenetic tree (Fig. 3A, Supplementary Fig. S5). A similar grouping of human Cisd3 with proteobacterial Class II CDGSH proteins was also observed when all archaeal and bacterial CDGSH proteins were included in the phylogenetic analysis (Supplementary Fig. S6). The findings presented in Fig. 3A, Supplementary Figs S5 and S6 support the notion that the origin of the human Cisd3 protein is bacterial in nature and could have emerged as a consequence of the endosymbiotic transfer event that gave rise to mitochondria.

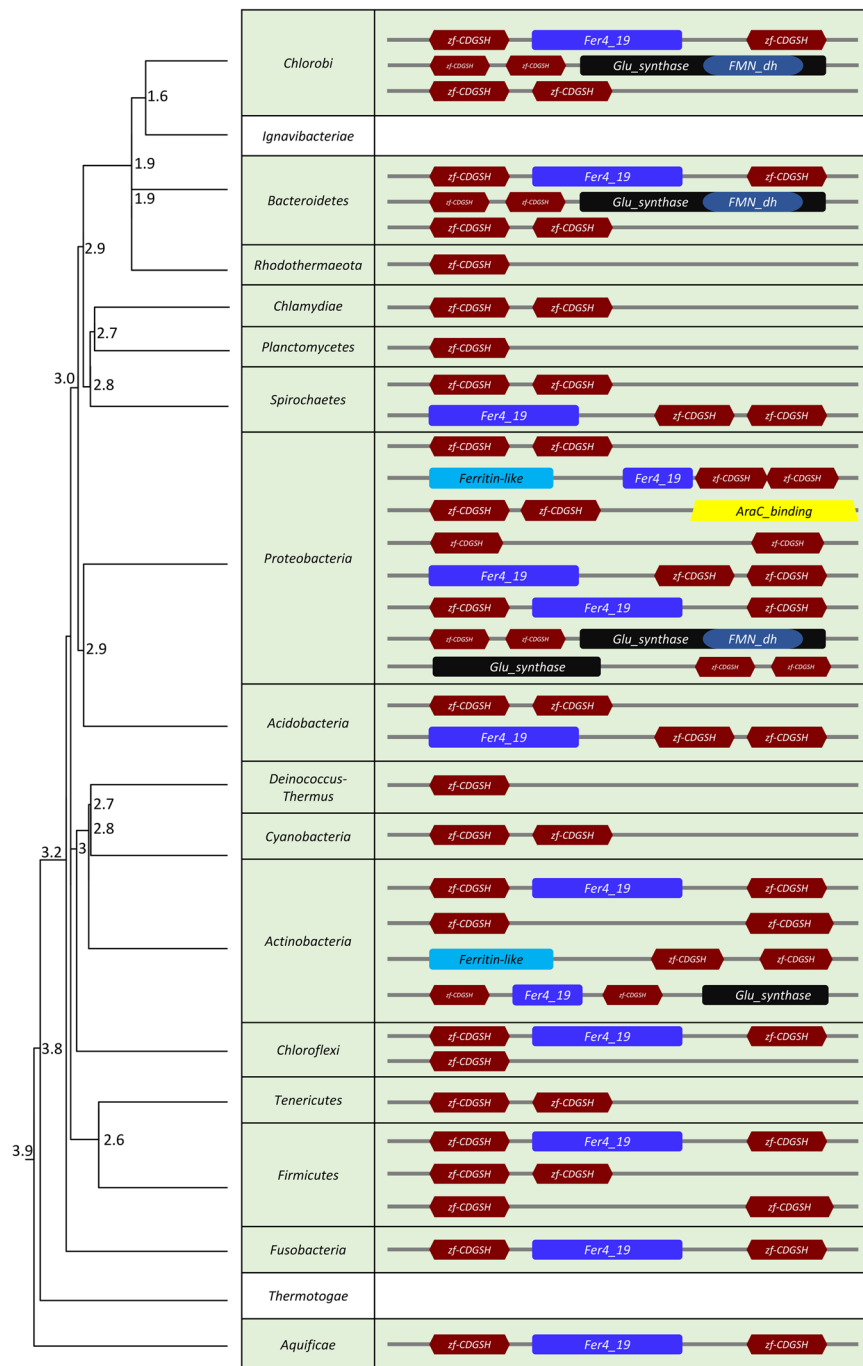


Figure 2. Occurrence and organization of the CDGSH domain in Bacteria. A time tree for all bacteria phylum was obtained from TimeTree.org. The presence or absence of the CDGSH domain and its organization within each phylum was determined using the *Dictyostelium* Cisd (XP_647247.1) sequence as the query sequence to perform PSI-BLAST. Bacterial Cisd homologs from each phylum were subjected to PFAM domain analysis to identify the domain organization (see methods). Node ages are represented in billion years ago (BYA).

Because archaea and eukaryotes contain Class I CDGSH proteins, whereas bacteria contain primarily the Class II CDGSH protein (Figs 1 and 2), we constructed a phylogenetic tree of bacteria and archaea CDGSH proteins with human Cisd2 (a Class I CDGSH protein; Fig. 3B, Supplementary Fig. S7). As shown in Fig. 3B, human Cisd2 appears to be more similar to archaeal CDGSH proteins, than to their bacterial counterparts. A similar grouping of human Cisd2 with archaeal CDGSH proteins was also observed when all archaeal and bacterial CDGSH proteins were included in the phylogenetic analysis (Supplementary Fig. S8). This finding could suggest that the human Class I CDGSH proteins (represented by mNT and NAF-1) could trace their origin to an archaeal ancestor (Fig. 3B), whereas the human Class II CDGSH protein (Miner 2) could trace its origin to a bacterial one (Fig. 3A).

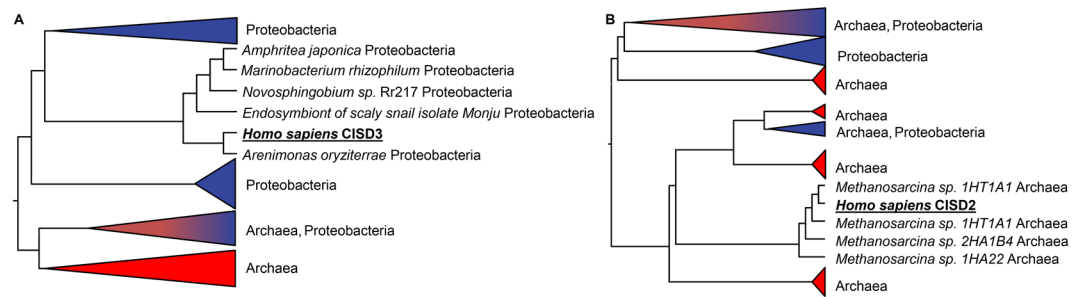


Figure 3. Homology of the human Class II (CISD3) and Class I (CISD2) proteins with representative archaeal and proteobacteria Cisd proteins. **(A)** Phylogenetic tree of human CISD3, Archaea, and proteobacteria Cisd proteins showing homology between the human CISD3 protein and proteobacteria Cisd proteins. For simplification purposes, the different clades were collapsed based on the grouping of archaeal and bacterial sequences. A full version of the tree with complete protein annotations and posterior probabilities (aBayes) values is shown in Supplementary Fig. S5. **(B)** Phylogenetic tree of human CISD2, Archaea, and proteobacteria Cisd proteins showing homology between the human CISD2 protein and archaeal Cisd proteins. For simplification purposes, the different clades were collapsed based on the grouping of archaeal and bacterial sequences. A full version of the tree with complete protein annotations and posterior probabilities (aBayes) values is provided as Supplementary Fig. S6. Representative proteobacteria and archaea sequences were obtained as described in the Methods section. Multiple sequence alignment of the human CISD3/CISD2, archaeal, and proteobacteria Cisd sequences were performed using MUSCLE and a maximum likelihood tree was generated using PhyML (see methods).

