universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Succinate:quinone oxidoreductases: evolution of individual subunits"

verfasst von / submitted by

Valeriya Karavaeva, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von / Supervisor:	Assoz. Prof. Maria Filipa Baltazar de Lima de Sousa, PhD

Abstract / Zusammenfassung

Succinate dehydrogenases (SDHs) and fumarate reductases (FRDs) catalyze the interconversion of succinate and fumarate. Complex II has been a focus of research for many years because this reaction is highly conserved in all domains of life, the cytoplasmic subunits of this complex are homologous to a variety of other proteins, and the enzymes are part of (an)aerobic electron transport chains and the TCA cycle. The current classification of SDH/FRDs is based on the structure of membrane anchor subunits and their cofactors. However, complex II has not been widely analyzed in the context of evolution or taxonomic distribution. In this work, a large-scale comparative genomics analysis of complex II addresses the questions of taxonomic distribution and phylogeny. Our findings report that for types C, D, and F, the 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, a new evolutionary scenario is proposed, in which a primordial soluble module, corresponding to the cytoplasmic subunits, would give rise to the current diversity via several independent membrane anchor attachment events. The results of this project also emphasize the necessity of further biochemical characterization of taxonomically diverse SDH/FRDs of different types, which is currently lacking.

Succinatdehydrogenasen (SDHs) und Fumaratreduktasen (FRDs) katalysieren die gegenseitige Umwandlung von Succinat und Fumarat. Komplex II steht seit vielen Jahren im Mittelpunkt der Forschung, da diese Reaktion in allen Domänen des Lebens hochgradig konserviert ist, die katalytischen Untereinheiten dieses Komplexes homolog zu einer Vielzahl anderer Proteine sind, und diese Enzyme ein Teil von den (an)aeroben Elektronentransportketten und dem Citratzyklus sind. Die aktuelle Klassifikation von SDH/FRDs basiert auf der Struktur von Membrananker-Untereinheiten und ihren Cofaktoren. Die Forschung von Komplex II wurde jedoch im Zusammenhang mit der Evolution oder der taxonomischen Verbreitung nicht umfassend analysiert. In dieser Arbeit befasst sich eine groß angelegte vergleichende Genomanalyse von Komplex II mit den Fragen der taxonomischen Verteilung und Phylogenie. Unsere Ergebnisse zeigen, dass für die Typen C, D und F strukturelle Klassifizierung und Phylogenie Hand in Hand gehen, während die Situation für die Typen A, B und E komplexer ist, was die Möglichkeit ihrer Klassifizierung in Untergruppen hervorhebt. Basierend auf diesen Erkenntnissen wird ein neues Evolutionsszenario für diese Enzyme vorgeschlagen, in dem ein primordiales lösliches Modul, das den zytoplasmatischen Untereinheiten entspricht, die aktuelle Diversität über mehrere unabhängige Membrananker-Anheftungsereignisse hervorrufen würde. Die Ergebnisse dieses Projekts betonen auch die Notwendigkeit einer weiteren biochemischen Charakterisierung von taxonomisch diversen SDH/FRDs verschiedener Typen, die derzeit fehlt.

Acknowledgments

I would like to express my gratitude to the people who helped me undertake this research.

First and foremost, I would like to thank my supervisor, Dr. Filipa Sousa, for all your dedicated and enthusiastic guidance. Your continuous support kept me inspired and fascinated. Thank you for making me fall in love with the origin of life research. I am extremely grateful you took me on as a student, and helped me become a better scientist.

Second, I would like to thank my teammates and mentors, Sinje Neukirchen and Anastasiia Padalko, for answering my countless questions and providing invaluable insights into the methods used in the project. I also would like to thank them and the other members of the team, Ulf Naumann, Edith Pichlhöfer, and our “honorary Geeg” Rafael Ponce for making me feel like I truly belong in this team. Thank you for making me feel like a real “Geeg”!

Finally, I would like to thank my mom Natalya and my partner Brendan for their unwavering faith in me. I simply could not have done this without you.

Abbreviations

SDH	Succinate dehydrogenase
FRD	Fumarate reductase
ETC	Electron transport chain
TCA cycle	Tricarboxylic acid cycle
TFR	Thiol:fumarate reductase
MFR	Methylmenaquinol:fumarate reductase
FAD	Flavin adenine dinucleotide
NadB	L-aspartate oxidase
GLP	*Sn*-glycerol-3-phosphate dehydrogenase
APR	Adenosine-5'-phosphosulfate reductase
GLC	Glycolate oxidase
HDR	Heterodisulfade reductase
LLD	L-lactate dehydrogenase
kstD	3-oxosteroid 1-dehydrogenase
SOX	Sarcosine oxidase
ThiO	Glycine oxidase
QMO	Quinone-modifying oxidoreductase
GcvT	Glycine cleavage system T protein
DadA	Glycine/D-amino acid oxidase (deaminating)
QCR	Complex III (Cytochrome bc_1 complex)
ThiG	Thiazole synthase
CobZ/TcuA	Tricarballylate dehydrogenase
FccA/FrdA	Soluble fumarate reductase (*Shewanella*)

Table of Contents

Table of Contents

1. Introduction

Succinate:quinone oxidoreductases (SDHs, EC 1.3.5.1) are membrane enzyme complexes that catalyze the succinate to fumarate conversion and link chemiosmotic coupling to carbon metabolism via the TCA cycle (Hederstedt 2003). These complexes are closely related and homologous to quinol:fumarate reductases (FRDs, EC 1.3.5.4), which catalyze the conversion of fumarate to succinate. 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 (Lemos *et al.* 2002; Lancaster 2002a; Lancaster 2011b).

1.1. What is the function of succinate: quinone oxidoreductases?

These enzymes are a part of aerobic mitochondrial-like (Fig. 1) and anaerobic fumarate-utilizing (Fig. 2) respiratory electron transport chains (ETCs) as well as the only membrane component of the TCA cycle. They are anchored either in the cytoplasmic membrane of prokaryotes or in the inner mitochondrial membrane of eukaryotes, with the catalytic domain in the cytoplasm or mitochondrial matrix, respectively (Lancaster 2002a).

Functionally, SDH/FRDs form 3 classes based on the reaction they perform *in vivo* and the type of the quinone they use: class 1 is SDHs that oxidize succinate and reduce a high-potential quinone (*e.g.* ubiquinone); class 2 FRDs reduce fumarate using a low-potential quinol (*e.g.* menaquinol); class 3 enzymes oxidize succinate with the help of a low-potential quinone. So far, it is impossible to tell which reaction a certain SDH/FRD enzyme would catalyze without *in vivo* tests (Hägerhäll 1997; Lemos *et al.* 2002; Lancaster 2002a; Lancaster 2011b).

1.1.1. Aerobic mitochondrial-like electron transport chain

Aerobic mitochondrial-like electron transport chains use oxygen as the final electron acceptor. Such electron transport chains have five complexes involved in electron transfer: NADH dehydrogenase (Complex I), succinate dehydrogenase (Complex II), cytochrome *bc1* (Complex III), cytochrome *c* oxidase (Complex IV), and an ATP synthase (Complex V). The electron donors are NADH for complex I and succinate for complex II. Electrons flow across three complexes until they reach complex IV, where they are passed onto the O_2 molecule to yield water. In aerobic mitochondrial-like ETC, SDH is the only complex that does not directly pump protons but some types of this enzyme might participate in enhancing the proton gradient by supplying reducing equivalents from succinate metabolism (charge separation; Lancaster 2002b). These reducing equivalents are transported through the ubiquinone pool to complexes III and IV, and these complexes, in turn, extrude protons (Moosavi *et al*. 2019). The generated proton gradient is used by the ATP synthase to produce ATP (Lancaster 2002a).

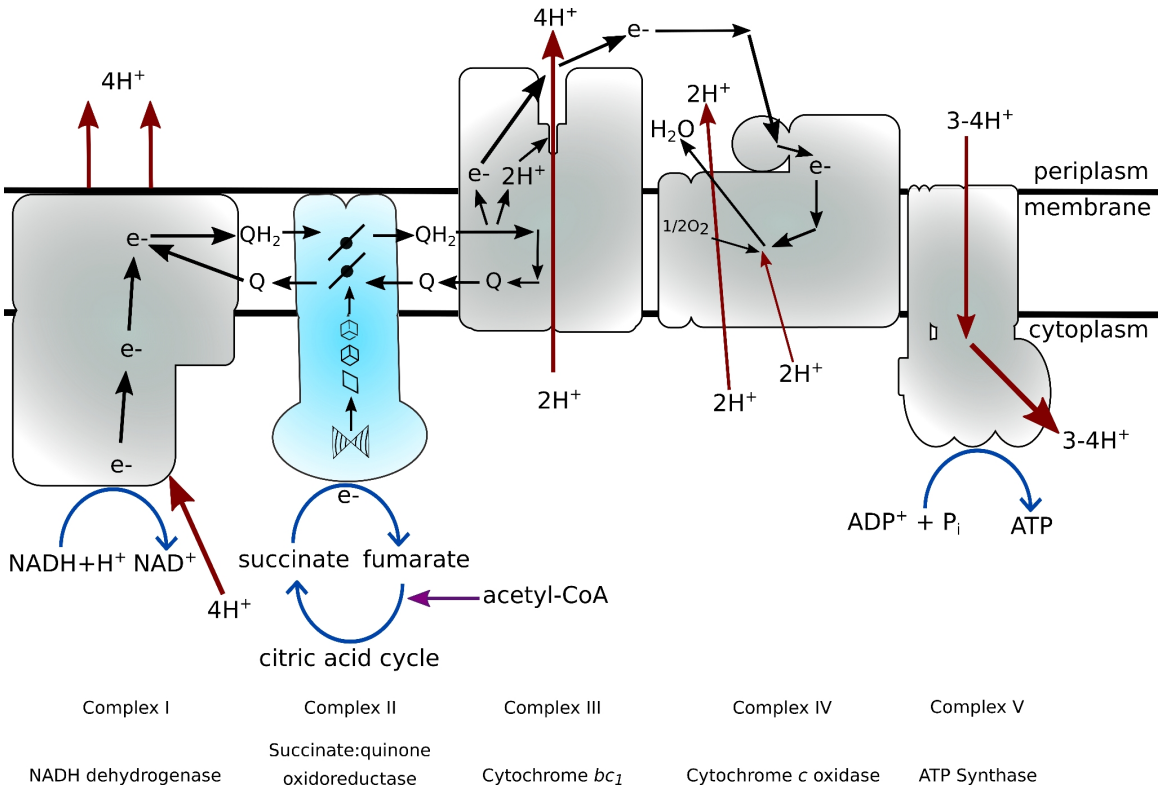


Fig. 1. Mitochondrial-like aerobic electron transport chain. *Succinate dehydrogenase (complex II) is shown in blue. This scheme is based on Fig. 1 from Lancaster 2002a.*

1.1.2. Anaerobic electron transport chain

Anaerobic electron transport chains using fumarate as terminal electron acceptor are found in diverse prokaryotes (Fig. 2). Using as example the one present in *Wolinella (W.) succinogenes*, it consists of three enzymatic complexes (as well as an ATP synthase), all containing two cytoplasmic subunits and a membrane anchor: formate dehydrogenase, fumarate reductase, and hydrogenase. Electrons flow from both hydrogenase and formate dehydrogenase to fumarate reductase via soluble quinones (usually menaquinone), where fumarate is reduced to succinate. Proton gradient generated by these complexes is used by the ATP synthase to generate ATP. For a detailed review see Kröger *et al*. 2002.

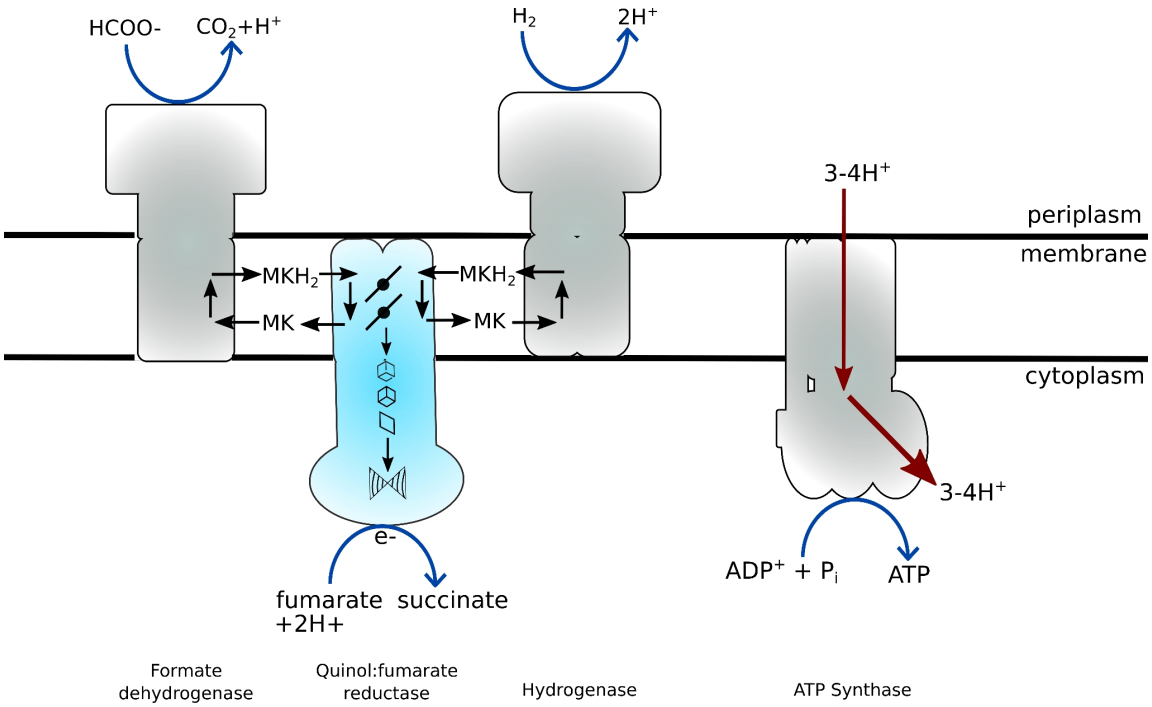


***Fig. 2. Anaerobic electron transport chain (fumarate respiration)**. Fumarate reductase is colored in blue. This scheme is based on Fig. 1 from Lancaster 2002a.*

1.2. Structure of SDH/FRD complex

Structurally, SDH/FRDs are composed of either three or four subunits that can be divided into the cytoplasmic part (catalytic SdhA and electron-transporting SdhB subunits) and the membrane anchor part composed of one or two subunits. Subunit A (Fig. 3) is a flavoprotein subunit that contains a dicarboxylate binding site where the succinate and fumarate interconversion takes place. This subunit is soluble and exposed to the cytoplasm, and contains one FAD cofactor (covalently bound in most organisms (Lancaster 2002a)). The FAD group serves as a first electron acceptor and passes electrons onto the other subunits (Hägerhäll 1997; Lancaster 2011b).

