Modular structure of complex II: An evolutionary perspective

Val Karavaeva, Filipa L. Sousa

Department of Functional and Evolutionary Ecology, University of Vienna, Djerassiplatz 1, 1030 Wien, Austria

ARTICLE INFO

Keywords:
Succinate dehydrogenase
Fumarate reductase
Comparative genomics
Succinate:Quinone oxidoreductase
Quinol:Fumarate reductase
Electron transport chain

ABSTRACT

Succinate dehydrogenases (SDHs) and fumarate reductases (FRDs) catalyse the interconversion of succinate and fumarate, a reaction highly conserved in all domains of life. The current classification of SDH/FRDs is based on the structure of the membrane anchor subunits and their cofactors. It is, however, unknown whether this classification would hold in the context of evolution. In this work, a large-scale comparative genomic analysis of complex II addresses the questions of its taxonomic distribution and phylogeny. Our findings report that for types C, D, and F, structural classification and phylogeny go hand in hand, while for types A, B and E the situation is more complex, highlighting the possibility for their classification into subgroups. Based on these findings, we proposed a revised version of the evolutionary scenario for these enzymes in which a primordial soluble module, corresponding to the cytoplasmatic subunits, would give rise to the current diversity via several independent membrane anchor attachment events.

1. Introduction

Succinate dehydrogenases (succinate:quinone oxidoreductases, SDHs, EC 1.3.5.1) are membrane complexes that catalyse the conversion of succinate to fumarate and link chemiosmotic coupling to carbon metabolism via the TCA cycle [1]. These complexes are closely related and homologous to fumarate reductases (quinol:fumarate reductases, FRD, EC 1.3.5.4) [2]. While succinate dehydrogenases oxidize succinate to fumarate, reducing usually high potential ubiquinone to ubiquinol, fumarate reductases reduce fumarate to succinate, usually oxidizing low potential menaquinol back to a menaquinone [3–5].

1.1. Function of SDH/FRD enzymes

These enzymes are part of aerobic and anaerobic respiratory electron transport chains (ETCs) as well as the only membrane component of the TCA cycle, and have been extensively studied since the beginning of the 20th century [4,6–72]. They are anchored in the cytoplasmic membrane of prokaryotes or in the inner mitochondrial membrane of eukaryotes, with the catalytic domain in the cytoplasm or mitochondrial matrix side, respectively [4]. SDH complexes might enhance the proton gradient by supplying reducing equivalents from succinate metabolism [73,74]. In addition, it is proposed that some SDH/FRD complexes might be able to translocate protons across the membrane [73–76], further contributing to the establishment of pmf, or on the contrary, as in the case of Bacillus subtilis [54,55,76–78], dissipate pmf by operating in reverse. In aerobic mitochondrial-like ETC, these reducing equivalents are transported through the ubiquinone pool to complexes III and IV, and these complexes, in turn, extrude/pump protons [79].

1.2. Subunit composition and current classification

Functionally, SDH/FRDs form three classes based on the reaction they perform in vivo and the type of the quinone they use: class 1 SDHs oxidize succinate and reduce a high-potential quinone (e.g. ubiquinone); class 2 FRDs reduce fumarate using a low-potential quinol (e.g. menaquinol), and class 3 enzymes oxidize succinate with the help of a low-potential quinone [2]. So far, it is not possible based on the primary sequence alone to identify the reaction a certain SDH/FRD enzyme would catalyse without in vivo tests [2–5].

Structurally, SDH/FRD complexes have a variable number of subunits, with prokaryotic complexes being composed of three to four subunits and eukaryotic complexes having between four and twelve subunits as in the case of Trypanosoma [80,81]. Interestingly, within Viridiplantae the number of subunits is not conserved, with reports of Brassicaceae and monocotyledonous plants having 8 different subunits while other Embryophyta having only 7 of those subunits, and Chlorophyta only containing four traditionally conserved subunits [82–84]. In this analysis, we only focus on the three to four modules that are conserved across the three domains of life, and can be divided into the
catalytic or cytoplasmic part (SdhA/FrdA and SdhB/FrdB) and the anchor module composed of one or two subunits [3] (Fig. 1). Subunit A is a flavoprotein that contains a dicarboxylate binding site where the succinate to fumarate conversion takes place. This subunit is soluble and exposed to the cytoplasm, and contains one FAD cofactor (covalently bound in most organisms) [4]. The FAD group serves as the first electron acceptor and passes electrons onto the other subunits [2,5]. Electrons flow from FAD into the next electron-accepting soluble subunit (SdhB) that contains three iron-sulfur centers with different compositions: S1 center ([2Fe–2S]$^{2+}$), S2 center ([4Fe–4S]$^{2+}$), and S3 ([3Fe–4S]$^{1+}$) [4]. In the succinate oxidation reaction, the S1 center is the first to accept electrons from FAD [2,5].

In case of fumarate reduction, the order of electron acceptors is reversed. The anchor part is composed of membrane subunits with varied cofactor content and structural motifs that served to establish the structural classification of this protein family [2,3,5,85]. For convenience, the structural type information is given as subscript in the subunit abbreviation (e.g. SdhA of type C is indicated as SdhA$_C$). SDH/FRDs belonging to structural type A contain two separate membrane subunits (SdhC$_A$ and SdhD$_A$), both with three transmembrane helices where two heme groups - a high redox midpoint potential heme group (b$_A$) and a low redox midpoint potential heme group (b$_B$) are bound [76]. For a long time, enzymes of this type were assumed to be a hallmark of archaea [3,71], although later, bacterial type A complexes were characterised as well [86–88]. Interestingly, the available structure of M. smegmatis SDH2 of type A indicates a presence of a potential small third anchor subunit, titled SdhF$_A$ by the authors [88]. This subunit is 32 amino acids long and bears no relation to the SdhF$_B$ first characterised in Acidianus ambivalens [89].

So far, enzymes of this type have been known to only have the succinate dehydrogenases activity, since at least in the conditions tested with the Mycobacterium smegmatis SDH2, fumarate reductase activity was not determined. [4,87]. However, due to the low number of studies measuring fumarate reductase activity in type A enzymes, it is not clear if type A enzymes function strictly in one direction.

In type B enzymes, on the other hand, only one large membrane subunit (SdhC$_B$) with five transmembrane helices is found. Similarly to type A, type B also binds two hemes (b$_B$ and b$_B$). Type B enzymes were shown to be able to catalyse the reaction in both directions, i.e. being either SDHs, FRDs, or bifunctional depending on the in vivo function [4,74,76–78,90–96].