Conservation of the CDGSH domain between bacteria, archaea and human. The finding of the CDGSH domain in extant genomes of organisms from all three domains of life, could potentially trace the origin of this domain to ~4 BYA when life originated (see, e.g., Class I CDGSH in *Thermoprotei*, an archaeon, Class II CDGSH in *Aquificae* a bacterium, and both types of CDGSH proteins in *Homo sapiens*; Figs 1–4, Supplementary Figs S1–S8). This intriguing possibility prompted us to assess how conserved the CDGSH domain is between these distinct prokaryotic organisms and human. We therefore performed multiple sequence alignment analysis comparing representative CDGSH proteins from *Aquificae*, *Thermoprotei*, and human. As shown in Fig. 4, the CDGSH domains of the representative archaeal and bacterial organisms chosen for this test are highly conserved with the CDGSH domains of the human CISD1–3 proteins (65% conservation; 30% identity). These findings also suggest that the regions surrounding the CDGSH iron-sulfur binding domain are highly conserved and that the canonical 3Cis-1His coordinating structure of CDGSH proteins is similar between these organisms representing different domains of life. In addition, our analysis revealed that the CDGSH domain of the Class I CDGSH proteins included in the analysis (i.e., mNT, NAF-1 and the representative archaeal sequences from *Thermoprotei*) was more similar to the CDGSH domain that is closer to the N-terminal of the Class II proteins (i.e., human CISD3 and the representative bacterial sequences from *Aquificae*), than to the CDGSH domain that is closer to the C-terminal of Class II CDGSH proteins (Fig. 4). Because the human mNT and NAF-1 proteins contain a transmembrane (TM) domain at their N-terminal^{4,5,39,63}, and this domain plays an important role in their function⁶³, we searched for a transmembrane domain in the *Thermoprotei* and *Aquificae* sequences. However, a TM domain could not be found in these proteins, as well as in human CISD3, suggesting that the TM domain of human CISD1/2 proteins originated later in evolution (Fig. 4).

Molecular clock analysis of CDGSH evolution in eukaryotes. Our previous analysis revealed that in eukaryotes Class I CDGSH domain proteins evolved into human NAF-1 and mNT, and Class II CDGSH proteins evolved into human CISD3¹. However, the evolutionary timing of these events was not determined, as well as the evolutionary timing for the appearance of eukaryotic CDGSH proteins. To address these questions we used the BEAST cross-platform program for Bayesian analysis of molecular sequences using Markov Chain Monte Carlo MCMC^{64–66}. Using two different models (see materials and methods section), we generated two rooted phylogenetic time trees that were very similar in their topology (Fig. 5, Supplementary Figs S9–S10). According to both trees, a distinct set of four major clades represents the eukaryotic CDGSH proteins. These include a Class II CISD3-like clade, and three Class I clades: a CISD1-like clade, a CISD2-like clade, and a clade we term CISD that contains the remainder of the Class I proteins. The CISD clade contains at least two other major sub clades with two different divergence points (Fig. 5, Supplementary Figs S9–S10). According to our analysis, the Class I and Class II eukaryotic CDGSH proteins diverged from their most recent common ancestor ~2.3–2.6 BYA, a time frame that puts this separation event at or close to the emergence of eukaryotic organisms on Earth³⁹, as well as to the great oxidation event⁶⁷. This finding could also support the notion that the progenitor of the eukaryotic Class II proteins is bacterial (Fig. 3A, Supplementary Figs S5–S6) and could have been a consequence of an endosymbiotic event that is inferred to have occurred during the early evolution of eukaryotes (Kurland and Andersson 2000; Hedges and Kumar 2009; Pittis and Gabaldón 2016). Interestingly, a sub-group of Class II CDGSH proteins from the slime molds *Dictyostelium discoideum* and *Acytostelium subglobosum* appears to precede the separation and/or endosymbiotic event that distinguished between Class I and Class II CDGSH proteins. This could suggest that multiple origins could exist for eukaryotic CDGSH proteins, potentially arising from different endosymbiotic