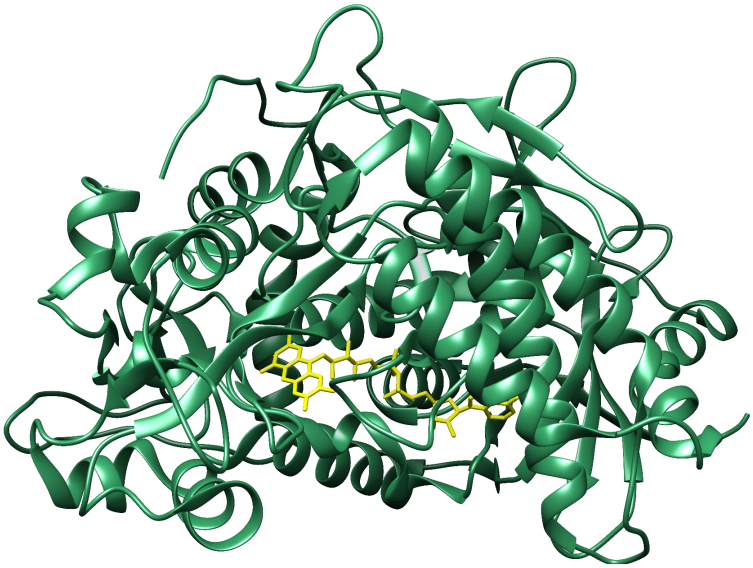


***Fig. 3. Structure of E. coli SDH subunit A (green) with bound FAD (yellow)**. RCSB PDB code: 1NEK. Depicted using UCSF Chimera software.*

Electrons flow from FAD into the next electron-accepting soluble subunit (SdhB; Fig. 4), which contains three iron-sulfur (FeS) centers with different compositions: S1 center ($[2Fe\text{-}2S]^{2+,1+}$), S2 center ($[4Fe\text{-}4S]^{2+,1+}$), and S3 center ($[3Fe\text{-}4S]^{1+,0}$) (Lancaster 2002a). In the succinate oxidation reaction, the S1 center is first to accept electrons from FAD and pass them onto the next centers (Hägerhäll 1997; Lancaster 2011b). The geometry of each FeS center is shown in Fig. 5, [2Fe-2S] is rhombus-shaped (top), [4Fe-4S] is cube-shaped (center), and [3Fe-4S] resembles an irregularly shaped cube due to the loss of one Fe atom compared to [4Fe-4S] (bottom right).

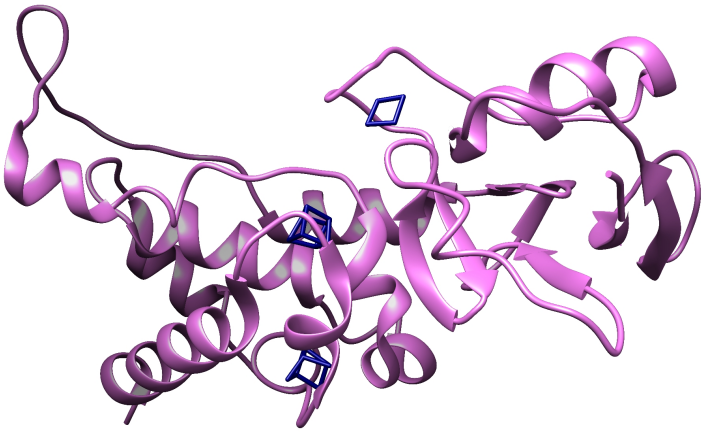


Fig. 4. Structure of E. coli SDH subunit B (purple) with three FeS clusters shown in dark blue. *RCSB PDB code: 1NEK. Depicted using UCSF Chimera software.*

The anchor part is composed of membrane subunits (C and D, or E and F, depending on a structural type) with various cofactor content (Fig. 5).

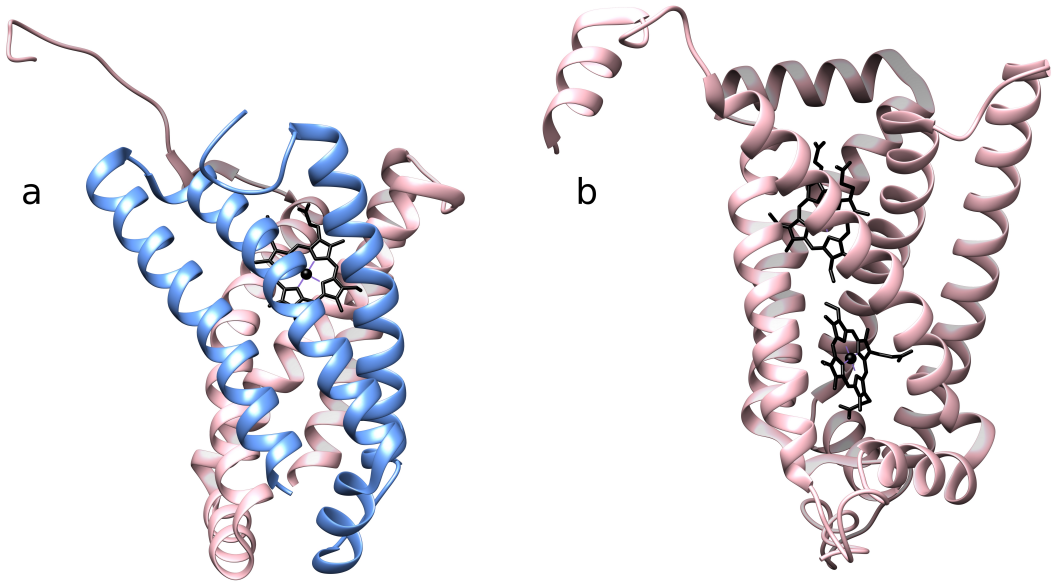


Fig. 5. Membrane components of SDH/FRD complexes. *a) E. coli SDH subunit C (pastel pink, left) and subunit D (pastel blue, left), b) W. succinogenes SdhC of type B (pastel pink, right). Heme groups (black) are located between the two subunits or, in the case of a single subunit, in the space between the inner helices. In W. succinogenes, SdhC binds two hemes, and in E. coli, subunit C and subunit D harbor one heme between them. RCSB PDB codes: 1NEK; 2BS2. Depicted using UCSF Chimera software.*

1.3. Current structural classification of SDH/FRDs

The structural classification of SDHs is based predominantly on the structure of the membrane anchor subunits and their cofactor content (Hägerhäll 1997; Lemos *et al.* 2002; Jardim-Messeder *et al.* 2017; Lancaster 2011b). Fig. 6 provides a summary of these structural types. SDH/FRDs belonging to the structural type A contain two separate membrane subunits (C and D), both with three transmembrane helices (similarly to the structures shown in Fig. 5 and Fig. 6), where two heme groups - a high redox midpoint potential heme group (b_H) and a low redox midpoint potential heme group (b_L) are bound (Hägerhäll 1997; Lancaster 2013). So far, enzymes of this type have been known to only have succinate dehydrogenase activity (Hägerhäll 1997; Lancaster 2002a).

In type B enzymes, on the other hand, only one large membrane subunit (SdhC), with five transmembrane helices is found. Similarly to type A, type B also binds two hemes (b_H and b_L). Type B enzymes were shown to be able to catalyze the reaction in both directions, *i.e.* being either SDHs or FRDs depending on the *in vivo* function (Hägerhäll 1997; Lancaster 2002a; Lancaster 2003).

Type C and D enzymes are very similar to type A, with the differences lying in the number of heme groups: type C has only one heme group (b_H), and in type D, no heme groups are present. The well-studied *E. coli* SDH belongs to the structural type C (Fig. 4-6). Interestingly, while a cysteine acts as a third ligand of the [2Fe-2S] center present in the subunit B of types A, B, and D, within type C enzymes it can be replaced with

aspartate or serine residue (Lemos *et al*. 2002). Also, and while type C complexes are known to be able to perform the succinate and fumarate conversion in both directions, known enzymes of type D were found to act only as fumarate reductases (Hägerhäll 1997; Lancaster 2002a).

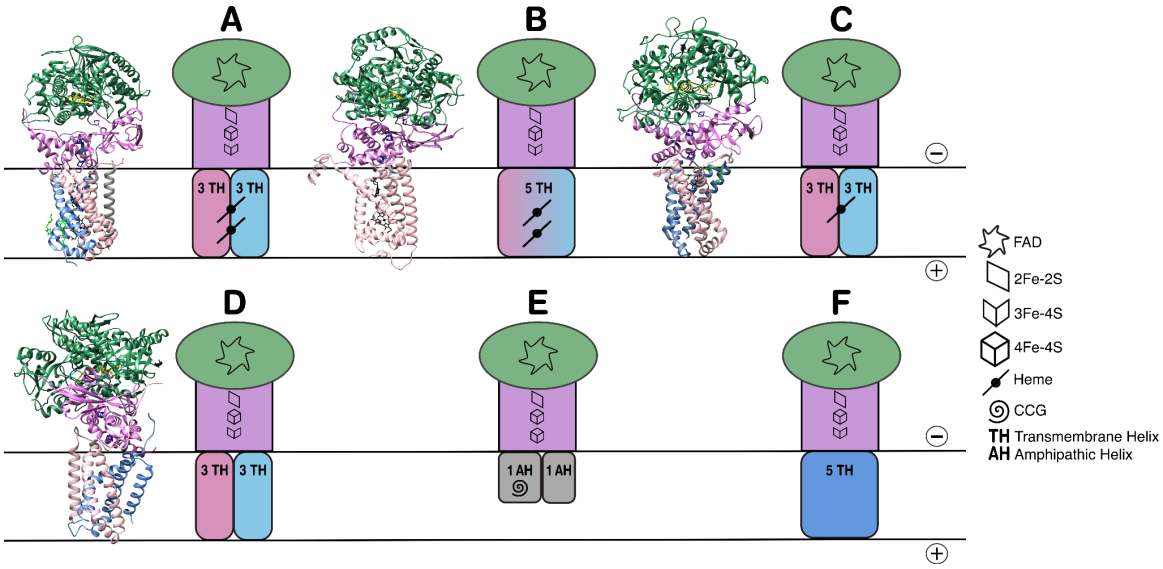


Fig. 6. Schematic representation of structural SDH types with their respective structure *(no structure for types E and F is available). From top left to bottom right: Type A, B, C, D, E, and F. Subunit A is colored in green, subunit B in purple, subunit C in pastel pink, and subunit D in light blue. Subunits E and F are colored in gray. The subunit C of type F 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. "-" and "+" indicate the cytoplasmic and periplasmic sides of the membrane. X-Ray crystallography structures of complex II (from type A to type D, respectively): Structure of M. smegmatis succinate dehydrogenase 2 (6LUM, Gong et al. 2020); structure of W. succinogenes fumarate reductase (2BS2, Madej et al. 2006); structure of E. coli succinate dehydrogenase (1NEK, Yankovskaya et al. 2003); structure of E. coli fumarate reductase (3P4P, Tomasiak et al. 2011).*

In 2001, a new type of SDH complexes (type E) was functionally characterized from the membranes of *Acidianus ambivalens* (Lemos *et al.* 2001; Lemos, *et al.* 2002). This type, known to be present in some Sulfolobales, is more dissimilar to the types described so far, having two amphipathic membrane subunits, E and F, which do not bind heme groups, have no predicted transmembrane helices, and bind specific sulfur-containing caldariella quinone. The SdhE subunit of this type is also known to bind a [4Fe-4S] center, and no cofactor was identified in SdhF (Hamann *et al*. 2009). In addition, its subunit B contains a second [4Fe-4S] center instead of a [3Fe-4S] center. The third ligand of the [2Fe-2S] center is a cysteine, similarly to types A, B, and D. This type catalyzes the oxidation of succinate to fumarate, with no known enzyme so far able to perform fumarate reduction (Lemos *et al.* 2001; Lemos *et al.* 2002; Lancaster 2002a). Aside from Sulfolobales, other organisms, such as Cyanobacteria and Aquificae, reportedly have SDH complexes with a membrane anchor subunit E but lacking subunit SdhF (Lancaster 2011a; Lemos *et al*. 2002). In this thesis, such complexes are marked as type E*. This type also includes the homologous methylmenaquinol:fumarate reductase complex (MFR; Juhnke *et al.* 2009; Guccione *et al*. 2010), which is located in periplasm and upregulated under high oxidative conditions. MFR does not perform

succinate to fumarate conversion, and although being able to catalyze fumarate to succinate reduction, it is likely to have some other function in the organism (Weingarten *et al*. 2009).

Recently, a new structural type (type F) has been proposed (Hards *et al.* 2019). So far, only one representative of this type has been characterized, the SDH2 from *Mycolicibacterium smegmatis* (Pecsi *et al.* 2014; Hards *et al*. 2019). This type is characterized by having only one membrane subunit with no bound hemes and no detected similarity to the membrane subunits present in types A to D. Similarly to the membrane subunit of type B, which is assumed to be the result of a fusion event between the subunits C and D of type A (Hägerhäll, Hederstedt 1996), Hards *et al* proposed that type F membrane anchor subunit rearrangement is the results of a potential fusion, this module being most similar to the type D anchor subunits (Hards *et al*. 2019). For convenience, structural type information is given as subscript in the subunit abbreviation (*e.g.* SdhA of type C is indicated as $SdhA_C$).

1.4. SDH/FRD homologous proteins and evolution

In most prokaryotes, SDH subunit genes are found in the vicinity of each other (Hägerhäll, 1997). Subunit A is usually found upstream of subunit B, and the membrane subunits tend to be usually clustered together upstream or downstream of the clustered A and B subunits (Lemos *et al.* 2002; Jardim-Messeder *et al.* 2017). The soluble subunits are homologous to the other enzymes that share the same domains, such as L-aspartate oxidase (NadB), which catalyzes the oxidative deamination of L-amino acids to alpha-keto acids and contains a flavoprotein subunit homologous to subunit A of SDH/FRDs (Mattevi *et al.* 1999), or the anaerobic *sn*-glycerol-3-phosphate dehydrogenase, in which GlpB is homologous to SdhA, and GlpC is homologous to SdhB (Cole *et al.* 1988). The anaerobic *sn*-glycerol-3-phosphate dehydrogenase uses *sn*-glycerol-3-phosphate as an electron donor and fumarate as an electron acceptor, and together with a fumarate reductase forms a short anaerobic electron transport chain (Cole *et al.* 1988). Also the subunit A (AprA) of the adenosine-5'-phosphosulfate reductase, an enzyme involved in dissimilatory sulfate reduction, shares homology with SdhA (Jardim-Messeder *et al*. 2017). In thiol:fumarate reductase, a soluble fumarate reductase that uses coenzyme M and coenzyme B as electron donors, subunit A (TfrA) is homologous to SdhA, while subunit B (TfrB) is homologous to both SdhB and SdhE (Heim *et al*. 1998). The glycolate oxidase, which catalyzes the oxidation of glycolate to glyoxylate, contains two subunits (GlcD and GlcE) which are homologous to SdhA, and the subunit F (GlcF) that shares homology to SdhB and SdhE (Pellicer *et al*. 1996). Additionally, heterodisulfide reductase subunits B and D (HdrB, HdrD) are homologous to SdhB and SdhE (Lemos *et al*. 2002). There are no homologous proteins of membrane subunits SdhC and SdhD belonging to canonical types of SDHs as well as there are no homologous proteins described so far for SdhF. The homologous relationships of the flavoproteins led Jardim-Messeder *et al* to propose the classification of the "fumarate reductase superfamily" (Jardim-Messeder *et al.* 2017).

The conversion of succinate to fumarate (and vice versa) is highly conserved among the three domains of life (Hederstedt, Rutberg 1981; Lemos *et al.* 2002; Jardim-Messeder *et al.* 2017) and, based on the organism and the environment it inhabits, these enzymes

participate in both respiration and fermentation (Hederstedt, Rutberg 1981; Hägerhäll 1997; Lancaster 2002a; Lancaster 2011b). Overall, understanding the evolution of succinate:quinone oxidoreductases will provide insights into the evolution of energy metabolism in general and of this family of enzymes in particular.

1.5. Goals of the thesis

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. The goals of this thesis included the reanalysis of the current structural classification in the framework of evolution, their current distribution across prokaryotic domains and the natural diversity of the complexes.

2. Methods

The research methodology used in this thesis is described below. Methodological scheme of methods pipeline is depicted in Supplementary Fig. 7.1.