Type C and D enzymes are very similar to type A, with differences relying on the number of hemes groups: type C contains only one heme group (b$_C$) and in type D, no heme groups are present. The existence of functional complexes devoid of hemes questions the functional role of these cofactors. The well-studied E.coli SDH enzyme [10,15,37,51,97–103] belongs to the structural type C, and the E.coli FRD enzyme [20,24,47,104–106] belongs to the type D. While both type C and D complexes were shown to be able to perform the succinate to fumarate conversion in both directions, in wild type E.coli cells type C SDH usually acts as a succinate dehydrogenase, while type D FRD acts as a fumarate reductase [2,4,99,107]. Interestingly, it also has been shown that in E.coli type D FRD (specifically, the FrdA subunit) not only participates in anaerobic respiration but also under aerobic conditions contributes to the shift of the direction in rotation of flagella [108,109], by interacting with FliG, a protein that is a part of the flagellar switch complex, and is responsible for switching to clockwise direction.

In the context of structural classification, so far all characterised eukaryotic enzymes belong to type C [21,41,81,82,110–112]. However, it is worth noting that besides numerous mutations within eukaryotic

---

**Fig. 1.** Schematic representation of SDH/FRD structural types with their respective crystallographic structure. SdhA is coloured in green, SdhB in purple, SdhC in pastel pink, and SdhD in light blue. SdhE$_A$ and SdhF$_A$ are coloured in grey. SdhC$_B$ is represented in a darker blue to indicate lack of homology to canonical SdhCs. Cofactors and numbers of membrane helices are indicated as described in the figure. The symbols “+” and “−” indicate the cytoplasmic and periplasmic sides of the membrane. X-Ray crystallography structure of type A is represented by the structure of M. smegmatitis succinate dehydrogenase 2 (PDB: 6LUM) [134]; type B by W. succinogenes fumarate reductase (PDB: 2BS2) [176]; type C by E.coli succinate dehydrogenase (PDB: 1NEK) [103]; type D by E.coli fumarate reductase (PDB: 3P4P) [105], and type F by M. smegmatitis succinate dehydrogenase 1 (PDB: 7D6V) [88]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
SDH enzymes that lead to various diseases [113,114], eukaryotic enzymes exhibit additional diversity, such as expressing different isoforms at different developmental stages [115–119] or in different tissues [120], having a subunit split into two [80], having a Tyr instead of His to coordinate heme in the anchor module [121], as well as having additional subunits in the complex [80,82–84,122]. For most of these additional subunits, in plants as well as Trypanosoma, the function is not known [80,83,84]. However, for plant subunits Sdh6 and Sdh7, it is proposed that they compensate for the lost helices in subunits Sdh3 and Sdh4 [84]. In what regards the common modules, eukaryotic enzymes across large phylogenetic distances are reported to have a high degree of identity (~80 %) for the FAD and iron-sulfur subunits [123], while as in prokaryotes, their anchor subunits have low identity [123,124].

In 2001, a new type of SDH complexes (type E) was functionally characterised from the membranes of *Acidilus ambivalens* [89,125]. This type, known to be present in some Sulfolobales, is more dissimilar to the types described so far, having two membrane subunits with amphipathic helices, named SdhF and SdhG [89], which do not bind heme groups. Instead, the SdhE/F complex possesses two Cysteine-rich domains, whereby one binds a [4Fe-4S] center [125] and the other serves as a zinc binding site [126]. So far, no cofactor was identified in SdhF and SdhG [125]. The subunit nomenclature of type E enzymes proposed in 2001 [89] overlaps with the nomenclature for succinate dehydrogenase assembly factor SdhE from *E.coli* [127,128]. Therefore, to avoid further confusion, this paper will follow the historical naming of the subunits and use the SdhE to refer to the anchor subunit, and not to the SdhE assembly factor, which is not the focus of this paper. In addition, its subunit B contains a second [4Fe-4S] center instead of a [3Fe-4S] center [89].

The *A. ambivalens* and the *S. acidocaldarius* enzymes most likely in vivo catalyse the succinate oxidation to fumarate [89,129], since both organisms have Caldariella quinone with a redox potential of +103 mV [130], which would create thermodynamic barriers for the reverse reaction, although in *vitro* at least the *S. acidocaldarius* enzyme is reversible [128] As aside from Sulfolobales, other organisms, such as Cyanobacteria and Aquificae, reportedly have SDH/FRD complexes (in here marked as type E) containing a membrane anchor subunit SdhE but lacking subunit SdhF [89,125]. In this type is also included the homologous methylmenaquinol:fumarate reductase complex (MFR; [131]), which is located in periplasm and is upregulated under high oxidative conditions [132].

Recently, a new structural type, type F, has been proposed [133]. So far, only one representative of this type has been characterised, the SDH1 from *Mycobacterium smegmatis* [88]. It is worth noting, that in addition to SDH1 of type F, this organism has SDH2 of type A [134], while other Mycobacterium organisms may in addition have type D enzymes as well [135]. Type F is characterised by having only one transmembrane subunit with five predicted transmembrane helices with no bound hemes but with a Rieske FeS center cofactor [88], and with no detected similarity to the membrane subunits present in types A to E. Similarly, to the transmembrane subunit from type B, which is assumed to be the result of a fusion event between subunits SdhC and SdhD [136], Hards et al proposed that type F membrane anchor subunit rearrangement is the result of a potential fusion, this module being most similar to the type D anchor subunits [133].

1.3. Homology relationships of SDH/FRD complexes

The SDH/FRD cytoplasmic subunits have homologues in other enzymes that share the same domains, such as *i*-aspartate oxidase (NadB), which contains a flavoprotein subunit homologous to SdhA [137], or the anaerobic *sn*-glyceraldehyde-3-phosphate dehydrogenase, in which GlpB is homologous to SdhA and GlpC is homologous to SdhB [138]. Also the adenosine-5′-phosphosulfate reductase subunit A (AprA) shares homology with SdhA [85], and in thiol:fumarate reductases, soluble fumarate reductases that use Coenzyme M and Coenzyme B as electron donors, subunit A (TfrA) is homologous to SdhA, while TfrB is homologous to both SdhB and SdhE [139]. Finally, the glycolate oxidase, which catalyzes the oxidation of glycolate to glyoxylate, contains two subunits (GlcD and GlcE) that are homologous to SdhA, and the subunit GlcF that shares homology to SdhB and SdhE [140]. These homologous relationships lead Jardim-Messeder et al to propose the classification of the “fumarate reductase superfamily” [85].