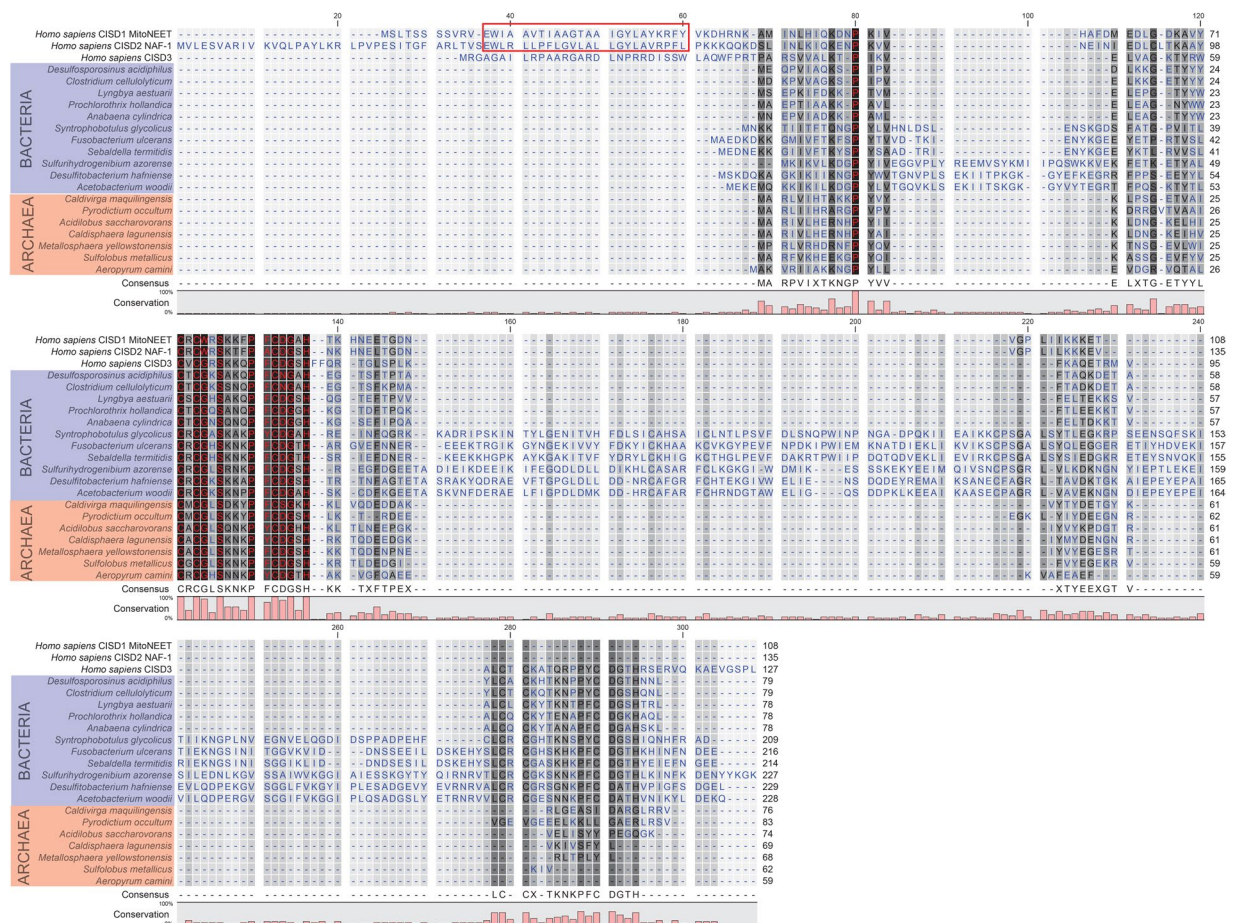


Figure 4. Conservation of the CDGSH domain between bacteria, archaea and human. Multiple sequence alignment analysis comparing representative Cisd proteins from Aquificae, a bacterium, *Thermoprotei*, an archaeon, and human Cisd1–3 proteins. Multiple sequence alignments were performed using MUSCLE. Red box indicates the transmembrane domain of Cisd1 and 2. Bar graph under the aligned sequences indicates degree of conservation (%). Color legend: *Background:* White - Least conserved, Black - Most conserved; *Font:* Blue - Least conserved; Red - Most conserved. Representative bacteria and archaea sequences were obtained as described in the Methods section.

and/or lateral gene transfer events (Fig. 5, Supplementary Figs S9–S10). The separation of animal and plant CDGSH proteins appeared to have occurred about 1.5 BYA, and the separation of Class I Cisd1/2 and Cisd proteins appears to have occurred 1–1.1 BYA. As previously reported¹, plants do not contain a Class II CDGSH protein and it is possible that this class of CDGSH proteins was lost during their evolution.

The divergence of Cisd1 and Cisd2 from their common ancestor, that was previously postulated to coincide with the emergence of vertebrates on Earth¹, occurred 622–768 MYA. The latter time estimate is in accordance with the tree of life timeline for the appearance of vertebrates⁵⁹. Interestingly, in both of our trees representatives of the *Dictyostelium discoideum* and *Acytostelium subglobosum* Class I CDGSH proteins appeared within the Class II CDGSH clade (Fig. 5, Supplementary Figs S9–S10). This finding further supports the hypothesis that the slime mold CDGSH proteins could be similar in sequence to the ancient progenitor of eukaryotic CDGSH proteins¹.

Discussion

Our phylogenetic analysis of the CDGSH domain in prokaryotes revealed that it is highly conserved and widespread among many phyla of bacteria and archaea, suggesting that it evolved early during the emergence of life on Earth (Figs 1–4, Supplementary Figs S1–S8). Its apparent initial association with the *Fer4_19* (4Fe-4S binding) domain (Figs 1, 2, Supplementary Figs S1–S4) demonstrates strong association between the CDGSH domain and other Fe-S proteins. The finding that the 2Fe-2S CDGSH binding domain is ancient and appears in all domains of life is in agreement with the presence of high levels of iron and sulfur in the primordial oceans and the finding of many Fe-S proteins, some belonging to the *Fer4_19* family, in the inferred genome of LUCA last universal common ancestor^{55–58}. In contrast to the finding of the CDGSH domain in association with the *Fer4_19* domain in bacteria and archaea (Figs 1, 2, Supplementary Figs S1–S4), we could not find the *Fer4_19* domain in eukaryotes (not shown), suggesting that some aspects of CDGSH function could be different between prokaryotes and eukaryotes.