2.1. Literature review

An extensive literature search was performed, where over 150 scientific articles were analyzed (including articles related to homologous proteins, methods articles, as well as any other potentially relevant papers). This analysis allowed to collect information regarding microorganisms containing characterized or reported succinate dehydrogenases/fumarate reductases.

2.2. Query dataset

Query sequences of SDH/FRDs and homologous enzymes were retrieved from BRENDA database (release 2020.2, Chang *et al.* 2021), KEGG (release 95.0, Kanehisa, Goto 2000), and UniProt (The UniProt Consortium 2021) databases or from internal databases, as in the case of heterodisulfide reductases and adenosine-5'-phosphosulfate reductases (Neukirchen, Sousa, 2021). For anaerobic *sn*-glycerol-3-phosphate dehydrogenase, heterodisulfide reductase, and adenosine-5'-phosphosulfate reductase, non-homologous subunits GlpA, GlpD, HdrA, HdrC, HdrE, and AprB were gathered to have additional information for synteny analysis. In addition, all of the at the time available SDH structures were retrieved from the RCSB PDB (Berman *et al.* 2000) database, with the exclusion of *E. coli* mutants. Query sequences were mapped to complete genomes from the dataset using NCBI BLASTP+ (Camacho *et al.* 2009). Additionally, copies were retrieved by blasting the acquired query sequences back to the genomes they came from using DIAMOND Blastp (Buchfink *et al.* 2021). Gene copies were defined using the cutoff of >=70% local identity, >=50% query coverage, and an E-value lower than 10^{-10}. Retrieved sequences, including homologous enzymes, were functionally annotated using Pfam (Mistry *et al.* 2007), and the transmembrane helices predicted using TMHMM (version 2.0, Krogh *et al.* 2001).

X-ray crystallographic structures of SDH complexes and respective cofactors were visualized using UCSF Chimera (version 1.14, Pettersen *et al.* 2004). The *E. coli* SDH structure (PDB code: 1NEK) was chosen as an “average” representative of all structural types. *W. succinogenes* $SdhC_B$ (PDB code: 2BS2, chain C) was kept as representative for fused membrane anchor subunits of enzymes belonging to type B.

Protein complexes were checked for their completeness. In some cases, information related to all of the subunits was not available at BRENDA (release 2020.2, Chang *et al.* 2021) or KEGG (release 95.0, Kanehisa, Goto 2000), so a search for the potentially missing subunits was performed. First, under the assumption that SDH/FRD subunits tend to form syntenic blocks, feature tables from respective genomes were checked to determine if the missing subunits were marked as “pseudogenes”, and thus were absent from the proteomic assembly. This affected the case of *Sulfolobus acidocaldarius*

assembly, where SdhB and SdhE were marked as “pseudogenes”. In addition, a DIAMOND Blastp (Buchfink *et al.* 2021) search of all queries to the genomes of the complex with missing subunit(s) was performed, and hits with 70% identity, 50% query coverage, and an E-value lower than 10^{-10} were marked as the potentially “missing” subunits. They were further analyzed using Pfam (PfamScan.pl version 1.5, Mistry *et al.* 2007) and TMHMM (version 2.0, Krogh *et al.* 2001). The results were compared to the same SDH/FRD subunits from other query organisms, and hit sequences were added to the query list if Pfam domains and the number of predicted transmembrane helices matched the expected. In total, 69 SDH and FRD complexes belonging to 59 organisms were gathered, spanning 10 bacterial phyla and three archaeal phyla. Queries included both both succinate dehydrogenase and fumarate reductase complexes as well as epsilonproteobacterial methylmenaquinol:fumarate reductases from *Campilobacter (C.) jejuni* and *W. succinogenes*, and sequences from homologous enzymes. In total, 293 query sequences were used for further analysis (Supplementary Table 6.1).

To assess the homology between query sequences at the subunit level, an all versus all pairwise global alignment per subunit was performed using Needleall, which is an EMBOSS (version 6.6.0.0) implementation of the Needleman-Wunsch algorithm (Needleman, Wunsch 1970). The pairwise global identities for each subunit were plotted as a heatmap using R package pheatmap (version 1.0.12, Kolde 2019; R Core Team 2020).

2.3. (Meta)genomic dataset

A subset of our in-house dataset of over 190000 metagenomic assemblies (downloaded from NCBI on November 2019 with two *Acidianus ambivalens* assemblies added at a later date (Neukirchen, Sousa 2021)) was created by filtering genomic records based on previously mapped NCBI taxonomic information and genomic quality in terms of contamination and completeness calculated by the Rinke method (Rinke *et al*. 2013). The analysis in this thesis was conducted using a dataset that contained all genomes marked as “reference” or “representative”. In addition, to ensure the existence of at least one representative per species, additional genomic records (one per species) were added, given preference for complete genomes followed by higher quality assemblies. If a species had several “complete” genomic records available then the best quality complete genome, based on redundancy and completeness, was kept. If no records of complete genomes were available for a given species, the best quality metagenomic assembly, based on redundancy and completeness, was kept. If more than one genome had exactly the same values for redundancy and completeness, the first one in the list was kept. The finalized dataset contained 35017 metagenomic assemblies with 33683 belonging to 179 bacterial phyla and 1334 to 22 archaeal phyla. The full dataset with the description of metagenomic assemblies is given in Supplementary Table 6.2.

2.4. Similarity analysis

Similarity searches were performed using the reciprocal best blast hit approach (rBBH) (Wolf, Koonin 2012). The search was conducted using DIAMOND Blastp (Buchfink *et al*. 2021) in “ultra-sensitive” mode with the “-k 0” parameter to report all targets for which alignments were found. The ultra-sensitive mode was used to improve the sensitivity of

the search in the <40% identity range. The version of DIAMOND used in this project was v2.0.4.142. The tabular output format 6 was used with the custom column order. 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 cutoffs 25% identity and an E-value lower than 10^{-10}. Copies of hits were retrieved by aligning each genome against itself, and filtering the resulting hits for 70% identity, E-value lower than 10^{-10}, and at least 70% query coverage excluding self-hits. For the second direction, retrieved hits (including copies) were blasted against a DIAMOND database of query genomes using as cutoffs of 25% identity and an E-value lower than 10^{-8}. The E-value cutoff for the second direction was increased to account for the larger size of the database since large DIAMOND databases could lead to increased E-values. Best hits for each accession in both directions are retrieved, and those hits that are found in both directions were deemed reciprocal and kept for further analysis. A hit is reciprocal if it is the best hit for the accession in both directions, or if the copy of this hit is the best hit in one of the directions. Any copy sequences of kept reciprocal hits were added to the final list of hits. A total of 201016 unique reciprocal hits and their copies were retrieved from the genomes, and an all-vs-all BLAST was performed using NCBI BLASTP+ (Camacho *et al*. 2009), and filtered for 25% identity and an E-value lower than 10^{-10}. Unique pairs fulfilling the threshold were globally aligned using Needleall (Needleman, Wunsch 1970). The global alignment was filtered for 25% identity and parsed as “accession1 accession2 identity” format for further use in MCL clustering.

2.5. Functional annotation and homology distinction

The unique rBBH hit sequences were functionally annotated using the NCBI Conserved Domain Batch SEARCH (abbreviated as CD SEARCH; CDD database, in automatic search mode with an E-value lower than 0.01 as threshold, composition corrected scoring ON, the maximum number of hits = 500, including retired sequences, Standard Results mode; Lu 2020). The sequences were labeled as fusions if two or more non-overlapping CD SEARCH domains characteristic of different subunits or proteins were found. Prediction of transmembrane helices was performed with TMHMM (version 2.0, Krogh *et al*. 2001) and TMPred (Hofmann, Stoffel 1993). For a clearer differentiation between AprA sequences and SdhA sequences, DiSCo was used (Neukirchen, Sousa 2021). In addition, KOfam (kofam_scan with KEGG mapper format, version 1.3.0, Kanehisa, Goto 2000; using HMMER version 3.2.1, hmmer.org) and Pfam (PfamScan.pl version 1.5, Mistry *et al*. 2007) 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. Based on the information from the genomic dataset table, the corresponding taxonomic affiliation and the metagenomic assembly level (“Complete”, “Scaffold”, “Contig”, or “Chromosome”) were mapped to the retrieved hits.

2.6. Synteny analysis

Analysis of the syntenic arrangement of the retrieved sequences was performed using the feature table information of their respective genomes (only complete genomes were used for this analysis). The neighborhood of a hit was defined by two sequences

upstream and two sequences downstream of the hit sequence, as long as all were present on the same chromosome or contig. The start and end positions, as well as the direction on the chromosome were obtained from feature tables. Neighbor sequences not previously identified as rBBH hits were functionally annotated as described above. The syntenic patterns of SDH subunits were analyzed and representative organizations were plotted using R package genoPlotR (Guy *et al*. 2010; R Core Team 2020). The resulting plots were used to produce a summary figure for syntenic organizations per SDH/FRD type. Gaps between subunit genes indicate that subunits are not in synteny. Strand direction was not taken into account.

2.7. Clustering of rBBH hits via Markov Chain Clustering (MCL)

The rBBH Needleall relationships were clustered using Markov Chain Clustering (MCL, version 14-137; van Dongen 2000; Enright *et al.* 2002) with an inflation parameter of 1.2. This value was chosen instead of the default 2.0 to account for possible overclustering artifacts, which were encountered during test runs (data not shown). The initial MCL inflation parameter setting of -I 2.0 resulted in several intercluster relationships above the 25% threshold, some of which had global identities above 90%. The inflation parameter affects cluster granularity: the higher the inflation, the more clusters will be obtained (and *vice versa*). The optimal value for this parameter varies depending on the data characteristics (van Dongen 2000; Enright *et al.* 2002). Therefore, to minimize underclustering, the clustering was performed with -I 1.2. The cause of these artifacts was not established with certainty. However, it is possible that the MCL algorithm is not currently adapted to work with large metagenomic data (over 270000 hits), since this data often contains partial or misassembled sequences that can introduce errors in the clustering procedure.

2.8. Cluster annotation and redundancy filtering

Resulting MCL clusters were annotated by determining the number of sequences per cluster, percentages of KOs, Pfam domains, CD SEARCH domains, number of predicted transmembrane helices, DiSCo hits, number of query sequences per cluster and their subunit and type, the estimated SDH type per sequence as a percentage per type in the cluster, and percentage of different genomic assembly levels per cluster, as well as a percentage of taxa (phylum and class) per genome and per sequence (Supplementary Table 6.3). In addition, the percentage of unique syntenic arrangements of SDH subunits, was calculated per sequence and per genome.

To assess the quality of the clustering procedure, intercluster mean, median and maximum identities were calculated and analyzed via hierarchical clustering, and the heatmaps plotted using R (R pheatmap package, version 1.0.12, Kolde 2019, R Core Team 2020). Taxonomic distribution of cluster hits was calculated as a percentage of genomes from given taxa (class or order for Thaumarchaeota and Cyanobacteria) that have sequences in a given cluster. The resulting matrix was sorted by protein type for cluster columns (SDH/FRD subunit or other homologous protein according to the main KOfam and CD search annotations per cluster) and NCBI taxonomy for rows and used to

create a heatmap (R pheatmap package, version 1.0.12; Kolde 2019; R Core Team 2020) for further analysis.

To reduce redundancy, clusters with over 1000 SDH/FRD sequences were reclustered using MCL (version 14-137; van Dongen 2000; Enright *et al.* 2002). For this purpose, sequences from clusters with intracluster relationships >=90% global identity (including self-hits) were extracted and reclustered in MCL. A representative per subcluster was kept for further analysis.

2.9. Multiple sequence alignment, phylogenetic and network analyses

Multiple sequence alignments of clusters containing SDH/FRD subunits were performed using ClustalOmega (version 1.2.4, Sievers *et al*. 2011) with following parameters: "--max-guidetree-iterations=100 --max-hmm-iterations=100 --output-order=tree-order". Alignments were trimmed using TrimAl (version 1.2, Capella-Gutiérrez *et al*. 2009) with a -gapthreshold of 0.05 (1 - (fraction of sequences with a gap allowed)) and a minimum of 60% of the positions in the original alignment conserved (-cons 60 parameter). In addition, clusters containing sequences of the same subunits were pooled together to produce joined multiple sequence alignments. Sequences, in which SDH/FRD subunit fusions had been identified, were manually split. In cases, where a fusion involved additional domains unrelated to SDH/FRD complexes, these were trimmed from the alignment, whereby a split residue was determined using CD SEARCH domain assignment. Clusters containing membrane anchor subunits of types E and F were kept separate due to no homology to membrane subunits to other SDH/FRD types. Type B membrane anchor sequences were split into "SdhC" and "SdhD" based on the number of helices predicted by TMHMM (Hägerhäll 1997). Since the $SdhC_B$ sequences would only contain two helices, compared to three helices of SdhCs of other types, other SdhC sequences were split after the second helix. For some sequences, TMHMM failed to predict any helices, therefore TMPred (Hofmann, Stoffel 1993) was used. If TMPred predicted helices where TMHMM did not, the TMPred prediction was used to split or cut the sequences. This affected 8 sequences. Additionally, a structural alignment of $SdhC_F$ and $SdhC_B$ sequences was performed using Expresso mode of T-Coffee (Notredame, *et al.* 2000; Armougom, *et al.* 2006; *W. succinogenes* structure, RCSB PDB code: 2BS2).

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 analyzing the available SDH/FRD structures using UCSF Chimera (version 1.14, Pettersen *et al*. 2004). *E. coli* SDH structure was used as an example (RCSB PDB: 1NEK). Additionally, $SdhC_B$ from *W. succinogenes* were visualized to show the conserved Histidines that bind the second heme group that is absent in *E. coli* SDH. The resulting alignments were used to reconstruct maximum-likelihood phylogenies in Iqtree (Minh *et al*. 2020) with 1000 ultrafast bootstraps (Hoang *et al.* 2018; significance threshold >= 95) and best model selection "-m TEST" (Kalyaanamoorthy *et al*. 2017) to resemble jModelTest/ProtTest (Darriba *et al*. 2012).

Phylogenetic reconstructions were rooted using the minimal ancestor deviation (MAD) method (version 2.22, Tria *et al*. 2017) with a modified script to keep bootstrap values (kindly provided by Giddy Landan). SDH/FRD type, functional, and taxonomic annotations were added to phylogenies, and the analysis was conducted using FigTree (version v.1.4.4, tree.bio.ed.ac.uk/software/figtree). Several phylogenetic reconstructions were performed for both SdhA and SdhB subunits with different sequence datasets to access the robustness of the MAD rooting position. First, long branches were excluded to assess the effect of the presence of a very short metagenomic sequence on the topology of SdhA phylogeny, and the presence of two MvhD sequences on the topology of SdhB phylogeny. In a different reconstruction, the entire clade of type B was removed to see if the dissimilarity of the cytoplasmic subunits of this type to the other types could skew the topology of the phylogenetic reconstructions. In a third reconstruction, the clade containing type F, TFR, and other homologous proteins was removed. Finally, a reconstruction in with the combined removal of type B clade and homologous clade was computed. In addition, phylogenetic reconstructions were redone with cluster reduction based on >=90% identity and same genus affiliation.

Similarity networks based on global identities of SDH/FRD membrane subunits were computed and visualized using Cytoscape (Shannon *et al*. 2003). To plot large networks, global identity relationships were reduced by 70%, keeping one representative sequence per genus. Additionally, all relationships below 30% were excluded from the networks. SdhE and SdhF proteins were excluded due to their non-homologous amphipathic nature, and were separately analyzed within their cluster (SdhE together with HdrB sequences). $SdhC_F$ sequences were kept.