1.4. Re-evaluation of SDH/FRD current diversity

The increase in the number of genomes available combined with the important role of SDH/FRDs within the metabolism of the three domains of life calls for a re-evaluation of its distribution and evolution. In this article, we analysed the taxonomic distribution of the different SDH/FRD types and observed that with few exceptions, the different types are not confined to a specific taxonomic group. The phylogenetic and similarity network analyses showed the existence of several prokaryotic events of membrane anchor replacements occurring within the history of these complexes, with eukaryotic diversity arising by the existence of isoforms and additional subunits that increase the complexity of the enzyme. Finally, the underlying analysis can serve as basis for selecting natural variants as candidates for future functional studies.

2. Materials and methods

2.1. Query dataset

A thorough literature search allowed gathering information regarding microorganisms containing characterised or genomic reported succinate dehydrogenases/fumarate reductase complexes (Table S1). Query sequences of succinate dehydrogenases and homologous enzymes sequences were retrieved from BREnda (release 2020.2, [141]), KEGG (release 95.0, [142]) and UniProt [143] databases or from internal databases as in the case of heterodisulfide reductases and adenosine-5′-phosphosulfate reductases [144]. Protein complexes were checked for their completeness and a search for missing subunits was performed by BLASTing all queries against the genomes of the complex with missing subunit(s), using as cut-off identity of at least 70 %, query coverage of 50 % or higher and an E-value lower than 10^-10. Retrieved hits were further analysed using Pfam (PfamScan.pl version 1.5 [145]) and TMHMM (version 2.0 [146]). The results were analysed and sequences added to the query list if the Pfam domains and the number of predicted transmembrane helices matched the expected. In addition, the feature tables from the respective genomes were inspected to determine if the missing subunits were classified as “pseudogenes” and thus, absent from the proteomic assembly. This affected one case, in the genome of *Sulfolobus acidocaldarius* DSM 639. In total, 69 SDH and FRD complexes belonging to 59 organisms were gathered, spanning 10 bacterial phyla and three archaeal phyla. Queries included both succinate dehydrogenases and fumarate reductases complexes as well as epsilon-proteobacterial methylmenaquinol:fumarate reductases from *Campylobacter* (C.) *jejuni* and *Wolinella* (W.) *succinogenes*, sequences from homologous enzymes representatives from *i*-aspartate oxidase [137,147–150], thiol:fumarate reductase [139,151], anaerobic *sn*-glyceraldehyde-3-phosphate dehydrogenase [138,152], glycolate oxidase [140,153,154], adenosine-5′-phosphosulfate reductase [155,156], heterodisulfide reductases HdrABC [156-158] and HdrDE [159]. In total, 293 query sequences were used for further analysis (Table S1). In addition, eukaryotic SDH sequences from 41 representative (Table S2) were acquired from KEGG (release 95.0, [142]).

2.2. Genomic dataset

A subset of our in-house dataset of over 190,000 metagenomic assemblies (download from NCBI November 2019 with two *Acidianus ambivalens* assemblies added at a later date [144]) was created by
filtering genomic records based on previously mapped NCBI taxonomic information and genomic quality in terms of completeness and contamination calculated with the Rinke method [160]. For this study, all NCBI reference and representative genomes as well as genomes containing queries were kept. Moreover, to ensure the existence of at least one representative from each species, additional genomic records (one per species) were added given preference for complete genomes followed by higher quality assemblies. The finalized genomic dataset contained 35,017 (meta)genomic assemblies, from which 33,683 belong to 179 bacterial phyla and 1334 to 22 archaeal phyla (Table S3).

2.3. Similarity analysis

Similarity searches were performed using the reciprocal best blast hit approach (rBBH, [161]). The search was conducted using DIAMOND Blastp (v2.0.4.142) [162,163] in “ultra-sensitive” mode with “-k 0”. The first direction of rBBH consisted of blasting each protein from the genomes of the dataset against a database composed of all query sequences as one DIAMOND database using as cut-offs 25% identity and an E-value lower than 10^{-10}. Copies were identified by DIAMOND BLASTing each genome against itself, and filtering the results for identity higher than 70%, E-value lower than 10^{-10}, and at least 70% query coverage. Retrieved hits (including copies) were then blasted against a DIAMOND database of query genomes using as cut-offs 25% identity and an E-value lower than 10^{-8}. The E-value cut-off for the second direction was increased to account for the larger size of the database, which could lead to increased E-values. The 201,016 unique reciprocal hits and their copies were retrieved from the genomes and first an all vs all BLAST was performed using NCBI BLASTP+ [164] and filtered for 25% identity and 10^{-12} E-value followed by a global alignment using Needleall [165] (same identity cut-off). The rBBH relationships were clustered using Markov Chain Clustering (MCL, version 14–137, [166,167]) with an inflation parameter of 1.2. This value was chosen instead of the default 2.0 to account for possible over-clustering artefacts which were encountered during test runs (data not shown). The cause of these artefacts was not established with certainty; however, it is possible that MCL algorithm is not currently adapted to work with large metagenomic data (over 270,000 hits), since this data often contains partial or mis-assembled sequences that can introduced errors in the clustering procedure. Intercluster mean, median and maximum identities were calculated and analysed via hierarchical clustering and heatmaps plotted using R (R pheatmap package, version 1.0.12, [168,169] , and R gplots package, version 3.1.2) [170]. To reduce redundancy, clusters with over 1000 SDH/FRD sequences were reclustered in MCL (version 14–137, [166,167]) using as cut-off a global identity of at least 90%. A representative sequence per cluster was kept for further analysis.

2.4. Functional annotations and syntenic rearrangements

The rBBH sequences were functionally annotated using the NCBI Conserved Domain Batch SEARCH (abbreviated as CD SEARCH; CDD database, in automatic search mode with E-value lower than 0.01 as threshold, composition corrected scoring ON, maximum number of hits = 500, including retired sequences, standard results mode [171]). Sequences were labelled as fusions if two or more non-overlapping CD SEARCH domains characteristic of different subunits/proteins were found. Prediction of the number of transmembrane helices was performed with TMHMM (version 2.0 [146]) and TMPred [172]. For a clearer differentiation between ApRA and SDHα subunits, DSCO was used [144]. In addition, KOFAM (kofam, scan version 1.3.0, [142]; using HHMNER version 3.2.1, hmmer.org] and Pfam (PfamScan.pl, version 1.5, [145]) annotations were performed for all genomic records and filtered for hits of interest. Sequences with multiple significant KOs assignments were checked based on their CD SEARCH and Pfam annotations (where possible) since kofam,scan output does not include start and end positions of KOs assignments.