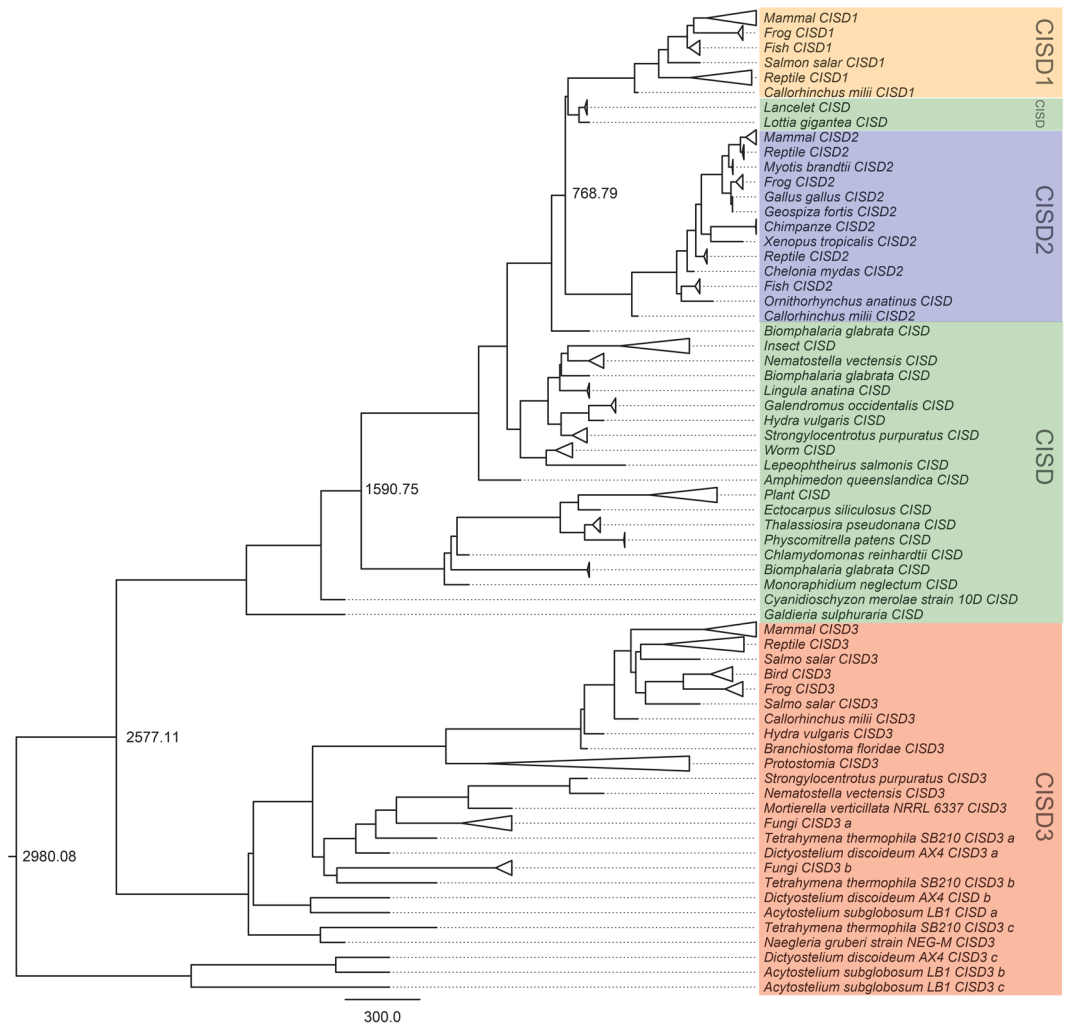


Figure 5. Molecular clock analysis of CDGSH evolution in eukaryotes. A time tree of divergence among eukaryotic CISTD proteins constructed using the BEAST (Bayesian Evolutionary Analysis Sampling Trees) software. Eukaryotic sequences were obtained as described in the Methods section. A total of 150 sequences were subjected to Bayesian analysis using two model combinations (Supplementary Table S3, Supplementary Material) - a constant (shown here and Supplementary Fig. S9) and an exponential (Supplementary Fig. S10) population size model with a relaxed uncorrelated log-normal clock. The divergence times for Class I/II and Class III, Class I and Class II, and the putative ancestor of eukaryotic CISTD proteins are shown. An expanded version of the two trees (constant and exponential) with complete protein annotations and estimated node ages is shown in Supplementary Figs S9 and S10.

Structural studies conducted on the human CDGSH proteins mNT, NAF-1^{3,12,15,24} and on the human and bacterial CISTD3 proteins^{2,68}, our previous phylogenetic analysis of eukaryotic CISTD proteins¹, and our current analysis of these proteins in archaea, bacteria and eukaryotic organisms, reveal an interesting property of CDGSH proteins. When appearing as a Class I single domain CDGSH proteins such as mNT and NAF-1, CDGSH proteins function as homodimers. In contrast, when appearing as a Class II CDGSH proteins that have two CDGSH sequences on the same polypeptide, CDGSH proteins function as a monomer. Furthermore, CDGSH proteins with 3 or more CDGSH domains on the same polypeptide were not found in our current or previous analysis of CDGSH proteins in genomes from different life domains¹. It is therefore possible that CDGSH proteins require two CDGSH 2Fe-2S clusters in close proximity to each other to be able to function in different biological systems. Although this hypothesis, which is based on structural and phylogenetic studies, is highly speculative and would require additional structural and evolutionary studies to be validated, it nevertheless bears importance when attempting to speculate on the origins and functions of ancient CDGSH proteins. Did these proteins originate as a Class I single domain, or did they originate as a Class II double domain? Although we may never know the answer to this question, our findings that bacteria primarily contain Class II CDGSH domain proteins, and that the ancient archetype of CDGSH proteins in bacteria could potentially be a Class II protein associated with a *Fer4_19* domain (Fig. 2, Supplementary Figs S3, S4), suggest that the Class II domain organization structure might have an initial evolutionary advantage, explaining its retention in many eukaryotic organisms and bacteria. Of course, to generate a double CDGSH domain Class II protein,

an initial duplication event of a single domain was required. In this respect it should be noted that, of the two CDGSH domains of Class II proteins, the CDGSH domain closer to the N-terminal of these proteins appears to have a higher degree of homology to the CDGSH domain of Class I proteins (Fig. 4), suggesting that the Class I proteins could have emerged after a deletion of the distal part of the ancient Class II gene encoding C-terminal CDGSH domain, or that the ancient Class II CDGSH proteins emerged after a duplication of the sequence encoding the N-terminal CDGSH domain in an ancient Class I gene.

Our findings that proteobacterial Class II CDGSH proteins are more similar to human and slime mold Cisd3 proteins than to archaeal CDGSH proteins (Fig. 3A, Supplementary Figs S5–S6), suggest that human and perhaps other eukaryotic Cisd3 proteins could trace their origin to the ancient proteobacterial genome that gave rise to mitochondria through the endosymbiotic transfer event. In contrast, the finding that the human Class I Cisd2 protein is more closely related to archaeal Class I and II CDGSH proteins than to proteobacterial Class II CDGSH proteins (Fig. 3B, Supplementary Figs S6–S8) suggests that the ancestor of human Class I CDGSHs protein evolved after the radiation of bacteria and archaea/eukaryotes. If this hypothesis holds true, then the origins of eukaryotic Class I single domain proteins could be distinct from that of eukaryotic Class II proteins (Fig. 3, Supplementary Figs S5–S8). Further studies are required to address this possibility. Additionally, our phylogenetic analysis revealed some interesting instances of possible horizontal gene transfer between archaea and bacteria. For example, CDGSH containing protein sequence of *Asticcacaulis benevestitus* (Proteobacteria, WP_018079727.1) was embedded within an archaeal clade, with *Halobaculum gomorrense* (Archaea, WP_073307495.1) as the nearest neighbor (Supplementary Fig. S6). This clade had a bootstrap confidence of 94.36%, providing support to the possibility of inter-domain gene transfer from an archaeon (*Halobaculum gomorrense*) to a bacterium (*Asticcacaulis benevestitus*). Furthermore, as both these strains are aquatic isolates and dwell in hypersaline environment, their shared ecology might have facilitated gene exchange including of those harboring CDGSH domains^{69,70}.

The evolutionary trajectory of the CDGSH domain is proposed in Fig. 6. In this model, it is hypothesized that a prototype of Class II CDGSH protein is the last common ancestor of all CDGSH proteins. This Class II double domain protein originated from an early duplication event and was retained in the genomes of representative organisms from all domains. As speculated above, this type of CDGSH protein (Class II) provided an adaptive advantage owing to its role in Fe-S driven reactions in early organisms and therefore the archetypal Class II CDGSH gene was selected for and retained in the course of evolution, and the genomes of almost all extant organisms from bacteria to archaea to eukaryotes harbor the Class II CDGSH gene. The appearance of the Class I single domain CDGSH protein that has been reported to function as a homodimer^{3,12,15,24} might have independently occurred after bacterial and archaeal/eukaryotic lineages diverged from their common ancestor, and is currently found primarily in archaea and eukaryotes. Because many of the Class II CDGSH proteins of both bacteria and archaea contain the *Fer4_19* domain, but eukaryotic Class II CDGSH proteins do not, it is possible that archaea and bacteria Class II proteins are related. In contrast, all eukaryotic Class II CDGSH proteins could have evolved from an ancient proteobacteria Class II CDGSH protein that might have lost the *Fer4_19* domain or more likely, this domain got lost after the primary endosymbiotic gene transfer event. It is also possible that once Class I CDGSH proteins evolved, some organisms, for example plants, or certain bacterial and archaeal lineages, lost the Class II domain protein and retained only the Class I CDGSH protein. The model described above suggest that the evolution of a two domain CDGSH protein via domain duplication preceded the evolution of the single domain CDGSH protein that requires a homodimeric structure to function. Because a simple domain duplication rather than the emergence of mechanisms for two identical proteins to dimerize (likely to require a stepwise evolutionary process involving changes to many different amino acids at the surface of the protein) appears more parsimonious and thus plausible. It is reasonable to speculate that the function of the Class II proteins was initially established through domain duplication. This event was then followed by the more complex process of Class I homodimer protein evolution (a single domain protein that could function as a homodimer). Once this new form of protein (Class I homodimer) was established, the Class II protein could have been lost in some lineages, as is likely the case in plants¹. The high prevalence of proteins containing the CDGSH domain in bacteria and archaea is similar to that of proteins containing other important domains such as the catalase heme or the *Fer4_19* domains, indicating that the CDGSH domain could have played an important role in evolution (Supplementary Figs S11, S12). Further studies are therefore required to address the origin and function of this fascinating and highly conserved CDGSH iron-sulfur binding motif.