2.10. SDH/FRD type classification

The SDH/FRD type for each sequence was determined using a combinatory analysis of synteny, best hit relationships, global identity to queries with a defined type, number of conserved heme-binding histidines in the anchor module, phylogenetic and similarity network analysis, and MCL cluster information. Complete complexes in sinteny were classified based on the number and type of membrane subunits, taking into account the number of histidines present within the membrane anchor. Cytoplasmic subunits in synteny with membrane subunits, their type was assigned according to the type of the membrane subunits. Otherwise, for cases where SDH/FRD cytoplasmic subunits were non-syntenic or cases where membrane anchors were not identified, the best reciprocal hit relationships of the query sequences to this sequence were inspected, and if all respective query sequences belonged to the same SDH type, the query sequence type was assigned . However, if the respective query sequences belonged to more than one SDH type, the SDH type was assigned based on the hierarchical clustering of global identity relationships between the SDH sequences of the same subunit. In this analysis, a matrix of the global identity relationships between all sequences of the same subunit was created, hierarchically clustered, and saved in the Newick format (using R, R Core Team 2020) to be analyzed in FigTree (version v.1.4.4, tree.bio.ed.ac.uk/software/figtree). Each protein was annotated with their SDH type or with a “No type” if the type was unknown. Based on the resulting clusters, the SDH type of “No type” sequences was resolved where it was possible, *i.e.* if a “No type” sequence is clustered strictly with type A sequences, it was annotated as type A.

However, SdhE sequences classification required additional steps. In such cases, KEGG assignment on its own is not a reliable parameter to classify a sequence as SdhE or HdrB. First, many SdhE sequences have no KO assigned and only include a vague description, *e.g.* “hypothetical protein”, “disulfide reductase”, “succinate dehydrogenase”, consistent with previous reports on cyanobacterial SdhE subunits (Lancaster 2011a). Second, although the HdrB sequences in here identified tend to be assigned as “HdrB2”, there is no specific KO for SdhE sequences, which results in the “HdrB2” KO assignment for some of these sequences as well. Moreover, they were found in the same cluster with true HdrB sequences, which further complicated the case. Therefore, these sequences were manually annotated as SdhEs only if they fulfilled certain conditions. First, the syntenic neighbors of the sequences were analyzed, and cases, in which soluble SDH/FRD subunits were present and HdrA and HdrC sequences were absent, were annotated as SdhE. In some cases, however, the genome contained more than one HdrB sequence, *e.g.* in Cyanobacteria genomes containing both SDH_E* and HDR complexes. The strategy to resolve such complex cases was to check if the HdrB sequences are in synteny with HdrAC subunits. In addition, these sequences were aligned with known HdrBs (*e.g.* from methanogens) and identified SdhEs (*e.g.* from *Acidianus ambivalens*), and a global identity analysis with hierarchical clustering was performed to aid in the distinction between SdhE and HdrB. If still no classification as SdhE was possible, the “HdrB” annotation was kept. The differentiation between $SdhE_E$ and $SdhE_{E*}$ was performed based on the absence of the SdhF subunit within the complex.

The SDH/FRD type annotation was used to analyze the taxonomic distribution of SDH/FRD complexes of different types, which was calculated as a percentage of genomes from a given taxon (phylum or class for Euryarchaeota, Crenarchaeota, Proteobacteria, and Firmicutes) that have an SDH/FRD complex of a given type. The resulting matrix was sorted by SDH type and NCBI taxonomy and used to create a heatmap (R pheatmap package, version 1.0.12; Kolde 2019; R Core Team 2020) for further analysis.

3. Results

The thesis consisted of the homology comparison of different structural SDH/FRD types, assessment of the taxonomic distribution of those types, and phylogenetic reconstructions of cytoplasmic subunits together with similarity comparison for membrane anchor subunits.

3.1. Similarity analysis

During the query gathering stage, sequences belonging to biochemically characterized enzymes were retrieved. To expand the query taxonomic diversity, additional sequences were collected from predicted genomic analysis. In total, 69 complexes (SDH and FRD) from 59 organisms were obtained, spanning 10 bacterial phyla (Bacteroidetes, Actinobacteria, Proteobacteria, Firmicutes, Deinococcus-Thermus, Cyanobacteria, Aquificae, Chlamydiae, Spirochaetes, Nitrospirae) and three archaeal phyla (Euryarchaeota, Crenarchaeota, Thaumarchaeota). Queries included both succinate dehydrogenase and fumarate reductase complexes, as well as epsilonproteobacterial methylmenaquinol:fumarate reductases from *C. jejuni* and *W. succinogenes*. A detailed description including references is given in Supplementary Table 6.1.

Sequences from homologous enzymes were also gathered. These included representatives from 7 homologous enzymes, namely L-aspartate oxidase (Sakuraba *et al*. 2002; Kawarabayasi *et al*. 1998; Sakuraba *et al*. 2008; Kawarabayasi *et al*. 2001; Mattevi *et al*. 1999), thiol:fumarate reductase (Bult *et al*. 1996; Heim *et al*. 1998), anaerobic *sn*-glycerol-3-phosphate dehydrogenase (Cole *et al*. 1988; Harrison *et al*. 2005), glycolate oxidase (White *et al*. 1999; Kaneko *et al*. 1996; Pellicer *et al*. 1996), adenosine-5'-phosphosulfate reductase (Meyer, Kuevert 2007; Pereira *et al*. 2011), and heterodisulfide reductases HdrABC (Buckel, Thauer 2013; Pereira *et al*. 2011; Ramos *et al*. 2015) and HdrDE (Buan, Metcalf 2010). A detailed description including references is given in Supplementary Table 6.1. Mapping of queries to their corresponding genomes showed that only in *Anaplasma phagocytophilum* and *Shewanella frigidimarina,* additional SDH/FRD complexes or duplicated subunits were found. The missing *Aquifex aeolicus* SdhA and *Mycolicibacterium smegmatis* SdhC of SDH1 subunits were found based on syntenic information from feature tables.

The functional annotation of queries allowed identification of the characteristic Pfam domains of different subunits, namely a FAD-binding domain (FAD_binding_2) and a fumarate reductase flavoprotein C-terminal domain (Succ_DH_flav_C) in subunit A, a [2Fe-2S] center binding domain (Fer2_3) and a [4Fe-4S] dicenter domain (Fer4_8, Fer4_17, Fer4_7, Fer4_10, Fer4, Fer4_21) in subunit B. $SdhC_{A\text{-}D}$ subunits and some of the $SdhD_{A,C}$ subunits contain a succinate dehydrogenase/fumarate reductase transmembrane subunit domain (Sdh_cyt). In addition, some type D subunits C and D contain a fumarate reductase subunit C (Fumarate_red_C) domain or a fumarate reductase subunit D (Fumarate_red_D) domain, respectively (Fig. 7). It is also possible for either subunit to not contain any detectable Pfam domains (type F, some $SdhD_{A,C}$ sequences). Finally, SdhE sequences contain a Cysteine-rich domain (CCG).

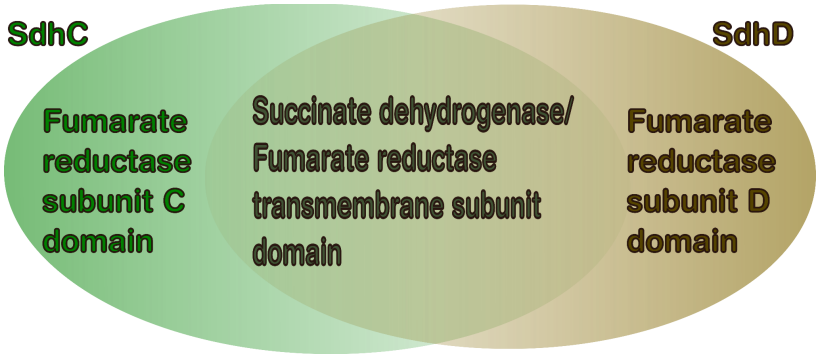


Fig. 7. Pfam domains of the canonical membrane subunits SdhC and SdhD (Type A to D).

Pfam annotation of homologous proteins showed that the characteristic for SdhA FAD-binding domain (FAD_binding_2, FAD_binding_4) is also present in SdhA homologous proteins. In addition, NadB, AprA, and TfrA sequences contain a flavoprotein C-terminal domain (Succ_DH_flav_C), indicating a possible higher level of homology to SdhA (Fig. 8).

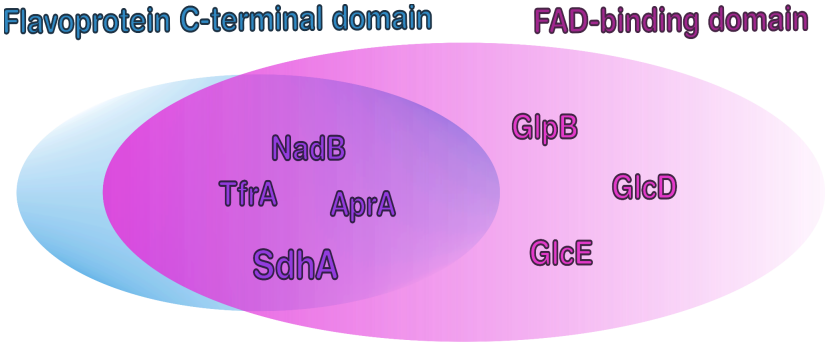


Fig. 8. Pfam domains of SdhA and its homologous proteins*. Proteins colored in pink contain a FAD-binding domain only. Proteins colored in purple contain both a flavoprotein C-terminal domain and a FAD-binding domain.*

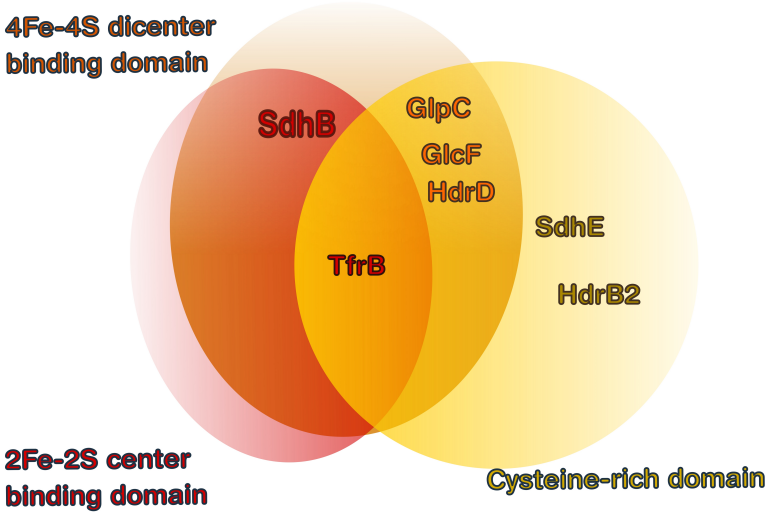


Fig. 9. Pfam domains of SdhB and SdhE and their homologous proteins. *Proteins colored in red contain both the 2Fe-2S center domain and the 4Fe-4S dicenter domain. Proteins colored in orange contain a 4Fe-4S domain and a Cysteine-rich domain. Proteins colored in yellow contain a Cysteine-rich domain only.*

SdhB homologous proteins share FeS domains of SdhB (a [2Fe-2S] center binding domain (Fer2_3) and a [4Fe-4S] dicenter domain (Fer4_8, Fer4_7, Fer4)) with TfrB and GlpC, having, in addition, a CCG domain similar to the one present in SdhE (Fig. 9).

Transmembrane helices prediction has shown that $SdhC_{A,C,D}$ (39 sequences) and $SdhC_{A,C,D}$ (38 sequences) contain three predicted transmembrane helices, which is in agreement with previous reports (Hägerhäll 1997; Lancaster 2011b). $SdhC_{B}$ (15 sequences) had five predicted transmembrane helices as expected due to the proposed fusion of SdhC and SdhD in this type (Hägerhäll, Hederstedt 1996; Hägerhäll 1997). However, in two SdhC sequences, there were also unconventional numbers of 2 and 4 helices predicted. Those are the cases of *Vibrio cholerae* $SdhC_{C}$, in which four transmembrane helices were predicted, and *Nitrososphaera viennensis* $SdhC_{A}$, where only two transmembrane helices were predicted. This could be due to losses or acquisition of helices within these subunits, prediction or misassembly errors. Due to the amphipathic nature of SdhE (13 sequences) and SdhF (5 sequences), no transmembrane helices were predicted for these proteins, as described in Lemos *et al*. 2001. Most of the cytoplasmic subunit sequences (SdhA and SdhB) have zero predicted transmembrane helices, which correlates with literature reports on their soluble nature and their cytosolic location (Hägerhäll 1997). However, in three SdhA sequences, one transmembrane helix was potentially identified. This is likely due to uncertainty in the prediction algorithm, assembly artifacts, or fusion events.

To estimate the homology levels between the query subunits, an all versus all pairwise global identities per subunit were calculated and hierarchically clustered (shown in Fig. 10 and 11 and Supplementary Fig. 7.2 and 7.3 for subunits A, C, B, and D, respectively). In the case of subunits A and B, sequences of the same type form clearly defined clades: two high identity groups of type C, one of type D, two of type E/E*, and three high identity clades in type A. Type B is divided into three groups (one includes one sequence of type A and one sequence of type F). For soluble subunits, the homology varies from high (~60% global identity) to low (25% global identity), sometimes even for the sequences of the same type, which indicates possible subgroups for types A, B, and C. For the membrane subunits C and D, in contrast to the cytoplasmic subunits, sequences of canonical types A to D do not form high identity groups. However, there is a high identity group formed by SdhE sequences, which is due to their less hydrophobic nature (and therefore, higher amino acid sequence conservation) in comparison to other SdhCs.