Analysis of the syntenic arrangement of the retrieved sequences was performed with the feature table information using a window of two upstream and two downstream proteins in the neighbourhood of the proteins of interest (same chromosome or contig). Neighbour sequences not previously identified as rBBH were functionally annotated as above. The syntenic patterns of SDH subunits were analysed, and this information was used for the functional annotation of MCL clusters.

2.5. Phylogenetic and network analyses

Multiple sequence alignments of clusters containing SDH subunits were performed using ClustalOmega (version 1.2.4, [173]) with: “-max-guidetree-iterations=100 -max-hmm- iterations=100 -output-order=tree-order” as parameters. A structural alignment of SdhCF and SdhCB sequences was performed using Expresso mode of T-Coffee [174,175] and W. succinogenes structure, (PDB code: 2BS2) [176], ClustalOmega alignments were trimmed using TrimAl (version 1.2, [177]) with a gap threshold of 0.05 and a minimum of 60% of the positions in the original alignment conserved. Clusters containing sequences of the same subunits were pooled together and a joined multiple sequence alignment produced. Sequences in which SDH subunit fusions had been identified were manually split. In cases where the fusion involved additional domains unrelated with the SDH/FRD complexes, these were trimmed from the alignment using information from the CD SEARCH domain assignments. Alignments of clusters containing transmembrane subunits were checked for the presence of the heme binding histidines and this information stored at each sequence level to aid in the classification into types. Clusters of type E and F membrane anchor subunits were kept separated due to lack of homology to the other membrane subunits. Type B membrane anchor sequences were split into “SdhC” and “SdhD” based on the number of helices predicted by TMHMM [2,146]. The alignment quality was assessed using information about the conserved catalytic and cofactor binding residues in SDH/FRD subunits retrieved from literature and further verified by analysing the available SDH/FRD structures using UCSF Chimera (version 1.14, [178]). Additional alignments were performed including eukaryotic sequences.

The resulting alignments were used to reconstruct maximum likelihood phylogenies in IQtree (version 2.1.2 [179]) with 1000 ultrafast bootstraps [180] and best model selection “-m TEST” [181]. Phylogenetic reconstructions were rooted using the minimal ancestor deviation (MAD) method (version 2.22, [182]) with a modified script to keep bootstrap values (kindly provided by Giddy Landan; Newick files S01 and S02, as well as their corresponding annotations S03 and S04, are provided in Supplementary information). Functional and taxonomic annotations were added to the phylogenies and the analyses performed in FigTree (version v.1.4.4, tree.bio.ed.ac.uk/software/figtree). Similarity networks based on global identities of SdhC and SdhD were visualised in Cytoscape [183]. Eukaryotic SdhC and SdhD similarity networks were kept separate due to the absence of similarity above 30% with prokaryotic enzymes. The global identity relationships were reduced by 70% keeping one representative sequence per genus. Additionally, all relationships below 30% were excluded from the networks. Similarity networks of SdhE proteins were analysed together with HdrB sequences.

2.6. Classification of SDH/FRD types

SDH/FRD types were determined using a combinatorial analysis of synteny, best hit relationships, global identity to queries with a defined type, number of histidines in the anchor module, phylogenetic and/or similarity network analysis. Syntenic complete complexes where classified based on the number and type of membrane subunits, taking into account the number of histidines present within the membrane anchor. For cases in which SDH/FRD catalytic subunits were non-syntenic or in which the membrane anchor(s) were not identified, the best reciprocal hit was inspected and a global identity matrix of the SDH/FRD subunit in
question was created and hierarchically clustered. Cases where a type could not be confidently determined were marked as “N.d.”.

The distinction between SdhE₈ andHdrB proteins included analysis of the syntenic region for presence and/or absence of HdrA orHdrC proteins. First, If the catalytic subunits were found in the genome but not in synteny with the potential SdhE₈ sequence, it was checked first for presence of HdrA andHdrC. If those were absent, the sequence was labelled as SdhE₈. In cases where the genome contained more than one HdrB sequence, the strategy consisted in checking the HdrB synteny withHdrAC subunits. In addition, these sequences were aligned with knownHdrBs (e.g. from methanogens) and identified SdhE₈ subunits (e.g. from *Acidimans ambiguus*) and a global identity analysis with hierarchical clustering performed to aid in the distinction of SdhE₈ andHdrB. If still no classification as SdhE₈ was possible, the “HdrB” annotation was kept. The differentiation between SdhE₈ and SdhE₉ was performed based on the absence of the SdhE₈ subunit within the complex.

3. Results and Discussion

The combined analysis regarding the distribution and phylogenies of SDH/FRD and close homologues is described below.

3.1. Distribution of SDH/FRD in prokaryotes

3.1.1. MCL cluster composition

The rBBH search for homologues identified 201,016 unique sequences (270,215 in total), with 87,278 sequences annotated as SDH/FRD subunits. After filtering for 25 % identity, the global alignment of these sequences produced over 1.5 billion relationship pairs. MCL analysis provided 105 clusters (Table S4), 27 of them containing SDH/FRD subunits, with the remaining containing other complexes used as queries. SDH/FRD sequences were identified in 77 % (26894) of genomes in the dataset, being absent from 8122. Most of these genomes are metagenomic assemblies, with different levels of completeness, so it is not clear whether these organisms have no SDH/FRD or if the assemblies are simply missing those sequences. Among the complete genomes devoid of SDH/FRD are several taxa known for having extremely reduced genomes ([184–187]) such as Nanoarchaeota (8 out of 8), Nanohaloarchaeota (8/8), the DPANN group (132/134), and the bacterial Tenericutes group (248/248). In addition, SDH/FRD were also absent within novel candidate phyla and in some genomes from phyla in which full SDH/FRD complexes were identified.

After functional annotation with KEGG and PFAM and inspection of the number of histidines in the membrane anchor module, the composition of clusters became clearer (Table S4). While SdhA and SdhB sequences, irrespectively of the type, were present in three and four different clusters, respectively, the membrane anchor subunits were grouped into a higher number of clusters: 12 in the case of SdhC and 8 in the case of SdhD. Membrane anchor subunits of type C are found in four of these clusters, and type B found in five. Type A membrane subunits are spread among 8 clusters, while only two clusters have type D sequences, being type F and E sequences each in separate clusters. The multitude of clusters for membrane anchor sequences of types A, B and C hints at the potential existence of subtypes within these groups. Interestingly, one cluster contained fusion sequences of SdhC and SdhD subunits, distinct from the canonical SdhCₓ, and containing 7 predicted transmembrane helices. These sequences were affiliated with the Chloroflexi phylum indicating a lineage specific fusion event. The non-canonical membrane subunits of amphipathic nature present in SdhE₈ (e.g. characterised *Acidimans ambiguus* SdhE₈ [89] and *C. jejuni* MfrE [131]) are found in a cluster containing also HdrB proteins. Within the largest cluster, in addition to SdhA sequences, other TfrAs and other closely related sequences which could not have been differentiated from SdhA by functional annotations are present.