Methods

Selection of organisms for analysis. For eukaryotes, we selected representative organisms from different lineages with fully sequenced and annotated genomes as described in¹. Briefly, human Cisd1, Cisd2, and Cisd3 were used as query sequences to perform a PSI-BLAST search to obtain the Cisd homologs from the genomes of the organisms selected for our analysis⁷¹. The default parameter setting of PSI-BLAST was used, with Expect threshold of 10 and PSI-BLAST threshold of 5. A total of 150 sequences were selected and the multiple sequence alignments were analyzed using BEAST (Bayesian Evolutionary Analysis Sampling Trees) for determining the divergence time of the Cisd genes^{65,66}. As previous analysis had reported *Dictyostelium discoideum* (a protist slime mold unicellular cell) as the possible representative of the most ancient Cisd gene in eukaryotes¹, we used the *Dictyostelium* Cisd (XP_647247.1) sequence as the query sequence to search for potential bacterial Cisd homologs in the non-redundant database using PSI-BLAST⁷¹. PSI-BLAST iterations were performed until no new BLAST hit was retrieved. All sequences thus obtained were subjected to domain analysis using PFAM, which utilizes profile hidden Markov model to predict the domain architecture of the protein sequences⁷². Sequences with at least one CDGSH domains were kept, and incomplete and partial sequences were discarded. A total of 494 bacterial sequences representing different lineages were selected for further analysis (Supplementary Table S1). In order to retrieve Cisd homologs in archaea, we performed a PSI-BLAST search against the non-redundant

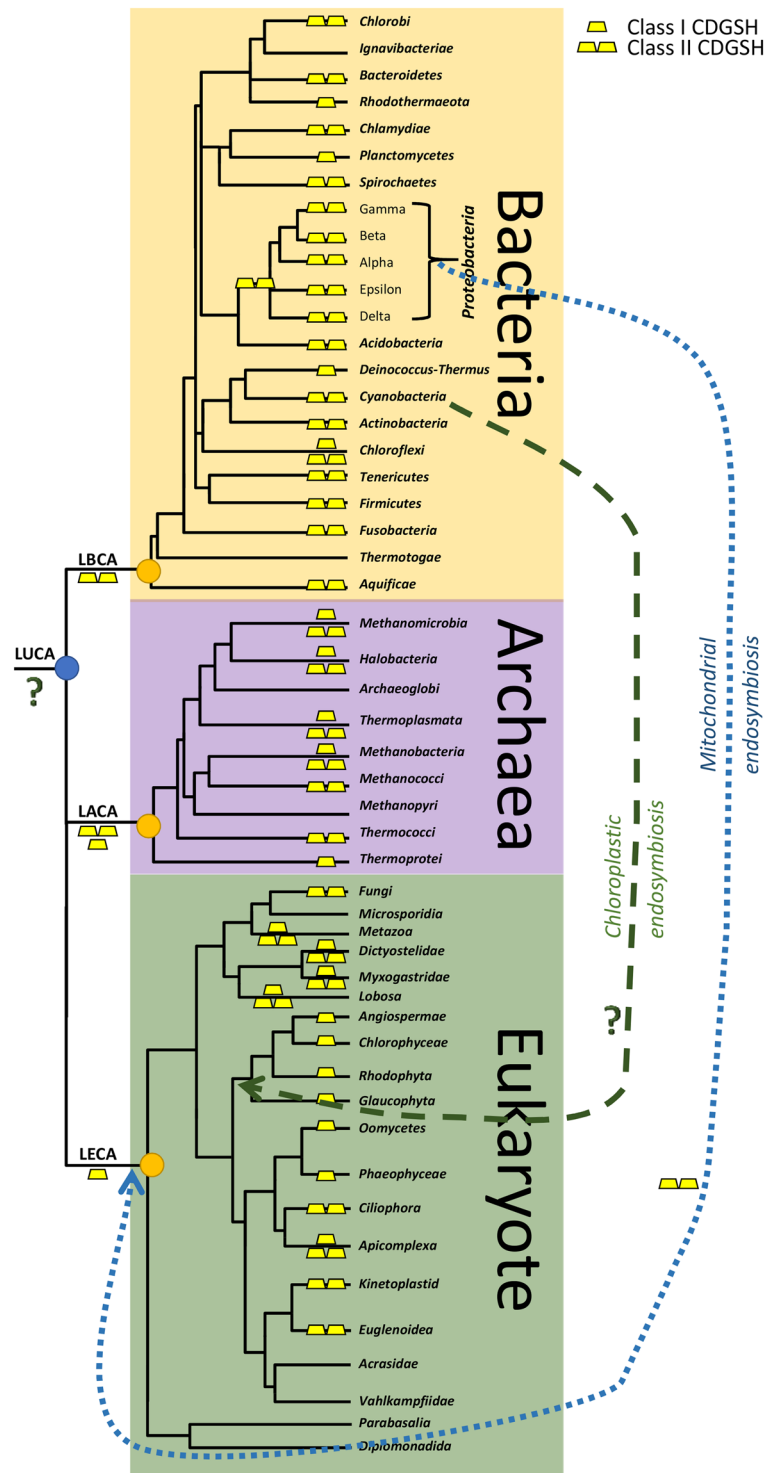


Figure 6. A hypothetical model for the evolution of CDGSH proteins. The occurrence of Class I and Class II CDGSH proteins is shown in major lineages from the 3 domains of life. Bacteria are shown to primarily contain Class II CDGSH proteins, whereas archaea and eukaryotes contain Class I and Class II CDGSH proteins. Eukaryote Class II CDGSH proteins are proposed to have originated from the endosymbiotic event that yielded the mitochondria. Class I CDGSH proteins are postulated to have evolved only after archaea and eukaryotes have radiated from bacteria. LUCA, last universal common ancestor, LACA, last archaeal common ancestor, LECA, last eukaryote common ancestor, LBCA, last bacterial common ancestor.

database using the same *Dictyostelium* CIRD sequence as the query. All sequences were again subjected to domain analysis using PFAM and sequences possessing at least one CDGSH domains were retained. A total of 191 archaeal sequences representing different lineages were selected for further analysis (Supplementary Table S2).

Sequence alignment. For the above three sets representing bacterial, archaeal, and eukaryotic CISD sequences, multiple sequence alignment was performed for each using command-line multiple alignment program MUSCLE with default options⁷³. trimAL was used (-automated1 option) to remove poorly aligned regions in order to obtain high quality alignments⁷⁴.

Divergence time estimation for eukaryotic sequences. Multiple sequence alignments of eukaryotic CISD sequences were analyzed using BEAST for the estimation of divergence times. Multiple combinations of population size change and molecular clock models were assessed in order to find the best-fit model. Among the models tested, the combination of a constant/exponential population size model and a relaxed uncorrelated log-normal clock with high estimated sample size (ESS) yielded the highest Bayes factor (Supplementary Table S3). Both selected models allowed the evolutionary rates to change among the branches of the tree and had the BLOSUM62 substitution model with γ correction for among-site rate variations⁶⁴. The time calibration points for each organism were obtained from the TimeTree website www.timetree.org⁵⁹.