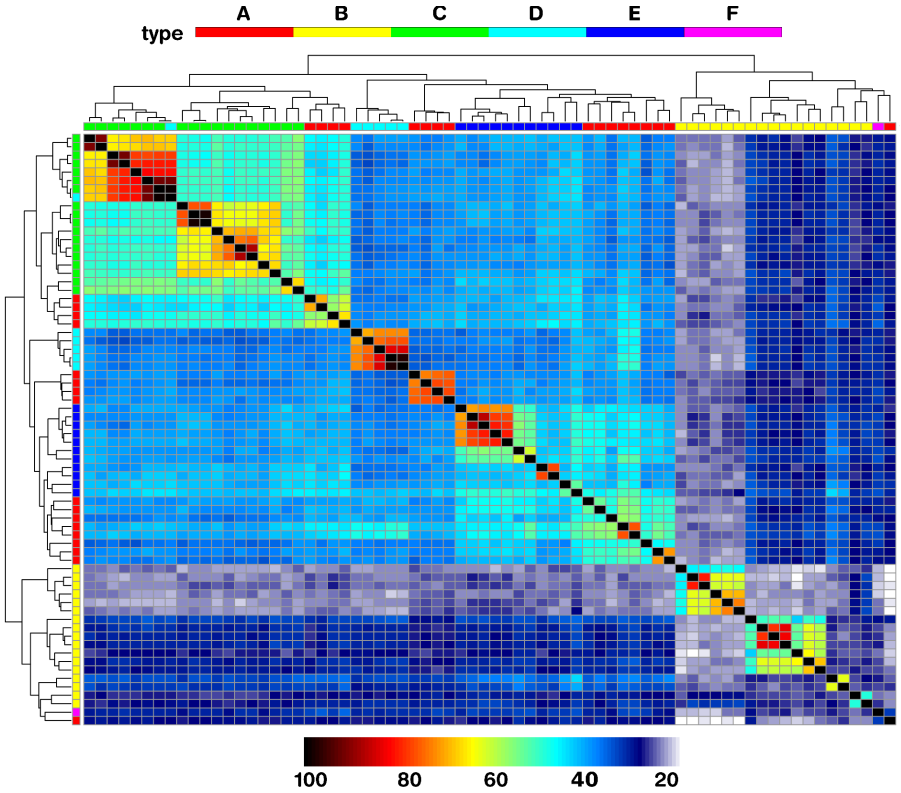


Fig. 10. All-vs-all global identity of subunit A query sequences. *Bars indicate sequence type with type A colored in red, type B in yellow, type C in green, type D in cyan, type E in blue and type F is colored in magenta. Bottom scale indicates percentage of identity.*

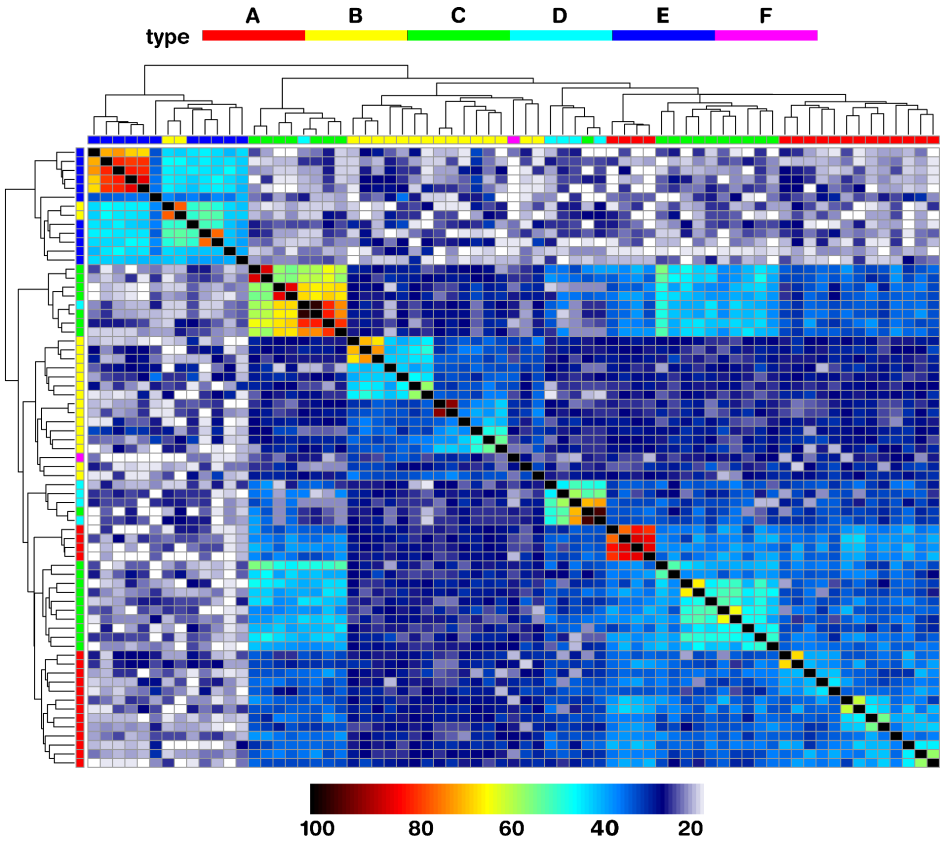


Fig. 11. All vs all global identity of subunit C query sequences. *Color code as in Fig. 10.*

Having characterized the different query sequences in terms of their domain content and similarity, a large scale search of SDH/FRD within the genomic dataset was performed using the reciprocal best blast hit approach. The rBBH search for homologous proteins identified 270215 sequences (201016 of those unique), with 87278 sequences being annotated as SDH/FRD subunits. After filtering for 25% identity, the global alignment of these sequences produced over 1.5 billion relationship pairs. SDH/FRD sequences were identified in 77% (26894) of genomes in the dataset, being absent from 8122. During this search, it was observed that in 4098 genomes (11.7% of the dataset) no hits were identified. Most of these genomes are metagenomic assemblies, with different levels of completeness, so it is not possible to say with certainty whether these organisms have no SDH/FRD and its homologous proteins, or the assemblies are simply missing those sequences. These cases could be divided into three groups. The first group includes well-studied taxa, such as, for instance, Firmicutes (921/5536), Fusobacteria (42/70 genomes), Proteobacteria (147/13364), Spirochaetes (23/216 genomes), Crenarchaeota (13/125), and 76/98 Thermotogae genomes. However, SDH/FRD and its homologous complexes were identified within these taxa suggesting possible gene losses or genomic incompleteness.

The second group consists of genomes belonging to archaeal DPANN group (138/140) and bacterial Tenericutes (248/248). These taxa are known for having extremely reduced genomes (Huber *et al*. 2003; Feng *et al*. 2021; Dombrowski *et al*. 2019; Sirand-Pugnet *et al*. 2007), therefore it was expected that they might not have any SDH/FRD complexes. The third group includes genomes affiliated with Candidate Divisions or *Candidatus* taxa. In these cases, further investigation is needed to resolve metabolic capabilities of these organisms.

3.2. MCL clustering and functional annotation of clusters

The MCL clustering produced 105 clusters, with the largest cluster containing 20582 sequences, and with 10 clusters containing only two sequences. SDH/FRD subunit were present in 27 clusters, while the remaining contained other complexes used as queries. Retrieved SDH/FRD sequences were preliminarily functionally classified to a type according to their best hit type and queries found in the respective cluster. For sequences of cytoplasmic subunits with relationships to queries of different types, this classification was performed by synteny analysis and/or global identity hierarchical clustering analysis (see Section 2.8). After functional annotation and inspection of the number of histidines in the membrane anchor module, the composition of clusters became clear (Table 1). While SdhA and SdhB sequences were grouped into three or four different clusters, respectively (regardless of the SDH type), the membrane anchor subunits were found into a higher number of clusters: 12 in the case of SdhC and 8 in the case of SdhD. Membrane type C sequences were found in four of these clusters, type B in five and type A membrane subunits spread among 8 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, fusion sequences of SdhC and SdhD subunits belonging to Chloroflexi were found in one cluster. Due to their high similarity, the non-canonical membrane subunits SdhE protein (this definition includes any membrane subunit of amphipathic nature, *e.g.* characterized *Acidianus ambivalens* SdhE and *Campylobacter jejuni* MfrE) were found in the cluster containing also HdrB

proteins. SdhF proteins formed their own single cluster. Within the largest cluster, in addition to SdhA sequences, also TfrAs and other closely related sequences which could not have been differentiated from SdhA by functional annotations were present. However, these sequences possibly do not belong to Sdh/FRD complexes since they are found in genomes that either do not contain any other subunits, or they contain additional complete syntenic complexes. The soluble fumarate reductase from *Shewanella* (FccA; Leys *et al*. 1999; Pankhurst *et al*. 2006) is found in cluster 13, however, little is known about the other proteins in this cluster. Non-SDH/FRD clusters were annotated by majority rule for simplicity (Table 1). The detailed cluster annotation is given in Supplementary Table 6.3 and the Taxonomic distribution of genomes (per class) per MCL cluster is shown in Supplementary Fig. 7.4.

Proteins in cluster	**Cluster numbers**
SdhA + TfrA, other close SdhA homologous sequences	1,12,58
SdhB, TfrB + SdhB-QcrB fusions	3,8,30,49,64
SdhC	15,17,22,23,24,29,34,36,40,43,50,94
SdhD	16,18,26,37,44,45,53,76
SdhE + HdrB, partial HdrBs	20
SdhF	67
SdhC and SdhD fusions	79
NadB, KstD, various FAD-binding proteins	2,27,48,55,56,70,71,72,92,96,102
GlcDEF	4,6,9,10,28,39,59,60,63,66,69,78,81,85,86,87,90,93
GlpABC/D	5,7,11,25,28,31,35,39,61,62,65,66,77,97,98,99,100,101
Flavocytochrome c + FccA/FrdA, cytochrome c3 fusions	13,82
HdrABC, QmoC, AprAB, HdrED	19,20,21,32,33,38,42,47,51,54,57,73,75,80,83,84,91,104,105
Other (LldEF, SoxB, ThiO, DadA,GcvT,ThiG,CobZ/TcuA,Ferredoxin)	14,41,46,52,68,74,88,89,95,103

Table 1. Functional annotation of MCL clusters*. Clusters containing SDH/FRD subunits are color-coded: SdhA in green, SdhB in pink, SdhC in orange, SdhD in blue, SdhE in yellow, SdhF in red, and SdhC+SdhD fusions in purple. Abbreviations: L-lactate dehydrogenase subunits E and F (LldE and LldF); 3-oxosteroid 1-dehydrogenase (KstD); Beta subunit of sarcosine oxidase (SoxB); Glycine oxidase (ThiO); Quinone-modifying oxidoreductase subunit C (QmoC); Glycine cleavage system T protein (GcvT); Glycine/D-amino acid oxidase (deaminating) (DadA); bc complex cytochrome b subunit (complex III) (QcrB); Thiazole synthase (ThiG); Tricarballylate dehydrogenase (CobZ/TcuA); soluble fumarate reductase from Shewanella (FccA/FrdA).*

Homologous proteins of subunits A and B annotated by KEGG and grouped by Pfam domains are shown in Fig. 12 and Fig. 13, respectively. Besides the homologous sequences already present in the query set, additional homologous to SdhA proteins were identified, such as urocanate reductase (UrdA; Bogachev *et al*. 2012), tricarballylate dehydrogenase (CobZ/TcuA; Lewis, Escalante-Semerena 2006), alkyldihydroxyacetonephosphate synthase (AgpS; Zomer *et al*. 1999), 3-oxo-5alpha-steroid 4-dehydrogenase (TesI; Florin *et al.* 1996), D-lactate dehydrogenase (Dld; Taguchi, Ohta 1991), and 3-oxosteroid 1-dehydrogenase (KstD; Plesiat *et al*. 1991) (Fig.

12). All of these homologs contain a FAD-binding domain and are dehydrogenases or oxidoreductases with the exception of AgpS, which is a synthase.

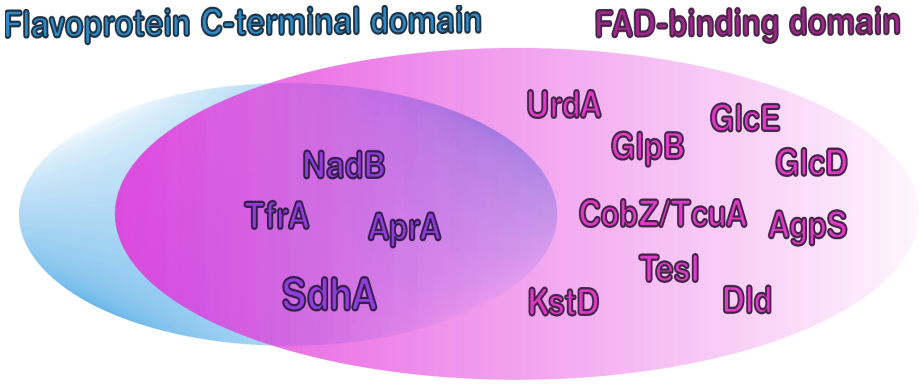


Fig. 12. SdhA homologous proteins, found via rBBH approach. *Annotated by KEGG, grouped by Pfam. Color code as in Fig. 8.*

Additional homologous proteins of SdhB include ferredoxin and HdrC2 proteins, both of which contain FeS clusters. In the case of SdhE, also LldE, a Cysteine-rich domain-containing protein involved in lactate utilization (Pinchuk *et al*. 2009) was identified as a homolog (Fig. 13).

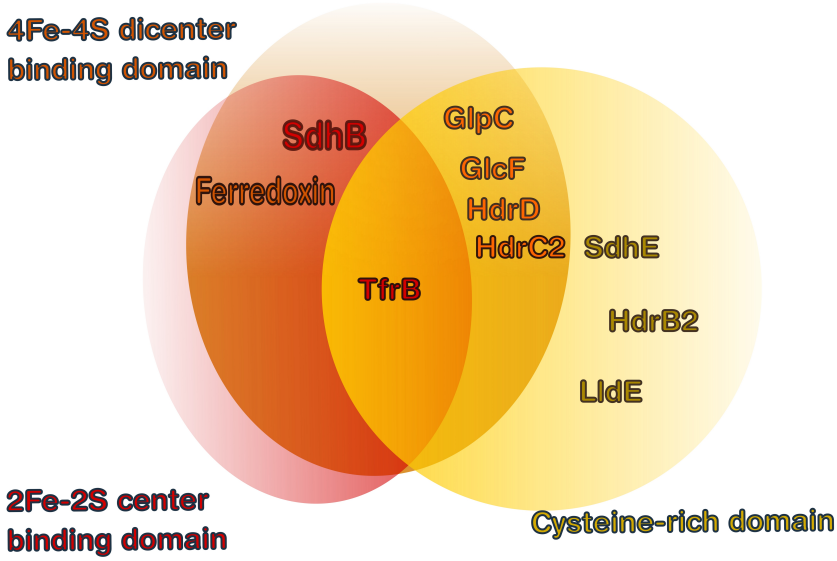


Fig. 13. SdhB and SdhE homologous proteins, found via rBBH approach*. Annotated by KEGG, grouped by Pfam. Color code as in Fig. 9.*

3.3. Mean intercluster global identity

To assess the performance of the clustering procedure, the mean, median and maximum global identities between clusters was calculated and hierarchically clustered. As can be seen in Fig. 14, this allowed the identification of several groups of clusters that share homology between themselves corresponding to either the same subunit or close homologues as the case of SdhA, TfrA, NadB, and FccA clusters. Clusters on the bottom right correspond to proteins not part of SDH/FRD complexes. Overall, sequences from each subunit form solid clusters, based on the mean intercluster identity. A similar

pattern is also observed in the median intercluster identity heatmap (Supplementary Fig. 7.5). As expected, the maximum intercluster identity values for clusters of subunits SdhA and SdhB are higher than the identities between the clusters of membrane subunits SdhC and SdhD (Supplementary Fig. 7.6). This analysis allowed to assess the quality of the clustering procedure and was also useful for the choice of clusters to combine for phylogenetic analysis (see below).