Besides homologues already present in the query set, additional homologues of SdhA were identified such as urocanate reductase (UrdA) [188], tricarboxylate dehydrogenase (ColbZ/TcuA) [189], alkaldihydroxyacetonephosphate synthase (Agps) [190], 3-oxo-Salphtet-sodium-4-dehydrogenase (Tesl) [191], D-lactate dehydrogenase (Dld) [192], and 3-oxoesterid 1-dehydrogenase (KsdD) [193]. Homologues of SdhB include several ferredoxin and HdrC2 proteins, both of which contain FeS clusters. In the case of SdhE₈, besides HdrB also Lde, a Cysteine-rich domain containing protein involved in lactate utilization, was identified.

3.1.2. Taxonomic distribution per SDH/FRD type

The classification of the sequences per type, shown in Fig. 2 and S1 allowed to analyse the overall distribution of complexes per type. In this dataset, 31,944 complete SDH/FRDs complexes and 2239 incomplete (lacking at least one subunit) were identified (Table S5). Inspection of the existence of pseudogenes within the genomic assemblies revealed that in 725 cases, the identification of incomplete complexes might be due to assembly artefacts. However, the remaining 1514 cases open the possibility of the existence of novel modular architectures within this family and pinpoints enzymes to be biochemically characterised. In the case of 2002 sequences, no type classification could be attributed (see “Materials and methods”). The taxonomic distribution of each type is described in detail below.

3.1.2.1. Type A. Type A enzymes are the taxonomically most diverse type and include the characterised complexes from *Myocobacterium smegmatis* [134], *Halobacterium salinarum* [44,71,194], *Natronomonas pharaonis* [71,195], *Rhodobacter sphaeroides* [196], *Mycobacterium tuberculosis* [46], *Thermus thermophilus* [86], and *Thermoplasma acidophilum* [71,197]. These complexes are widespread in both Archaea and Bacteria, being present in 8 archaeal and 40 bacterial phyla. In Archaea, type A SDH complexes were identified in the majority of Archaeoglobi (8/10 genomes), Halobacteria (over 90 % of 388 genomes present in the dataset), Korarchaeota (five out of five genomes), and Thaumarchaeota (83 % of 88 genomes). SDH complexes of this type are also detected in at least 40 % of metagenomic assemblies affiliated with *Candidatus Heimdallarchaeota*, *Candidatus Marsarchaeota*, and *Crenarchaeota*, specifically within Thermoprotei. Interestingly, in *Acidimans ambivalens* genomes, besides the canonical type E experimentally characterised by Lemos et al [89], an incomplete SdhBCDₐ complex was identified. Inspection of the surrounding genes did not allow the identification of other proteins that could replace the flavin subunit.

In Bacteria, this type is widespread in several phyla such as in Actinobacteria (66 % of 5489 genomes), Deinococcus-Thermus (over 90 % of 95 genomes), Deferribacteres (11/11 genomes), Rhodothermaeota (6/11 genomes) and *Candidatus phyla* (8/13 genomes), *C. Tectomicrobia* (6/6 genomes), *C. Division Zixibacteria* (9/18 genomes), *C. Kryptonida* (four out of four genomes), and *C. Marinimicrobia* (6/12 genomes). In addition, type A SDH complexes are scarcely present in 30 bacterial additional phyla. Of note, the short length of 32 amino acids of the proposed SdhF₈ subunit present in *M. smegmatis* prevents this peptide from getting accurate BLAST hits, therefore it was excluded from the analysis. Moreover, this protein is not in synteny with the complex but 4038 genes downstream in one direction and 2773 genes upstream in the other.

The existence of different SDH/FRD types within closely related strains was also observed. While in this analysis, in *Rhodothermus marinus DSM 4252* genome an SDH of type A was identified, the characterised enzyme from *Rhodothermus marinus PRQ32B* albine strain belongs to type B ([75] and Miguel Teixeira personal communication). It would be of interest to compare the position within the phylogenies of these two complexes, but the lack of genomic records for strain PRQ32B impairs this analysis.

3.1.2.2. Type B. Type B complexes include the characterised complexes...
from *Bacillus subtilis* [30,50,54,55–61,78], *Bacillus cereus* [90], *Bacteroides fragilis* [91], *Bacteroides thetaiotaomicron* [92], *Helicobacter pylori* and *Campylobacter jejuni* [93], *Desulfovibrio gigas* [94], *Geobacter sulfurreducens* [95], and *Wolinella succinogenes* [74,96,176]. This type is widespread among bacterial lineages being present in a total of 49 bacterial phyla. Type B SDH/FRD complexes were identified in the majority of Acidobacteria (70 % of 130 genomes), Bacteroidetes (96 % of 2769 genomes), Balneolaeota (17/17 genomes), Chlorobi (62 % of 34 genomes), *Candidatus* Tectomicrobia (6/6 genomes), Chlamydiae (72 % of 65 genomes), Planctomycetes (64 % of 245 genomes), Verrucomicrobia (88 % of 70 genomes), Fibrobacteres (93 % of 44 genomes), *Candidatus* Tectomicrobia (6/6 genomes), Chlamydiae (72 % of 65 genomes), Planctomycetes (64 % of 245 genomes), Verrucomicrobia (89 % of 217 genomes), Actinobacteria (~50 % of 5489 genomes), Armatimonadetes (55 % of 33 genomes), and Bacilli (66 % of 3143 genomes). Interestingly, in Proteobacteria, SDH/FRDs of type B are widely distributed in some proteobacterial classes (5/5 genomes of *Candidatus* Lambdaproteobacteria, 64 % of 672 genomes of Deltaproteobacteria, 84 % of 431 Epsilonproteobacteria genomes, and 88 % of 68 Oligoflexia genomes) while being very scarce or entirely absent in others. In Cyanobacteria, SDH/FRD complexes of this type are present in only 15 % of the 556 metagenomic assemblies. However, type B are almost entirely absent from Archaea, with the exception of a few metagenomes from *Candidatus* Thorarchaeota (one out of five, only SdhA and SdhC subunits identified), Methanobacteria (four out of 66 genomes), and unclassified Euryarchaeota (one of 73 genomes).