All BEAST Monte Carlo Markov Chain (MCMC) simulations were run for at least 50 million steps, with subsampling at every 1,000 steps. The trees generated by BEAST were summarized by a single maximum clade credibility (MCC) tree using TreeAnnotator⁶⁴ with 20% of the MCMC steps discarded as burn-ins. Statistical uncertainty is represented by a 95% confidence interval (CI) calculated as the 95% highest posterior density (HPD) interval (upper-lower). The final MCC tree was visualized and edited with the program FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>)⁷⁵. The inferred time of divergence from an ancestral node is indicated next to each internal node in Fig. 5, Supplementary Figs S9–S10).

Maximum likelihood phylogenetic analysis of bacterial and archaeal sequences. PhyML version 3.0 was utilized to generate maximum-likelihood trees for bacterial and archaeal CISD sequences⁷⁶. For statistical reliability, the following tests were used: an approximate likelihood-ratio test (aLRT) based on logarithm of the ratio of likelihood computed for the current tree and that of the best alternative, and a Bayesian-like transformation of aLRT (aBayes). To estimate the optimal model of substitution, ProtTest was used for each alignment⁷⁷. ProtTest indicated the WAG amino acid model with gamma distribution shape parameter (WAG + G) and the WAG amino acid model with invariable gamma distribution shape parameter (WAG + I + G) as the best fitting models among the 112 examined evolutionary models, based on Akaike information criterion (AIC) statistics, for archaea and bacteria respectively. The trees were visualized and designed with iTOL (Interactive Tree of Life) web-server⁷⁸. The domain organization for each sequence was appended at the end of terminal branches, using iTOL, as shown in Supplementary Figs S1–S4.

Time-tree of organisms. To generate an evolutionary timescale for the bacterial and archaeal organisms represented in our analysis, we generated time-trees using TimeTree.org website. The complete lists of representative bacterial and archaeal organisms (Tables S1 and S2) were uploaded separately to generate a time-tree for each. The time-tree represents the estimated divergence time between species or groups of species or lineages based on literature records⁵⁹. However, the divergence times for multiple bacterial and archaeal organisms were not found in the TimeTree website. Nevertheless, we ensured that at least one organism from a class or phylum is represented in the time-tree.

Data availability statement. All data used in this study is publically available. All data or tools generated by this study will be made available upon request.

References

- Inupakutika, M. A. *et al.* Phylogenetic analysis of eukaryotic NEET proteins uncovers a link between a key gene duplication event and the evolution of vertebrates. *Scientific Reports* **7**, 42571, <https://doi.org/10.1038/srep42571> (2017).
- Lin, J., Zhang, L., Lai, S. & Ye, K. Structure and Molecular Evolution of CDGSH Iron-Sulfur Domains. *PLoS ONE* **6**, e24790, <https://doi.org/10.1371/journal.pone.0024790> (2011).
- Paddock, M. L. *et al.* (Protein Data Bank, Rutgers University, 2007).
- Tamir, S. *et al.* Structure–function analysis of NEET proteins uncovers their role as key regulators of iron and ROS homeostasis in health and disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1853**, 1294–1315, <https://doi.org/10.1016/j.bbamcr.2014.10.014> (2015).
- Wiley, S. E., Murphy, A. N., Ross, S. A., van der Geer, P. & Dixon, J. E. MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. *Proceedings of the National Academy of Sciences* **104**, 5318–5323, <https://doi.org/10.1073/pnas.0701078104> (2007).
- Colca, J. R. *et al.* Identification of a novel mitochondrial protein (“mitoNEET”) cross-linked specifically by a thiazolidinedione photoprobe. *American Journal of Physiology-Endocrinology and Metabolism* **286**, E252–E260 (2004).
- Mittler, R. *et al.* NEET proteins: A new link between iron metabolism, ROS and cancer. *Antioxidants & Redox Signaling* (2017).
- Ferecatu, I. *et al.* The Diabetes Drug Target MitoNEET Governs a Novel Trafficking Pathway to Rebuild an Fe-S Cluster into Cytosolic Aconitase/Iron Regulatory Protein 1. *Journal of Biological Chemistry* **289**, 28070–28086, <https://doi.org/10.1074/jbc.M114.548438> (2014).
- Geldenhuis, W. J., Leeper, T. C. & Carroll, R. T. mitoNEET as a novel drug target for mitochondrial dysfunction. *Drug Discovery Today* **19**, 1601–1606, <https://doi.org/10.1016/j.drudis.2014.05.001> (2014).
- Habener, A. *et al.* MitoNEET Protects HL-1 Cardiomyocytes from Oxidative Stress Mediated Apoptosis in an *In Vitro* Model of Hypoxia and Reoxygenation. *PLOS ONE* **11**, e0156054, <https://doi.org/10.1371/journal.pone.0156054> (2016).
- He, Q.-Q. *et al.* MicroRNA-127 targeting of mitoNEET inhibits neurite outgrowth, induces cell apoptosis and contributes to physiological dysfunction after spinal cord transection. *Scientific Reports* **6**, <https://doi.org/10.1038/srep35205> (2016).
- Hou, X. *et al.* Crystallographic Studies of Human MitoNEET. *Journal of Biological Chemistry* **282**, 33242–33246, <https://doi.org/10.1074/jbc.C700172200> (2007).
- Kusminski, C. M. *et al.* MitoNEET-Parkin Effects in Pancreatic α - and β -Cells, Cellular Survival, and Intracellular Cross Talk. *Diabetes* **65**, 1534–1555, <https://doi.org/10.2337/db15-1323> (2016).