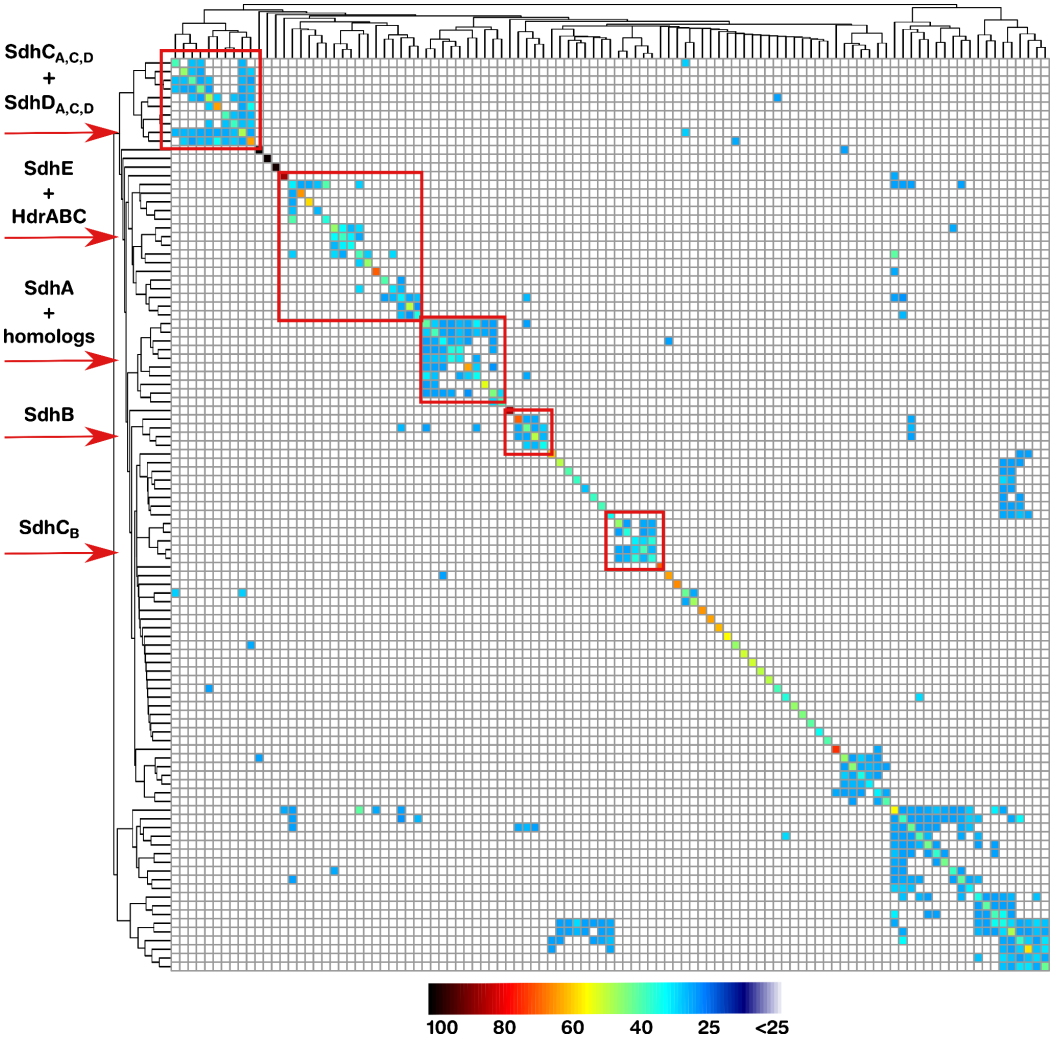


Fig. 14. Mean intercluster global identity, hierarchically clustered.

3.4. Taxonomic distribution of SDH/FRDs

The classification of sequences allowed to analyze the overall distribution of complexes per type (Fig. 15 and Supplementary Table 6.4). In this dataset, 31944 complete SDH/FRDs complexes and 2239 incomplete (lacking at least one subunit) were identified (Supplementary Table 6.5). Inspection of the existence of pseudogenes within genomic assemblies revealed that in 725 cases, the identification of incomplete complexes might be due to assembly artifacts. However, the remaining 1514 cases open the possibility of the existence of novel modular architectures within this family and pinpoints enzymes to be biochemically characterized. For 1780 sequences, no type classification could be attributed (see Methods). The taxonomic distribution of each type is described in detail below.

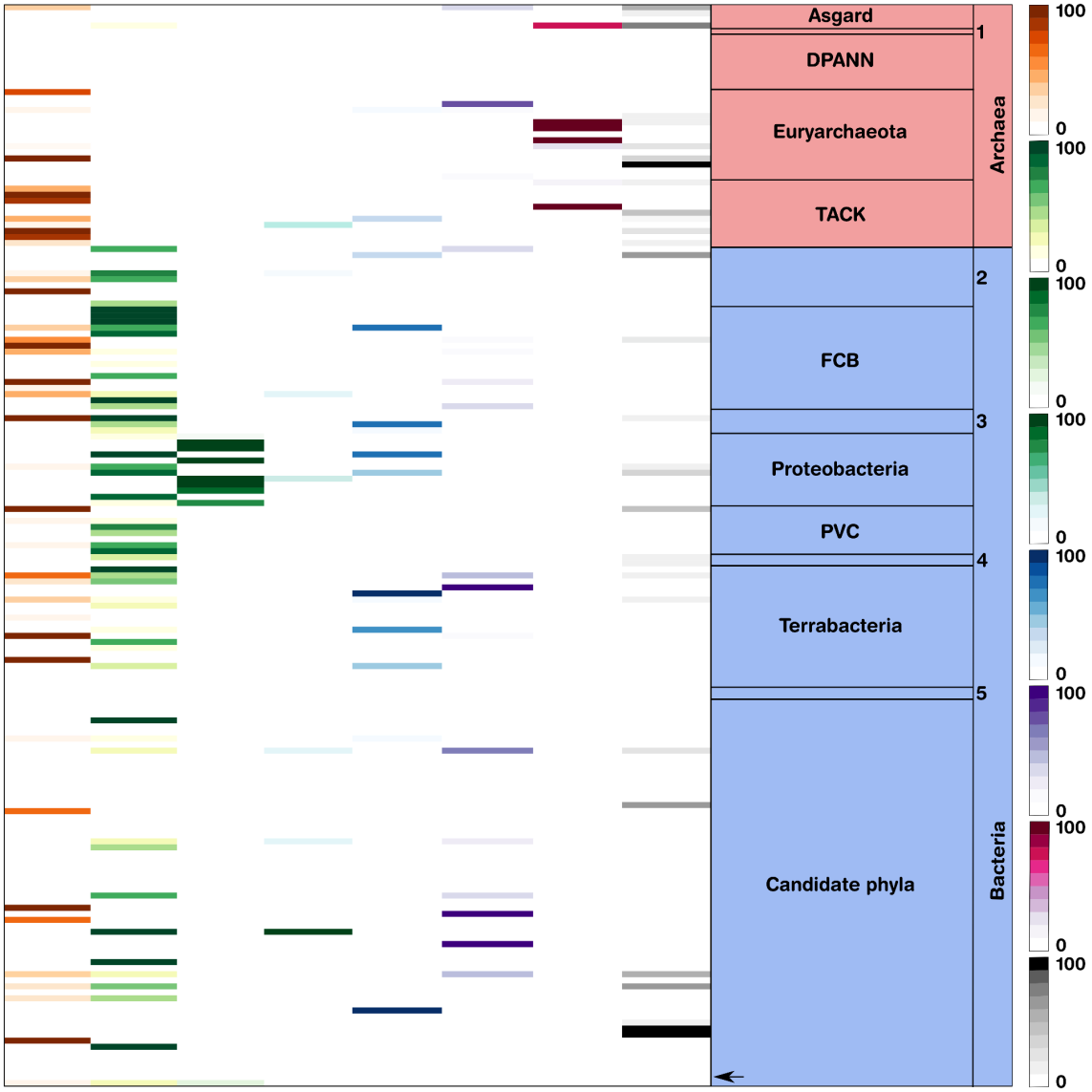


Fig. 15. Taxonomic distribution of SDH/FRD complexes per type*. Type A is colored in orange, type B in yellowish-green, type C in dark green, type D in light blue, type E in dark blue, type F in purple, and Tfr in pink. Proteins for which the type classification was not possible are grouped in the "No type" column and colored in gray. White color indicates an absence of type in the lineage. Taxonomic affiliations are phylum or class. Taxa are sorted by NCBI taxonomy from the NCBI newtaxdump version of 2020. Taxonomic supergroups are labeled where applicable. Rows containing taxa without a supergroup affiliation are numbered. Row group "1" contains Candidatus Hydrothermarchaeota. Row group "2" contains unclassified Archaea, Acidobacteria, Aquificae, Caldiserica, Candidatus Cryosericota, Calditrichaeota, Chrysiogenetes, Coprothermobacterota, Deferribacteres, Dictyoglomi, and Elusimicrobia. Row group "3" contains Fusobacteria, Candidatus Tectomicrobia, Nitrospinae, and Nitrospirae. Row group "4" contains Spirochaetes and Synergistetes. Row group "5" contains Thermodesulfobacteria and Thermotogae. A row marked with an arrow represent collapsed rows of Candidate phyla with no SDHs detected. The last row of Archaea section contains genomes affiliated with unclassified Archaea annotation, and the last row of Bacteria contains genomes affiliated with unclassified Bacteria annotation.*

3.4.1. Type A

Type A enzymes are taxonomically the most diverse type, and include characterized complexes from *Mycolicibacterium smegmatis* (Hards *et al*. 2019), *Halobacterium salinarum* (Gradin *et al*. 1985), *Natronomonas pharaonis* (Scharf *et al*. 1997), *Rhodobacter sphaeroides* (Barassi *et al*. 1985), *Micrococcus luteus* (Crowe *et al*. 1983), *Thermus thermophilus* (Kolaj-Robin *et al*. 2011), and *Thermoplasma acidophilum* (Anemüller *et al*. 1995). These complexes are widespread in both Archaea and Bacteria, being present in 8 archaeal phyla and 40 bacterial phyla. In Archaea, type A SDH/FRD complexes were identified in the majority of Archaeoglobi (8/10 genomes), Halobacteria (over 90% of 388 genomes present in the dataset), Korarchaeota (all 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 (four out of five genomes), *Candidatus* Marsarchaeota (11/14 genomes), and Crenarchaeota (over 40% of 125 genomes). Interestingly, in *Acidianus ambivalens* genomes, besides the canonical type E complex experimentally characterized by Lemos *et al* 2001, an incomplete SdhBCD complex of type A was identified. Inspection of the surrounding genes did not allow identification of other proteins that could replace the flavin subunit. A full list of archaeal phyla containing SDH complexes of type A is given in Supplementary Tables 6.4 and 6.5. In Bacteria, this type is widespread in Actinobacteria (66% of 5489 genomes), Deinococcus-Thermus (over 90% of 95 genomes), Deferribacteres (11/11 genomes), Chlorobi (35% of 34 genomes), Rhodothermaeota (6/11 genomes), *Candidatus* Aminicenantes (8/13 genomes), *Candidatus* Tectomicrobia (6/6 genomes), *Candidate* Division Zixibacteria (9/18 genomes), *Candidatus* Kryptonia (in all four genomes), and *Candidatus* Marinimicrobia (6/12 genomes). In addition, type A complexes are scarcely present (less than 20% of genomes) in 30 phyla (full list in Supplementary Tables 6.4 and 6.5). 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 characterized enzyme from *Rhodothermus marinus* PRQ32B albine strain belongs to type B (Fernandes *et al*. 2001; 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.

3.4.2. Type B

Type B complexes include the characterized complexes from *Bacillus cereus* (Garcia *et al*. 2008), *Bacteroides fragilis* (Baughn, Malamy 2003), *Bacteroides thetaiotaomicron* (Lu, Imlay 2017), *Helicobacter pylori* and *Campylobacter jejuni* (Mileni *et al*. 2006), *Desulfovibrio gigas* (Guan *et al*. 2018), *Geobacter sulfurreducens* (Butler *et al*. 2006), and *Wolinella succinogenes* (Lancaster *et al*. 1999). This type is widespread among bacterial lineages, being present in a total of 49 bacterial phyla. Type B SDH/FRDs 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), Ignavibacteriae (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, succinate dehydrogenases of type B are widely distributed in some 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 scarcely present or entirely absent in others. In Cyanobacteria, complexes of this type are present in only 15% of 556 metagenomic assemblies. A full list of bacterial phyla containing SDH complexes of type B is given in Supplementary Tables 6.4 and 6.5. However, type B SDH/FRDs are almost entirely absent from Archaea, with the exception of a few metagenomes from *Candidatus* Thorarchaeota (one out of five), Methanobacteria (three out of 66 genomes), unclassified Euryarchaeota (one of 73 genomes), and unclassified Archaea (two of 49 genomes).

3.4.3. Type C

Type C enzymes include the well-studied *E. coli* SDH (Yankovskaya *et al*. 2003) and mitochondrial enzymes (Bezawork-Geleta *et al*. 2017). Besides eukaryotes, these complexes are mainly present in Proteobacteria, with exception of one metagenome from Bacilli, one Bacteroidetes, one Actinobacteria, and ~20% of unclassified Bacteria genomes. The Bacilli metagenome has been in between reclassified as Gammaproteobacteria (NCBI, 2020) and the other two lineages are represented by thousands of genomes, none of which contain a type C SDH, which makes it possible that these single type C occurrences are contamination in the metagenomic assembly.

3.4.4. Type D

Type D, which *E. coli* FRD belongs to (Cecchini *et al.* 2002), has a restricted taxonomic distribution when compared to type A or B. In Bacteria, with exception of Gammaproteobacteria, where it is present in 30% of 5777 metagenomic assemblies, type D is scarcely present across 17 phyla, such as Acidobacteria, Calditrichaeota, Bacteroidetes, *Candidatus* Marinimicrobia, Gemmatimonadetes, and Nitrospirae. The full list of type D-containing taxa is given in Supplementary Tables 6.4 and 6.5. Although in this analysis, 11 complete complexes of type D were identified in Archaea, the actual presence of this type in Archaea is in question, as it was only found in unclassified metagenomes (four out of 73 unclassified Euryarchaeota genomes, four out of 8 unclassified Crenarchaeota genomes, and four out of 49 unclassified Archaea genomes). Thus, it is not clear if type D is truly present in Archaea, potentially due to the lateral gene transfer events, or these results are a consequence of assembly artifacts.

3.4.5. Type E/E*

Canonical type E complexes (*i.e.* with two amphipathic membrane anchor subunits SdhE and SdhF) were identified only in Archaea, in the Thermoprotei class of Crenarchaeota (~32% of 117 genomes), in the Thermoplasmata class of Euryarchaeota (~14% of 58 genomes), and in one unclassified Euryarchaeota genome. Interestingly, a variation of type E architecture, containing an SdhE subunit but lacking SdhF (in here denoted as type E*), were found in Bacteria. This type was found in 15 bacterial phyla, predominantly in Aquificae (50% of 40 genomes, excluding NADH-dependent soluble FRDs), Chlorobi (74% of 34 genomes), Nitrospinae (74% of 19 genomes),

Cyanobacteria (62% of 556 genomes), Negativicutes (42% of 189 genomes), and Epsilonproteobacteria (48% of 431 genomes). The MFR complexes of Epsilonproteobacteria were not distinguishable from other complexes of this type, and therefore are not differentiated. The full list of taxa containing type E* complexes is given in Supplementary Tables 6.4 and 6.5.

3.4.6. Type F

The newly discovered type F (Hards, *et al*. 2019) was identified in four archaeal lineages and 28 bacterial lineages. In Archaea, it is mostly present in *Candidatus* Poseidoniia (80% of 15 metagenomic assemblies), and Euryarchaeota (14% of 58 Thermoplasmata and 17% of 73 unclassified genomes). In Bacteria, this type was predominantly detected in at least 30% of metagenomic assemblies affiliated with Actinobacteria, Acidobacteria, Gemmatimonadetes, Candidate Division NC10, and *Candidatus* Rokubacteria. The full list is given in Supplementary Tables 6.4 and 6.5. Of note, in the multiple sequence alignment of $SdhC_F$ sequences, four strictly conserved histidines were identified. This type was reported not to contain heme cofactors and has been only studied in one organism so far (Hards *et al*. 2019). Thus, it was not clear whether these histidines could have been heme-binding, as in the case of other SDH/FRD types. This was further investigated by performing a structural alignment of $SdhC_F$ and $SdhC_B$ proteins using *W. succinogenes* structure as template (Magej *et al*. 2006). The resulting alignment has shown that the type F conserved histidines do not align with the heme-binding histidines of $SdhC_B$, being located at a different structural position, and therefore unlikely to be related relics of histidine ligands of the hemes present in other types.

3.4.7. Thiol:fumarate reductase (TFR)

Thiol:fumarate reductases are enzymes that perform the conversion of fumarate to succinate using Coenzyme M - Coenzyme B as an electron donor (Heim *et al*. 1998). 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 reductase was detected predominantly in Archaea (four phyla) and scarcely identified in 9 bacterial phyla. In Archaea, TFR (and TFR-like) complexes mainly occur in Euryarchaeota (mostly in methanogens, but also in some Halobacteria and Archaeoglobi organisms), as well as *Candidatus* Thorarchaeota and *Candidatus* Bathyarchaeota. In Bacteria, this complex is only found in 19 metagenomic assemblies, most of them belonging to *Candidatus* Roizmanbacteria and Deltaproteobacteria. The full list (including bacterial TFR-containing phyla) is given in Supplementary Tables 6.4 and 6.5.