3.1.2.4. Type E.Canonical type E complexes (i.e. with two amphipathic membrane anchor subunits SdhE and SdhF that include characterised complexes from *A. ambivalens* [71,89], *S. acidocaldarius* [71,130] and *S. tokodaii* [71,199]) were identified only in Archaea, in the Thermoprotei class of Crenarchaeota (~32 % of 117 genomes), the...
Thermoplasmata class of Euryarchaeota (~14% of 58 genomes), and in one unclassified Euryarchaeota genome. Interestingly, a variation of type E SDH architecture, containing the SdhE subunit but lacking SdhF, (in here denoted as type E*, with characterised representatives from C. jejuni MFR [131,200] and Synechocystis sp. SDH [201]), were found in Bacteria. This type was found in 15 bacterial phyla, predominantly in Aquificae (50% of 40 genomes, excluding soluble NADH-dependent fumarate reductases [202]), Chlorobi (74% of 34 genomes), Nitrospinae (74% of 19 genomes), Cyanobacteria (62% of 556 genomes), and Negativicutes (45% of 189 genomes).

3.1.2.5. Type F. The newly discovered type F [88,133] was thought to be exclusively present in Actinobacteria. In our analysis, this complex was identified in four archaeal lineages and 28 bacterial lineages. In Archaea, it is mostly present in Candidatus Poseidonia (80% of 15 metagenomic assemblies), and Euryarchaeota (14% of 58 Thermoplasmata and 17% of 73 unclassified genomes). In Bacteria, this type was found in 15 bacterial phyla, predominantly in Aquificae (50% of 40 genomes, excluding soluble NADH-dependent fumarate reductases [202]), Chlorobi (74% of 34 genomes), Nitrospinae (74% of 19 genomes), Cyanobacteria (62% of 556 genomes), and Negativicutes (45% of 189 genomes).

3.1.2.6. Thiol:Fumarate reductases (TFR) and “N.A.” sequences. Thiol: fumarate reductases, first discovered in M. marburgensis str. Marburg [49,139], are enzymes that in methanogenic archaea perform the conversion of fumarate to succinate using Coenzyme M. They are an electron donor [139]. These soluble enzymes contain a flavin subunit (TfrA), and an iron-sulfur subunit with a CCG domain (TfrB). Due to the closer homologous relationship of the flavin, iron-sulfur and CCG subunits of SDH/FRD, thiol: fumarate reductases are also included in this analysis. Soluble thiol: fumarate reductases were identified predominantly in Archaea (four phyla) and scarcely identified within 9 bacterial phyla. In Archaea, TFR complexes mainly occur in Euryarchaeota (methanogenic lineages, but also one Archaeoglobi) and C. marburgensis, as well as in Candidatus Thrarchaeota and Candidatus Batharchaeota. In Bacteria, this complex is only found in 19 metagenomic assemblies, most of them belonging to Candidatus Roizmanbacteria and Deltaproteobacteria. Although Coenzyme M is present in some bacteria [203], to our knowledge, Coenzyme M was not detected in Archaeoglobi or any of the bacterial lineages, therefore the cases identified in these lineages either use a different interaction partner or are a result of contamination or misassembly artefacts. As for the sequences identified as TFR in Thorarchaeota and Batharchaeota, this computational analysis cannot tell whether or not they are true TFRs or close homologues, although in some of these lineages, at least the Wood-Ljungdahl pathway was identified [204–207].

In addition to complexes of types A to F and thiol:fumarate reductases, there are 1780 SdhA or SdhB sequences for which no type could be assigned. For such cases, no membrane subunits were identified. Many of such SdhAB complexes are close to type A, B or E/E* by both global identity and phylogenetic analysis (see below), and often were found in genomes that also contain a full SDH/FRD complex of an identified type. The cases without type classification were found in 8 archael and 30 bacterial phyla.

3.1.3. Phylogenetic analysis of SdhA

In the joined maximum likelihood phylogenetic reconstruction of SdhA (Figs. 3 and S2), well supported monophyletic clades for sequences of types C, D, canonical E, and F are observed, while sequences of types A and E* are intercalated. Such intercalated nature of type A clades, as well as differences in sequence composition of the enzymes within these clades (Data not shown), hint at the existence of possible subgroups within this type. In this reconstruction, at least four type A clades can be distinguished, one stemming out of the largest TFR clade (see below) and the remaining intercalated with E/E* enzymes. The archael SdhAA sequences are present in three of the four clades, one containing exclusively Halobacterial sequences, other retrieving the grouping observed in [3], and the Heimdallarchaeal sequences are grouped in a clade containing type A bacterial proteins. The remaining type A clade closer to the large TFR clade is composed mainly of proteobacterial sequences and has low bootstrap support (Fig. 3). Of note, this SdhAA clade has its respective iron-sulfur center subunit located within the large bacterial/Heimdallarchaeal SdhBx clade (Fig. 4). A closer inspection of the SdhAA sequences revealed that the ones closer to the TFR clade are missing conserved two residues in the vicinity of the FAD cofactor (H46 and T255 M. smegmatis numbering). Besides this, with exception of random insertions, no other major difference was identified between the sequences.

The existence of distinct clades of type E*, is in agreement with literature reports [131,200,201] that suggest that this type contains functionally diverse enzymes. In contrast, canonical type E sequences constitute a single well-supported clade containing only archael sequences.

In this phylogeny, TFR sequences are found in three non-monophyletic clades. The larger TFR clade, closer to the root, contains 97 proteins associated with various taxa. The remaining clades contain proteins from archael metagenomic assemblies, one composed of five proteins affiliated with unclassified Euryarchaeota (located between canonical type E and the archael type A clade), and the last, located next to the cyanobacterial type E* clade, contains four proteins belonging to Candidatus Batharchaeota, as well as truncated proteins belonging to Deltaproteobacteria (1) and Methanobacteria (1). These two last TFR clades might correspond to specific adaptations occurring within these phyla or be a result of assembly artefacts, since the fused subunit characteristic of TFR was also identified. Moreover, SdhF proteins are grouped in a highly supported monophyletic clade branching close to TfrA proteins, and might represent a recent adaptation, with a membrane anchor replacement occurring within bacterial organisms. Within type F clade, archael sequences are present in two subclades (one with C. Poseidonia, C. Heimdallarchaeota, and unclassified Euryarchaeota sequences, and the other with Thermoplasmata proteins). This is an indication of two potential interdomain lateral gene transfer events.