14. Kusminski, C. M. *et al.* MitoNEET-driven alterations in adipocyte mitochondrial activity reveal a crucial adaptive process that preserves insulin sensitivity in obesity. *Nature Medicine* **18**, 1539–1549, <https://doi.org/10.1038/nm.2899> (2012).
15. Lin, J., Zhou, T., Ye, K. & Wang, J. Crystal structure of human mitoNEET reveals distinct groups of iron sulfur proteins. *Proceedings of the National Academy of Sciences* **104**, 14640–14645, <https://doi.org/10.1073/pnas.0702426104> (2007).
16. Salem, A. F., Whitaker-Menezes, D., Howell, A., Sotgia, F. & Lisanti, M. P. Mitochondrial biogenesis in epithelial cancer cells promotes breast cancer tumor growth and confers autophagy resistance. *Cell Cycle* **11**, 4174–4180, <https://doi.org/10.4161/cc.22376> (2012).
17. Sohn, Y.-S. *et al.* NAF-1 and mitoNEET are central to human breast cancer proliferation by maintaining mitochondrial homeostasis and promoting tumor growth. *Proceedings of the National Academy of Sciences* **110**, 14676–14681, <https://doi.org/10.1073/pnas.1313198110> (2013).
18. Takahashi, T. *et al.* A Novel MitoNEET Ligand, TT01001, Improves Diabetes and Ameliorates Mitochondrial Function in db/db Mice. *Journal of Pharmacology and Experimental Therapeutics* **352**, 338–345, <https://doi.org/10.1124/jpet.114.220673> (2014).
19. Yuan, H., Li, X., Zhang, X., Kang, R. & Tang, D. CISD1 inhibits ferroptosis by protection against mitochondrial lipid peroxidation. *Biochemical and Biophysical Research Communications* **478**, 838–844, <https://doi.org/10.1016/j.bbrc.2016.08.034> (2016).
20. Zuris, J. A. *et al.* Facile transfer of [2Fe-2S] clusters from the diabetes drug target mitoNEET to an apo-acceptor protein. *Proceedings of the National Academy of Sciences* **108**, 13047–13052, <https://doi.org/10.1073/pnas.1109986108> (2011).
21. Chang, N. C., Nguyen, M., Germain, M. & Shore, G. C. Antagonism of Beclin 1-dependent autophagy by BCL-2 at the endoplasmic reticulum requires NAF-1. *The EMBO journal* **29**, 606–618, <https://doi.org/10.1038/emboj.2009.369> (2009).
22. Chen, B. *et al.* CISD2 associated with proliferation indicates negative prognosis in patients with hepatocellular carcinoma. *International journal of clinical and experimental pathology* **8**, 13725 (2015).
23. Chen, Y. F. *et al.* Cisd2 deficiency drives premature aging and causes mitochondria-mediated defects in mice. *Genes & development* **23**, 1183–1194, <https://doi.org/10.1101/gad.1779509> (2009).
24. Conlan, A. R. *et al.* (Protein Data Bank, Rutgers University, 2009).
25. Du, X. *et al.* NAF-1 antagonizes starvation-induced autophagy through AMPK signaling pathway in cardiomyocytes. *Cell Biology International* **39**, 816–823, <https://doi.org/10.1002/cbin.10453> (2015).
26. Ge, Y.-Z. *et al.* Pathway analysis of genome-wide association study on serum prostate-specific antigen levels. *Gene* **551**, 86–91, <https://doi.org/10.1016/j.gene.2014.08.044> (2014).
27. Holt, S. H. *et al.* Activation of apoptosis in NAF-1-deficient human epithelial breast cancer cells. *Journal of Cell Science* **129**, 155–165, <https://doi.org/10.1242/jcs.178293> (2015).
28. Liu, L. *et al.* CISD2 expression is a novel marker correlating with pelvic lymph node metastasis and prognosis in patients with early-stage cervical cancer. *Medical Oncology* **31**, <https://doi.org/10.1007/s12032-014-0183-5> (2014).
29. Tamir, S. *et al.* Integrated strategy reveals the protein interface between cancer targets Bcl-2 and NAF-1. *Proceedings of the National Academy of Sciences* **111**, 5177–5182, <https://doi.org/10.1073/pnas.1403770111> (2014).
30. Tamir, S. *et al.* Nutrient-Deprivation Autophagy Factor-1 (NAF-1): Biochemical Properties of a Novel Cellular Target for Anti-Diabetic Drugs. *PLoS ONE* **8**, e61202, <https://doi.org/10.1371/journal.pone.0061202> (2013).
31. Wang, L. *et al.* Overexpressed CISD2 has prognostic value in human gastric cancer and promotes gastric cancer cell proliferation and tumorigenesis via AKT signaling pathway. *Oncotarget* **7**, 3791–3805, <https://doi.org/10.18632/oncotarget.6302> (2015).
32. Yang, L. *et al.* A novel prognostic score model incorporating CDGSH iron sulfur domain 2 (CISD2) predicts risk of disease progression in laryngeal squamous cell carcinoma. *Oncotarget* **7**, 22720–22732, <https://doi.org/10.18632/oncotarget.8150> (2016).
33. Yang, Y., Bai, Y.-S. & Wang, Q. CDGSH Iron Sulfur Domain 2 Activates Proliferation and EMT of Pancreatic Cancer Cells via Wnt/ β -Catenin Pathway and Has Prognostic Value in Human Pancreatic Cancer. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* **25**, 605–615, <https://doi.org/10.3727/096504016x14767450526417> (2017).
34. Amr, S. *et al.* A Homozygous Mutation in a Novel Zinc-Finger Protein, ERIS, Is Responsible for Wolfram Syndrome 2. *The American Journal of Human Genetics* **81**, 673–683, <https://doi.org/10.1086/520961> (2007).
35. Danielpur, L. *et al.* GLP-1-RA Corrects Mitochondrial Labile Iron Accumulation and Improves β -Cell Function in Type 2 Wolfram Syndrome. *The Journal of Clinical Endocrinology & Metabolism* **101**, 3592–3599, <https://doi.org/10.1210/jc.2016-2240> (2016).
36. Lu, S. *et al.* A calcium-dependent protease as a potential therapeutic target for Wolfram syndrome. *Proceedings of the National Academy of Sciences* **111**, E5292–E5301, <https://doi.org/10.1073/pnas.1421055111> (2014).
37. Mozzillo, E. *et al.* A novel CISD2 intragenic deletion, optic neuropathy and platelet aggregation defect in Wolfram syndrome type 2. *BMC Medical Genetics* **15**, <https://doi.org/10.1186/1471-2350-15-88> (2014).
38. Wiley, S. E. *et al.* Wolfram Syndrome protein, Miner1, regulates sulphhydryl redox status, the unfolded protein response, and Ca²⁺ homeostasis. *EMBO molecular medicine* **5**, 904–918, <https://doi.org/10.1002/emmm.201201429> (2013).
39. Nechushtai, R. *et al.* Characterization of Arabidopsis NEET Reveals an Ancient Role for NEET Proteins in Iron Metabolism. *The Plant Cell* **24**, 2139–2154, <https://doi.org/10.1105/tpc.112.097634> (2012).
40. Hart, T. *et al.* High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
41. Bak, D. W. & Elliott, S. J. Conserved Hydrogen Bonding Networks of MitoNEET Tune Fe-S Cluster Binding and Structural Stability. *Biochemistry* **52**, 4687–4696, <https://doi.org/10.1021/bi400540m> (2013).
42. Benson, S. K. *et al.* Multinuclear NMR and UV-Vis spectroscopy of site directed mutants of the diabetes drug target protein mitoNEET suggest that folding is intimately coupled to iron-sulfur cluster formation. *Inorganic Chemistry Communications* **63**, 86–92, <https://doi.org/10.1016/j.inoche.2015.11.022> (2016).
43. Bergner, M. *et al.* Model of the MitoNEET [2Fe-2S] Cluster Shows Proton Coupled Electron Transfer. *Journal of the American Chemical Society* **139**, 701–707, <https://doi.org/10.1021/jacs.6b09180> (2017).
44. Boyd, E. S., Thomas, K. M., Dai, Y., Boyd, J. M. & Outten, F. W. Interplay between Oxygen and Fe-S Cluster Biogenesis: Insights from the Suf Pathway. *Biochemistry* **53**, 5834–5847, <https://doi.org/10.1021/bi500488r> (2014).
45. Golinelli-Cohen, M.-P. *et al.* Redox Control of the Human Iron-Sulfur Repair Protein MitoNEET Activity via Its Iron-Sulfur Cluster. *Journal of Biological Chemistry* **291**, 7583–7593, <https://doi.org/10.1074/jbc.m115.711218> (2016).
46. Landry, A. P., Cheng, Z. & Ding, H. Reduction of mitochondrial protein mitoNEET [2Fe-2S] clusters by human glutathione reductase. *Free Radical Biology and Medicine* **81**, 119–127, <https://doi.org/10.1016/j.freeradbiomed.2015.01.017> (2015).
47. Landry, A. P. & Ding, H. Redox Control of Human Mitochondrial Outer Membrane Protein MitoNEET [2Fe-2S] Clusters by Biological Thiols and Hydrogen Peroxide. *Journal of Biological Chemistry* **289**, 4307–4315, <https://doi.org/10.1074/jbc.m113.542050> (2014).
48. Lipper, C. H. *et al.* Cancer-Related NEET Proteins Transfer 2Fe-2S Clusters to Anamorsin, a Protein Required for Cytosolic Iron-Sulfur Cluster Biogenesis. *PLOS ONE* **10**, e0139699, <https://doi.org/10.1371/journal.pone.0139699> (2015).
49. Roberts, M. E. *et al.* Identification of Disulfide Bond Formation between MitoNEET and Glutamate Dehydrogenase 1. *Biochemistry* **52**, 8969–8971, <https://doi.org/10.1021/bi401038w> (2013).
50. Tamir, S. *et al.* A point mutation in the [2Fe-2S] cluster binding region of the NAF-1 protein (H114C) dramatically hinders the cluster donor properties. *Acta Crystallographica Section D Biological Crystallography* **70**, 1572–1578, <https://doi.org/10.1107/s1399004714005458> (2014).
51. Tan, G. *et al.* His-87 ligand in mitoNEET is crucial for the transfer of iron sulfur clusters from mitochondria to cytosolic aconitase. *Biochemical and Biophysical Research Communications* **470**, 226–232, <https://doi.org/10.1016/j.bbrc.2016.01.040> (2016).
52. Darash-Yahana, M. *et al.* Breast cancer tumorigenesis is dependent on high expression levels of NAF-1 and the lability of its Fe-S clusters. *Proceedings of the National Academy of Sciences* **113**, 10890–10895, <https://doi.org/10.1073/pnas.1612736113> (2016).