3.4.8. “No type” SDH/FRD hits

In addition to complexes of types A to F and TFR, there were 1780 cases of SdhA and SdhB sequences, for which no type could be assigned. Such sequences were either in synteny (SdhAB) but had no membrane subunits or were singular sequences on their own (SdhA only). Many of such SdhAB complexes were close to type A, type B or type E/E* by global identity and were placed close to or within the clades of these types in phylogenetic reconstruction. In addition, some were found in the clade containing TFRs

and NadBs. Often, these hits were found in genomes that also contain a full SDH complex of a defined type. Some of the genomes containing these “No type” sequences were annotated as “Complete”, therefore the SDH type could not be assigned with certainty in such cases. Overall, proteins, for which “No Type” classification was attributed, were found in 8 archaeal and 30 bacterial phyla. A full list of phyla containing “No type” sequences is given in Supplementary Table 6.4 and 6.5. Interestingly, NADH-dependent fumarate reductase complexes previously characterized in Aquificae (Miura *et al*. 2008; EC 1.3.1.6) were found in 14% of Aquificae genomes, being identified by the presence of the FrdCDE subunits. However, by global identity relationships their corresponding FrdA and FrdB subunits are closely related to type E* SdhA and SdhB.

3.5. Genomic organization of SDH/FRD complexes

In terms of syntenic arrangement of SDH/FRD subunits, these complexes tend to be organized into mostly syntenic blocks (Baughn, Malamy 2003; Butler *et al.* 2006; Guccione *et al*. 2010; Janssen *et al.* 1997; Park *et al*. 1995; Pecsi *et al*. 2014; Schäfer *et al*. 2002; Westenberg, Guerinot 1999; Hägerhäll 1997; Jardim-Messeder *et al*. 2017; Lemos *et al*. 2002). However, variations in the non-syntenic arrangement are also possible (Hederstedt 2002; Massung *et al*. 2008; Jardim-Messeder *et al*. 2017). In this analysis, SDH/FRD complexes from Aquificae, Cyanobacteria, as well as the TFR complexes found in methanogens were found to be predominantly non-syntenic, while within other archaeal and bacterial phyla they were organized into syntenic blocks, in agreement with previous reports (Jardim-Messeder *et al*. 2017; Schäfer *et al.* 2002). In these blocks, the cytoplasmic subunits are in proximity to each other, followed (or preceded) by the membrane subunits. It is rare to find the cytoplasmic subunits being interrupted by membrane subunits or being surrounded by a membrane subunit from each side. Moreover, the neighborhood of SDH/FRD complexes varies strongly from class to class and phylum to phylum. Fig. 16 shows syntenic arrangements of SDH/FRD subunits for each structural type. Enzymes of types A, B, C, and E* have various syntenic arrangements, while for those in types D, E, and F only one genomic arrangement was found. For comparison, genomic rearrangement of TFR complexes is shown, where it can be observed that subunits of this complex can be localized in the same genomic region or be non-syntenic.

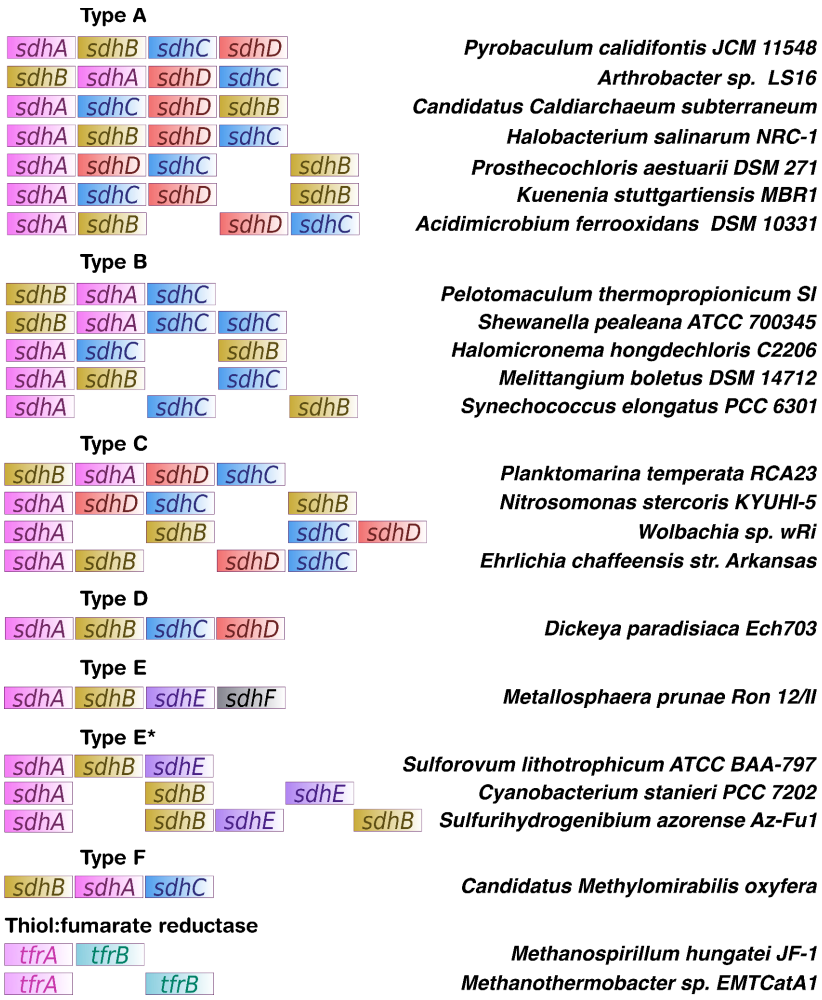


Fig. 16. Syntenic arrangement of SDH/FRD and TFR subunits with representative organisms per structural type. *Subunit A is colored in magenta, subunit B in yellow, subunit C in blue, subunit D in red, subunit E in purple, subunit F in gray, subunit TfrA in light pink, subunit TfrB in light blue. Type E* indicates complexes that contain SdhE subunit, but no SdhF subunit was identified.*

3.6. Phylogenetic reconstructions

This section describes the maximum likelihood phylogenetic reconstructions for cytoplasmic subunits as well as similarity network analysis of membrane anchor subunits.

3.6.1. Multiple sequence alignment

In order to build phylogenetic reconstructions, clusters were collapsed on the basis of redundancy reduction, as described in Methods. Sequences with >=90% relationships from SDH/FRD clusters that contained more than 1000 sequences were reclustered, and one representative sequence per subcluster was used for the alignment (Supplementary Table 6.6). Reduced clusters were used to produce multiple sequence alignments, and their quality was manually assessed (see Methods), using data from X-ray structures to determine important residues as well as the position of the histidine ligands of the hemes. Fig. 17 shows the available X-Ray crystallographic structures of complex II from the different types, some of which were used to check the quality of the alignments. It

can be seen that *E. coli* succinate dehydrogenase of type C contains one heme group, bound between SdhC and SdhD, while *E. coli* fumarate reductase of type D contains no heme groups, as expected. Both *W. succinogenes* and *D. gigas* FRDs of type B contains one large membrane subunit with two hemes bound in the center of it. The SDH2 of *M. smegmatis* (type A) contains two hemes, bound between the $SdhC_A$ and $SdhD_A$ subunits, and an additional single alpha-helix subunit SdhF, as proposed by Gong *et al.* 2020.

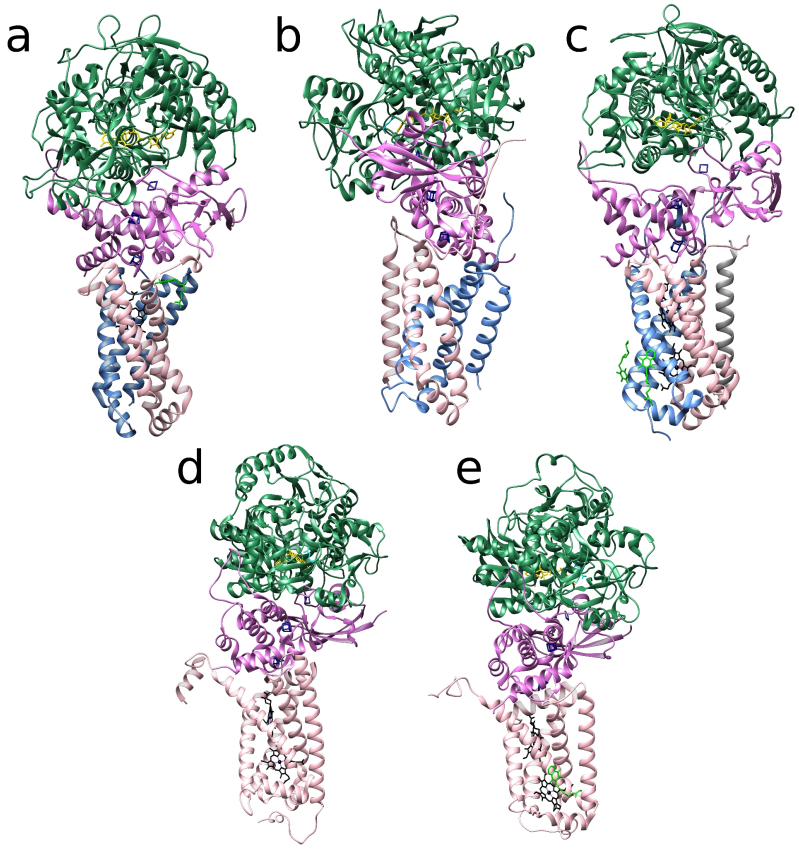


Fig. 17. X-Ray crystallography structures of complex II *(from a to e, respectively): Structure of E. coli succinate dehydrogenase (PDB: 1NEK, Yankovskaya et al. 2003); structure of E. coli fumarate reductase (PDB: 3P4P, Tomasiak et al. 2011); structure of M. smegmatis succinate dehydrogenase 2 (PDB: 6LUM, Gong et al. 2020); structure of W. succinogenes fumarate reductase (PDB: 2BS2, Madej et al. 2006); structure of D. gigas fumarate reductase (PDB: 5XMJ, Guan et al. 2018). The flavoprotein subunit SdhA is highlighted in green, with bound flavin adenine dinucleotide cofactor colored yellow; iron-sulfur subunit B is purple with FeS clusters colored dark blue; SdhC is light pink and SdhD is light blue, with heme groups (when present) colored black. In case of the M. smegmatis structure, the newly proposed SdhF subunit is colored gray. Some structures include fumarate and quinones, colored cyan and bright green, respectively.*

The conserved residues for each subunit that were used to check the quality of multiple sequence alignments are shown in Fig. 18 and 19. SdhA subunit has strong conservation in catalytic and FAD-binding region (Hägerhäll 1997; Cecchini *et al.* 2002; Iverson 2013; Hederstedt 2002; Lancaster 2002a) with several conserved residues (*E. coli* SDH structure numbering): His 45, His 242, His 354, Arg 286, Arg 399, and Thr 254 (Fig. 18a; Yankovskaya *et al*. 2003).

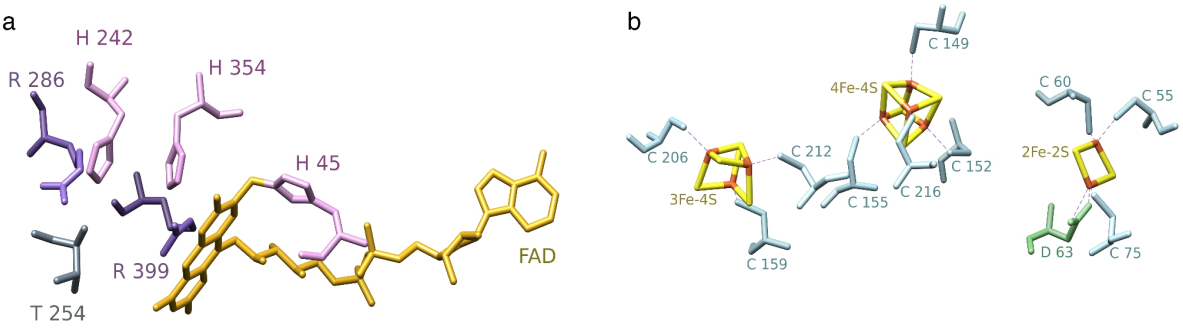


Fig. 18. Visualization of the conserved residues used to check multiple sequence alignments of SdhA and SdhB subunits. *SdhA E. coli SDH (PDB: 1NEK). Made in UCSF Chimera. As assigned by the one-letter amino acid code, H is Histidine (light pink), R is Arginine (purple), T is Threonine (gray). FAD is a flavin adenine dinucleotide cofactor (yellow). SdhB subunit, E. coli SDH (PDB: 1NEK). As assigned by the one-letter amino acid code, C is Cysteine (light blue), D is Aspartate (light green). FeS clusters are colored by the element, Fe atoms are orange, S is yellow.*

Since iron-sulfur centers in the SdhB subunit are bound by conserved cysteine residues (see Introduction; Hägerhäll 1997), the alignment of SdhB sequences was checked for conservation of these cysteines (Fig. 18b). Similar cysteine conservation was expected in SdhE (Lemos *et al*. 2001).As for transmembrane subunits, it was decided to check for alignment of heme-binding histidines, which are located in helix II of SdhC and helix II of SdhD (helix V of $SdhC_B$) for high-spin b_H heme, and in helix I of SdhC and helix I of SdhD (helix IV of $SdhC_B$) for low-spin b_L heme (Hägerhäll, Hederstedt 1996; Lancaster *et al*. 1999; Yankovskaya *et al*. 2003) (Fig. 19; PDB: 2BS2, Madej *et al.* 2006; PDB: 6LUM, Gong *et al*. 2020). In type C, containing a single heme, one histidine residue in each $SdhC_C$ and $SdhD_C$ (Cecchini *et al*. 2002, b_H) is expected. In type A and B, containing two heme groups, two histidines for each $SdhC_A$ and $SdhD_A$ (Schäfer *et al*. 2002) subunit, and four histidines for the fusion $SdhC_B$ subunit (Lancaster 2003) are expected respectively. In types D and F, no conserved histidine residues were expected since these types contain no hemes (Cecchini *et al*. 2002; Hards *et al*. 2019).