In the SdhA phylogeny, type B sequences are monophyletically organized, separated in at least three distinct groups and were initially separated by the root by MAD. A closer inspection of the MAD rooting results showed that this is an ambiguous root and that the support for the branch separating B type enzymes from the remaining is equally valid (Supplementary Figs. S2 and S3). This separation of type B into three groups is in agreement with what was previously observed [3,95,208], and may be the result of later enzymatic specializations. On one side of the root, two type B groups contain sequences from SDH or bifunctional complexes (Subgroup II and III in Table S6), while the third group, located on the other side of the root, contains FRD or bifunctional complexes (Subgroup I in Table S6). The distinction of the functional clades is supported by the CD SEARCH assignments which classifies subgroup I as fumarate reductase and subgroups II and III as succinate dehydrogenases/fumarate reductases. Although this separation is also
partially supported by the few available data on characterised type B enzymes and their *in vivo* wild-type predicted function (77,78,90–96,208–211), the lack of experimental characterization of other taxonomically diverse enzymes refrains us from putting forward any additional speculations regarding the enzymes’ *in vivo* function. Of note, type B enzymes are not taxonomically organized, with both prokaryotic domains represented in each of the three clades (Figs. 3 and 4, red and black branches). Additionally, one type B clade contains proteins from Clostridia and Methanobacteria that have a corresponding subunit B but are devoid of anchor subunits. Since some of these sequences are from complete genomes of cultivated organisms, this could be a case involving the recruitment of a common domain to perform a different function, or correspond to a soluble form of SDH/FRD enzyme, closely related to the membrane-anchored type B complex. Type B is not only placed as a basal type by MAD rooting in phylogeny of both SdhA and SdhB, with TFRs as basal in the sister clade of type B enzymes, but also it has the lowest mean/median global identity to other types (Table S7).

NadB sequences and/or SdhA protein homologues, for which no other SDH/FRD subunits were found (in grey in Fig. 3) form two large clades between type F and TFR clades. To our knowledge, none of these proteins, mainly present within Actinobacteria and Alphaproteobacteria, are so far characterised.

Although the phylogeny of catalytic subunits is discussed, it is observed that types with less than two hemes attached to the membrane subunits (C and D) stem out of distinct type A clades, suggesting that each of these types originated independently through the process of heme loss and that this signal is retained also at the level of the cytoplasmic subunits. This is also supported by the mean and median global identities between SdhA subunits of the different types (Table S7), where it can be seen that SdhAC is more similar to SdhAA than to SdhAD, further supporting the relatedness of the complexes belonging to type A and type C. Although type D is equally similar by global identity to types C and A, it forms highly supported clades within the SdhA and SdhB phylogenies. Interestingly, SdhAD and SdhAE are also similar (42.9 % mean global identity). In the case of type C, two prokaryotic subclades hint at the existence of two potential subgroups.

All of eukaryotic SdhAEUK sequences form a monophyletic well-supported clade within type C enzymes, having Alphaproteobacteria as basal. This is also supported by mean/median global identities (SdhAC

---

Fig. 3. – Maximum-likelihood reconstruction (LG amino acid substitution matrix four discrete gamma categories model; LG + G4) of SdhA proteins. Phylogenies were rooted using the minimal ancestor deviation method [182]. Black circles indicate significant ultrafast bootstrap (BT) support above 95. All other node support values are omitted for simplicity. The scale bar indicates the number of substitutions per site. Type A is coloured in orange, type B in green, type C in yellow, type D in light blue, type E/E* in indigo, type F in purple, TFR sequences are coloured in red while the clades of undetermined SDH/FRD type or homologous complexes are coloured in grey. Each clade containing characterised complexes is labelled with the corresponding organisms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Interestingly, with exception of *Drosophila melanogaster* sequences, all other isoforms are sisters of each other, some of them being truncated forms of the enzyme.

### 3.1.4. Phylogenetic analysis of SdhB

The overall topology of the joined maximum likelihood phylogenetic reconstruction of SdhB (Figs. 4 and S3) is similar to the one found in SdhA. In this phylogenetic reconstruction, type B subclades were not separated by the root by MAD (but still forming three subgroups). Further explanation of the MAD root separation is available in Supplementary information.

In comparison to the SdhA,E phylogeny, the SdhB,A,E sequences are organized in more distinct clades with a separation between E and E*. This is also observed in the intertype mean global identities between type SdhB and SdhB (intertype mean global identity of 26.6 %) and SdhA,A and SdhA,E (intertype mean global identity of 42.9 %). A possible explanation relies on the fact that the SdhB subunit is in contact with the membrane anchor module that differs significantly between these types (amphipathic nature of SdhE and the transmembrane nature of SdhCD).

The topology of SdhB clades differs slightly from what is observed in SdhA phylogeny. The low-supported clade, consisting mainly of proteobacterial sequences that in SdhA phylogeny was closer to TFRs, is found together with other bacterial and Heimdallarchaeal SdhB sequences. However, there is a small well-supported clade basal on one side of the root, consisting of some Deltaproteobacterial SdhB sequences, with the corresponding SdhA subunits found in a bacterial/Heimdallarchaeal clade. One explanation for different placement of the SdhB subunit in these cases could be the sequence differences necessary to accommodate a fused transmembrane anchor with 7 helices, which is found not as rBBH but as a syntenic neighbour in these genomes. A similar seven helix transmembrane subunit is also found in some of the Chloroflexi (as previously mentioned). These complexes are in here denoted as type A*.

In this phylogeny, the eukaryotic sequences are basal to a subclade of
type C including Alphaproteobacteria, opposed to what is observed in SdhA phylogeny. The mean global identity between SdhB type C including Alphaproteobacteria, opposed to what is observed in V. Karavaeva and F.L. Sousa. Based on the modular nature of the different types alone, several levels of annotations for each sequence (node), such as SDH/FRD number of membrane helices are indicated as described in Fig. 1. Complex X subunits of NADH-FRD are represented by three black ellipses. Cofactors and indicates lack of homology to canonical SdhCs. Additional non-homologous subunits of NADH-FRD are represented by three black ellipses. Cofactors and numbers of membrane helices are indicated as described in Fig. 1. Complex X indicates the soluble primordial module. Solid arrows indicate the direction of evolution of the types, while dotted arrows indicate an addition or a loss (based on the direction of the arrow) of a domain/cofactor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
evolution of this large family was elaborated (Fig. 5).

In this scenario, and considering the phylogenetic reconstructions topologies also observed in other reconstructions (clades of types A and B in [3], type B SDH and FRD clade separation in [208], three subclades in [95]), the first event would have been an ancient duplication of the FAD and iron-sulfur subunits, not yet associated with membrane anchors, that had undergone parallel evolution to give rise to type B on one hand, and TFR and the remaining SDH/FRD structural types on the other hand.