53. Bai, F. *et al.* The Fe-S cluster-containing NEET proteins mitoNEET and NAF-1 as chemotherapeutic targets in breast cancer. *Proceedings of the National Academy of Sciences*, 201502960, <https://doi.org/10.1073/pnas.1502960112> (2015).
54. Geldenhuys, W. J. *et al.* Identification of small molecules that bind to the mitochondrial protein mitoNEET. *Bioorganic & Medicinal Chemistry Letters* **26**, 5350–5353, <https://doi.org/10.1016/j.bmcl.2016.09.009> (2016).
55. Roche, B. *et al.* Reprint of: Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1827**, 923–937, <https://doi.org/10.1016/j.bbabi.2013.05.001> (2013).
56. Vinella, D., Brochier-Armanet, C., Loiseau, L., Talla, E. & Barras, F. Iron-Sulfur (Fe/S) Protein Biogenesis: Phylogenomic and Genetic Studies of A-Type Carriers. *PLoS Genetics* **5**, e1000497, <https://doi.org/10.1371/journal.pgen.1000497> (2009).
57. Weiss, M. C. *et al.* The physiology and habitat of the last universal common ancestor. *Nature Microbiology* **1**, 16116, <https://doi.org/10.1038/nmicrobiol.2016.116> (2016).
58. Xu, X. M. & Möller, S. G. Iron–Sulfur Clusters: Biogenesis, Molecular Mechanisms, and Their Functional Significance. *Antioxidants & Redox Signaling* **15**, 271–307, <https://doi.org/10.1089/ars.2010.3259> (2011).
59. Hedges, S. B. & Kumar, S. *The timetree of life*. (OUP Oxford, 2009).
60. Kurland, C. G. & Andersson, S. G. E. Origin and Evolution of the Mitochondrial Proteome. *Microbiology and Molecular Biology Reviews* **64**, 786–820, <https://doi.org/10.1128/mmb.64.4.786-820.2000> (2000).
61. Pittis, A. A. & Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature*, <https://doi.org/10.1038/nature16941> (2016).
62. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576–4579, <https://doi.org/10.1073/pnas.87.12.4576> (1990).
63. Karmi, O. *et al.* Interactions between mitoNEET and NAF-1 in cells. *PLOS ONE* **12**, e0175796, <https://doi.org/10.1371/journal.pone.0175796> (2017).
64. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS biology* **4**, e88 (2006).
65. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* **7**, 214 (2007).
66. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* **29**, 1969–1973 (2012).
67. Inupakutika, M. A., Sengupta, S., Devireddy, A. R., Azad, R. K. & Mittler, R. The evolution of reactive oxygen species metabolism. *Journal of Experimental Botany* **67**, 5933–5943, <https://doi.org/10.1093/jxb/erw382> (2016).
68. Lipper, C. H. *et al.* Structure of the human monomeric NEET protein Cisd3/MiNT and its role in regulating iron and ROS in cancer cells. *Proceedings of the National Academy of Sciences*. **Forthcoming** (2017).
69. Oren, A., Gurevich, P., Gemmel, R. T. & Teske, A. Halobaculum gomorrense gen. nov., sp. nov., a novel extremely halophilic archaeon from the Dead Sea. *International Journal of Systematic and Evolutionary Microbiology* **45**, 747–754 (1995).
70. Vasilyeva, L. V. *et al.* Asticcacaulis benevestitus sp. nov., a psychrotolerant, dimorphic, prosthecate bacterium from tundra wetland soil. *International journal of systematic and evolutionary microbiology* **56**, 2083–2088 (2006).
71. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389–3402 (1997).
72. Finn, R. D. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics* (John Wiley & Sons, Ltd, 2004).
73. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).
74. 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).
75. Rambaut, A. Fig Tree version 1.4.0. Available at <http://tree.bio.ed.ac.uk/software/figtree> (2012).
76. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307–321 (2010).
77. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
78. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *science* **311**, 1283–1287 (2006).

Acknowledgements

This work was supported by the National Science Foundation IOS-1557787 awarded to P.P. and R.M., the National Science Foundation MCB-1613462 awarded to R.M., R.N. and R.K.A., Israel Science Foundation - ISF 865/13 awarded to R.N., the National Institutes of Health DK54441 awarded to P.A.J., and funds from the University of North Texas College of Arts and Sciences awarded to P.P., R.M. and R.K.A. J.N.O. was supported by the Cancer Prevention and Research Institute of Texas (CPRIT - grant R1110), by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY- 1427654) and by NSF- CHE 1614101. The funders had no role in the design, data collection, analysis, decision to publish, or preparation of the manuscript.

Author Contributions

S.S.G. performed the experiments and analyzed the data, P.P., R.K.A., P.A.J., J.N.O. and R.M. analyzed the data and designed experiments. S.S.G., P.P., R.N., P.A.J., R.K.A. and R.M. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-23305-6>.

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018