If any of the membrane counterparts would already exist, then they would have been lost and regained several times. Moreover, this modular block organization composed of a FAD and an iron-sulfur subunit is observed in many other homologous complexes not addressed in here, such as the sn-glycerol-3-phosphate dehydrogenase (two subunits homologous to SdhA and SdhB [85,138], Apr reductase (one subunit homologous to SdhA and a second subunit containing iron-sulfur centers [85]), further supporting the joined evolution of the SDH/FRD soluble modules with the later attachment to the membrane [3,95,208,216,217]. The most likely function of the primordial enzymes would have been the reduction of fumarate for biosynthesis purposes.

Early in the evolution of this complex, the TFR branch recruited a CCG domain, homologous to the one present in heterodisulfide reductase subunit B, that in TFR was later fused with the iron-sulfur subunit. The HdrB domain, is a modular part of the HdrABC complex that is containing one iron-sulfur sulfur subunit.

‑ sulfur center, would have recruited an additional modular block organization composed of a FAD and an iron-sulfur subunit (SdhF), giving rise to type E enzymes.

modular block organization composed of a FAD and an iron-sulfur subunit (SdhF), giving rise to type E enzymes.

The intercalated nature of type A, type E and type E* as well as the presence of a type A clade close to TFRs suggests several independent associations with membrane anchors, followed by functional specializations, as in the case of MFR and the soluble Aquificae NADH-dependent fumarate reductase [202].

In this scenario, and as supported by phylogenetic analysis, type F would have been the result of a newer association with its specific membrane anchor, probably adapted to Actinobacterial polyketide quinones [133]. Type C and type D would have independently evolved from type A enzymes by losing one or two hemes respectively. Eu-
karyotes would have acquired SDH from the mitochondrial ancestor that currently is thought to be of Alphaproteobacterial origin.

The other side of the root only contains type B enzymes that evolved from a duplication of the primordial module. In this case, the same membrane anchor recruited by type A would have been recruited also by type B enzymes within bacteria. Through time, several events of inter-domain gene transfers would have occurred, with the gain of type B enzymes within several archaeal species, belonging to two phyla (Euryarchaeota and Candidatus Thorarchaeota). Due to the small size and functional characteristic of the transmembrane anchor module, it is not clear at this point, if the primordial module, would consist of one or two subunits. Since fusions tend to be more frequent than fissions [220], and type B taxonomic distribution is narrowed than that of type A, Occam's razor favours the former over the later hypothesis.

An idea, that can perhaps be put forward here, is that modular blocks that were involved in their evolution through time and are parts of larger protein families might be older than the ones less frequently used in biological networks, or in particular, in energy conserving electron transport chains. Of course, there will be exceptions, as in the case of the well conserved ATP synthase [221], but this trend might be valid for other modules such as for instance, iron-sulfur centers [222], or some enzymes belonging to the large family of Moco-enzymes [215]. Regarding the full SDH/FRD complex, its place in the major events in prokaryotic evolution was, so far, not clear. In the last years, laboratory experiments have achieved the synthesis of several intermediates of the TCA cycle, from where fumarate and succinate were present [216], which would be consistent with the existence of this substrate early in evolution. In addition, although the SDH/FRD reaction is not part of the 402 reactions of the biosynthetic core that trace to the last universal common ancestor [217], both fumarate and succinate are part of the metabolic network, so is the reaction nowadays catalyzed by the soluble version of NADH-fumarate reductase. In fact, the calculated deltaG of the reaction under alkaline vent conditions (−65.4 kJ·mol⁻¹) [217], one of the possible scenarios for the habitat of our last common ancestor [223–225] favours the conversion of fumarate to succinate in the reductive direction. However, the lack of a clear archaeal and bacterial separation in any of the protein types, as seen in the SdhA and SdhB phylogenies, might be an indication of a later event of membrane attachment giving rise to the full complex after the diversification of both prokaryotic domains.

With the attachment to the membrane, which occurred independently and many times, the complex became part of the anaerobic electron transport chains, and microorganisms were able to optimize their ATP production. Since the oxidation of succinate to fumarate depends on high potential electron acceptors [226], specialized SDHs only evolved at a later time, with the increase in oxygenation levels of the atmosphere [227]. These findings show a disconnection between the existing structural classification and the evolution of the modular structure of the complexes, partially contradicting the view established by Hägerhäll and Hederstedt paper [136], which hypothesized that type D resulted through the heme loss from type C, and type B resulted from fusion of anchor subunits present in type A.

4. Conclusions

The conversion of succinate to fumarate (and vice versa) is highly conserved among the three domains of life: Archaea, Bacteria, and Eukarya [2,3,85]. These enzymes participate in both respiration and fermentation based on the organism and the environment it inhabits [2,4,5]. The comparative genomic analysis of this superfamily across 35,000 genomic assemblies expanded the current taxonomic distribution of several of the types complex in prokaryotes and allowed the potential identification of novel subtypes, to be further experimentally characterised.

The combined analysis of the phylogenetic reconstruction and similarity networks allowed the elaboration of a scenario in which a pr-
mordial soluble module, composed of the common cytoplasmic subunits, underwent several independent events of membrane attachments, replacements, fusions and environmental adaptations to give rise to the current taxonomic distribution.

Supplementary data to this article can be found online at https://doi.

CRediT authorship contribution statement

FLS designed the research, VK performed the analysis. Both authors analysed the data and wrote the paper. All authors have seen and approved the final version submitted.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.
References


V. Karavaeva and F.L. Sousa

BBA - Bioenergetics 1864 (2023) 148916

13


https://doi.org/10.1093/oxfordjournals.jbchem.a124497.

Parasitol. 66 (1994) 273


https://doi.org/10.1016/S0021-9258(18)51526-X.

https://doi.org/10.1016/0166-6851(94)90163-5.


https://doi.org/10.1016/j.bbobi.2013.01.005.


https://doi.org/10.1016/j.bbi.2003.03.015.

https://doi.org/10.1074/jbc.M112764200.

https://doi.org/10.1016/j.bbrc.2005.03.015.

https://doi.org/10.1007/s11103-009-9573-z.

https://doi.org/10.1104/pp.108.134684.

https://doi.org/10.1039/A302917F.

https://doi.org/10.1016/0378-1119(92)90260-V.

https://doi.org/10.1016/j.bjp.2004.03.001.

https://doi.org/10.1007/s11103-010-0654-0.

https://doi.org/10.1073/pnas.2022308118.

https://doi.org/10.1016/j.bbabio.2013.01.005.


https://doi.org/10.1039/A302917F.


https://doi.org/10.1104/pp.108.134684.

https://doi.org/10.1039/A302917F.


V. Karavaeva and F.L. Sousa