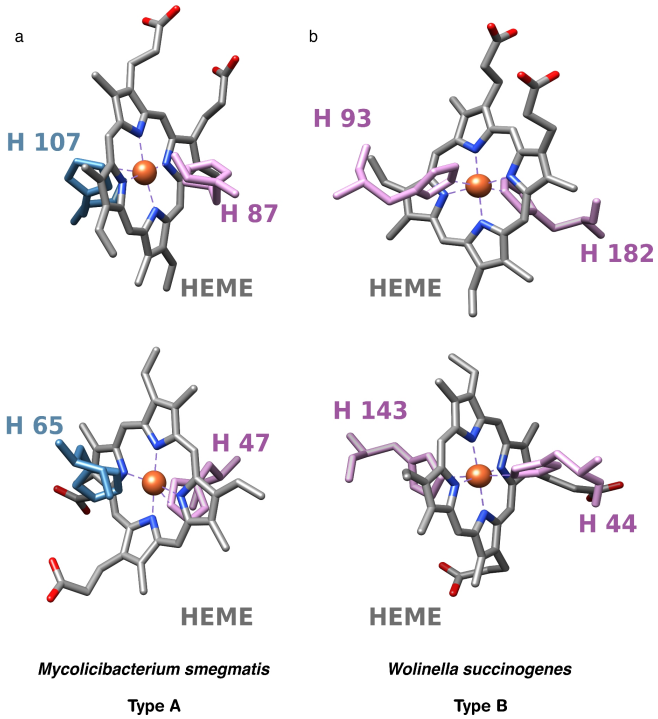


Fig. 19. Visualization of the conserved residues used to check the multiple sequence alignments of SdhC and SdhD subunits*. a) heme groups from M. smegmatis type A SDH; b) heme groups from W. succinogenes type B SDH. As assigned by the one letter amino acid code, H is histidine (pink for SdhC, blue for SdhD). Heme groups are colored by element.*

Quality check has led to the conclusion that in the multiple sequence alignments of the individual clusters, conserved residues were properly aligned. Therefore, multiple sequence alignments per subunit were calculated by joining clusters containing the same subunits. The SdhCs and SdhDs sequences, found by synteny analysis that were not found in the rBBH approach were also included. Clusters containing SdhE/HdrB, $SdhC_F$, and SdhF remained separated due to their low homology to other membrane subunits (See Fig. 14). Fusion sequences were removed if the fusion did not belong to an SDH/FRD subunit, or split and added to the respective group if the sequence is a fusion between two SDH/FRD subunits. Each fusion sequence was trimmed manually, based on the start and end positions defined by NCBI CD search annotations. An example is given in Fig. 20, of a fusion between $SdhC_A$ and $SdhD_A$ subunits that have well-defined CD search domains for each part .

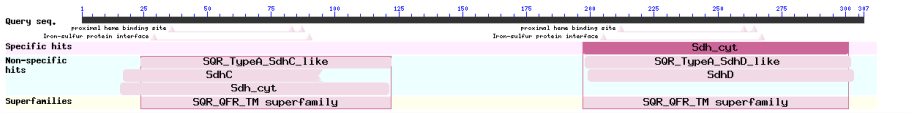


Fig. 20. A representative NCBI CD Search result with "Standard" output of fusion SDH sequences. *The accession of sequence used for this visualization is OGO49523.1.*

However, for the large $SdhC_B$ subunits, hypothesized to have resulted as fusions of ancestral SdhC and SdhD with a subsequent loss of the SdhC third helix (Hägerhäll 1997), it was not possible to use CD Search results as a guide for splitting, since the CD Search annotation for these sequences is a singular $SdhC_B$ domain (Fig. 21). Therefore, sequences were split before the third helix using the prediction of transmembrane helices (see Methods).

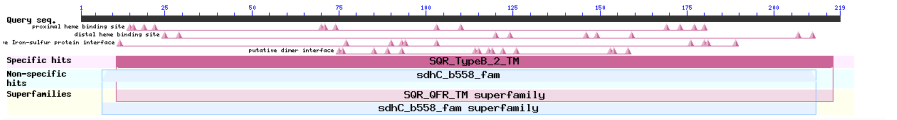


Fig. 21. A representative NCBI CD Search result with "Standard" output of SdhC$_B$ sequences. *The accession of sequence used for this visualization is KHD87266.1.*

At this stage, phylogenetic reconstructions of the resulting multiple sequence alignments per cluster and per subunit were produced and analyzed. Clades with an ultrafast bootstrap support equal or higher than 95 are statistically significant (Hoang *et al.* 2018), and only these were in here considered as highly supported.

3.6.2. Phylogenetic reconstruction of cytoplasmic subunits SdhA and SdhB

In the joined maximum likelihood phylogenetic reconstruction of SdhA sequences (Fig. 22A), sequences of types C, D, and F form well-supported monophyletic clades while sequences of types A and E/E* are intercalated. Non-monophyletic clades of type A also suggest the presence of several subgroups for this type (see below). Some clade of types A and E* contain sequences that were annotated as SdhA and had a corresponding SdhB subunit in synteny, but no respective membrane subunits were found. By global identity relationships they belong to the type of the sequences in the same clade (either, type A or E). However, due to the lack of membrane subunits and the fact that some of these sequences are from complete genomes, it was not possible to assign an SDH type for these sequences (See "No type" subsection of Section 3.5). It is worth noting a subclade in the clade of type E* contains NADH-dependent fumarate reductases from Aquificae (Miura *et al*. 2008; see "Type E*" subsection of Section 3.5). TFR sequences are found in three non-monophyletic clades. While most of the identified TFR complexes constitute a clade marked "TFR", TFR sequences from *Candidatus* Bathyarchaeota were found within a clade of type A (marked with an asterisk), and some TFRs from unclassified Euryarchaeota are found near the type D clade. Type F SdhA proteins are grouped in a highly supported monophyletic clade branching close to TfrA proteins. In addition, a clade containing type A protein belonging to Proteobacteria is stemming out of the largest TfrA clade. Two large red clades near the largest TFR clade contain either NadB sequences or sequences annotated as SdhA, for which no other SDH/FRD subunits were found. These sequences are most likely closely related to SdhA. To our knowledge, none of these proteins, mainly present within Actinobacteria and Alphaproteobacteria, are so far characterized. In the SdhA phylogeny, type B sequences are monophyletically organized, separated by the root, and form at least three distinct groups. This separation of type B into three groups is in agreement with what was previously observed (Butler *et al*. 2006; Kurokawa, Sakamoto 2005; Lemos *et al*. 2002). Starting from the root, two groups are likely to contain sequences from SDHs or bifunctional complexes (Subgroup II and III in Supplementary Table 6.7), while the third group contains FRDs or bifunctional complexes (Subgroup I in Supplementary Table 6.7). The distinction of functional clades is supported by CD Search annotation (sequences from the clade separated by the root are annotated as "Fumarate reductases" while other sequences are annotated as "Succinate dehydrogenases/fumarate reductases"), and by experimental characterization of

complexes from this clade (Mattson *et al*. 2000; Baughn, Malamy 2003; Garcia *et al*. 2007; Kurokawa, Sakamoto 2005; Butler *et al*. 2006; Lancaster *et al*. 1999; Lancaster 2001; Mileni *et al.* 2006; Weingarten *et al*. 2009). Additionally, one of the clades of type B contains some SdhA sequences from Clostridia, Coriobacteriia, and Methanobacteria that have a corresponding SdhB subunit but no membrane subunits. Some of the sequences were identified in complete genomes, therefore type B could not be assigned with certainty. Possibly, this clade contains complexes, the evolution of which involved the recruitment of a common domain to perform a different or similar function. Further taxonomic information regarding these three subgroups of type B is given in Supplementary Table 6.7. Interestingly, type B is not only placed as a basal type by MAD rooting in the phylogeny of both SdhA and SdhB, but also it has the lowest mean/median global identity to other types (Supplementary Table 6.8). A clade of type B (Subgroup III) additionally has its cytoplasmic subunits in separate MCL clusters, in contrast to other SdhAs and SdhBs being in the same cluster.

Although the phylogeny of only cytoplasmic 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 cytoplasmic subunits. The same is observed in the joined SdhB phylogenetic reconstruction (Fig. 22B). This is also supported by the mean and median global identities between SdhA subunits of the different types (Supplementary Table 6.8) where it can be seen that $SdhA_C$ is more similar to $SdhA_A$ than to $SdhA_D$ further supporting the relatedness of the complexes belonging to type A and type C. Although type D is equally similar by global identity to type C and A, it forms highly supported clades within SdhA and SdhB phylogenies. In addition, type C clade can be split into two subgroups, one containing predominantly sequences from Alphaproteobacteria and Gammaproteobacteria, while the other mostly includes Betaproteobacterial and Gammaproteobacterial sequences. $SdhA_A$ and $SdhA_E$ are also similar (42.9% mean global identity).

The overall topology of the joined maximum likelihood phylogenetic reconstruction of SdhB (Fig. 22B) is similar to the one found in SdhA. However, in this phylogenetic reconstruction, type B subclades are not separated by the root (but still form three subgroups). Moreover, in comparison to the SdhA phylogeny, the $SdhB_{A,E}$ sequences are organized in more distinct clades with a separation between types E and E*. A possible explanation relies on the fact that the $SdhB_{A,E}$ sequences (mean global identity is 26.9%, Supplementary Table 6.8) are more dissimilar than the corresponding SdhA sequences (mean global identity is 42.9%, Supplementary Table 6.8) since according to X-Ray structures, SdhB subunit is in contact with the membrane anchor subunits (receiving electrons from the hemes, or directly from the quinone in the cases of types D, E, and F), which differ significantly between these types, due to the amphipathic nature of SdhE and SdhF anchors and the transmembrane nature of the remaining SDH types. For both SdhA and SdhB phylogenies, the root placement and the overall tree topology remained unchanged in all cases of redoing the phylogenetic reconstruction with different parameters (various clade removal and different cluster reduction criteria as described in Methods).

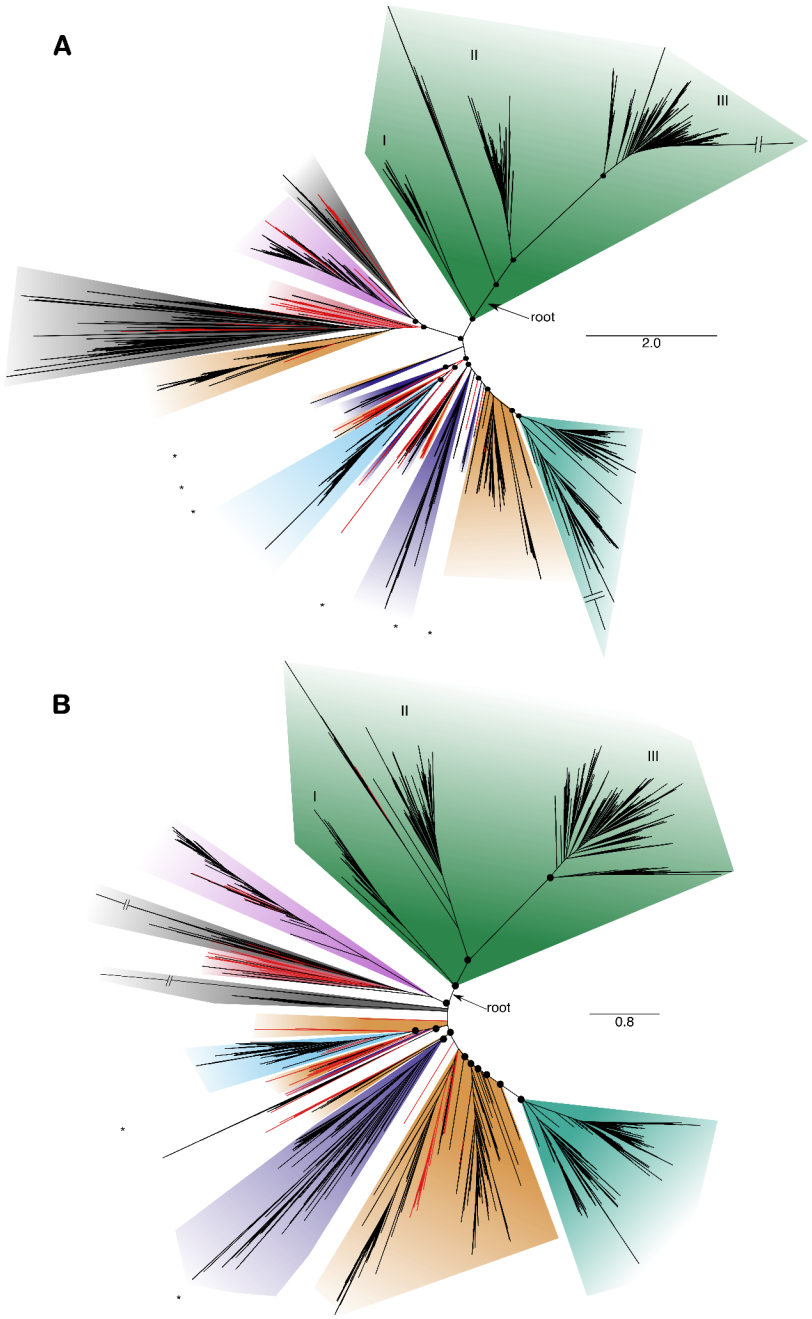


Fig. 22. Maximum likelihood phylogenetic reconstruction of SdhA (A) and SdhB (B) subunit sequences *(LG amino acid substitution matrix with four discrete gamma categories model; LG+G4). Radial view, colored by SDH type. Type A is colored in orange, type B in dark green, type C in dark cyan, type D in light blue, type E/E* in dark blue, type F in purple, TFR sequences are colored in pink while gray clades indicate clades of undetermined type or homologous complexes (see Methods and Section 3.3 of Results). The long interrupted type B branch belongs to a very short partial sequence (51 amino acids) belonging to the metagenomic assembly of unclassified Deltaproteobacteria. The long interrupted branch contains two sequences of MvhD from Candidatus Bathyarchaeota and Chloroflexi that were present in the largest SdhB cluster. Clades with Archaeal sequences are colored in red. Significant ultrafast bootstrap values (>= 95) are marked with a black dot. Subgroups of type B are number I, II, and III. Type E* clades are marked with an asterisk to differentiate from canonical type E enzymes. A clade of type A marked with an asterisk contains a TFR subclade of Candidatus Bathyarchaeota proteins.*

3.6.3. Transmembrane subunits SdhC and SdhD

Due to the low sequence conservation between the membrane subunits of the different types as well as their short sequence length (~100 amino acids after trimming), the phylogeny of membrane subunits was inconclusive (Data not shown). Instead, these proteins were analyzed in terms of similarity networks to evaluate how similar structural types are in relation to each other. Networks have the advantage of allowing several levels of annotations for each sequence (node), such as SDH type (Fig. 24 and Fig. 25), MCL cluster (Supplementary Fig. 7.7), CD search annotation (Supplementary Fig. 7.8), number of heme-binding histidines (Supplementary Fig. 7.9), sequence length (Supplementary Fig. 7.10), and number of predicted transmembrane helices (Supplementary Fig. 7.11), but have the disadvantage of losing the time connection (Bapeste *et al*. 2013). In the networks colored by structural type annotation, it can be seen that both membrane subunits from types A and B are each separated in at least three subclusters. Type B subgroups match those described for cytoplasmic subunits in Section 3.7.2. Interestingly, the Chloroflexi fusion membrane anchors, in here annotated as “Type B?” and colored in dark green, form their own cluster. This could be due to their fused nature, longer sequence length, and the number of predicted helices (7 vs 5 helices in types B and F, and three helices in other types). By the global identity analysis of soluble subunits, these fusions are closest to the type A complex.

As in type B, type A membrane anchors form three or four subgroups, following the overall clade organization found in phylogenetic reconstructions of SdhA/FrdA and SdhB/FrdB. However, it is noticed that Thermoplasmata, Thermoprotei, Archaeoglobi, and Thaumarchaeota tend to group together for all subunits, which could indicate that these sequences constitute a distinct subgroup, similar to what was observed in Lemos *et al*. 2002 phylogeny. The remaining archaeal proteins, belonging to Heimdallarchaeota and Halobacteria, are grouped within bacterial enzymes both in the phylogenetic reconstructions as well as in the network analyses. This could be an evidence for recent lateral gene transfer events within these lineages as reported for Halobacteria (Nelson-Sathi *et al*. 2012; Nelson-Sathi *et al*. 2015). Further analysis, including experimental characterization of these enzymes, is necessary to resolve the potential type A subgroups.

Moreover, while membrane anchors of types D and F form single clusters in both networks, type C sequences separate into two tightly connected subclusters, one composed mostly of Alphaproteobacteria and Gammaproteobacteria, and the other mostly of Betaproteobacteria and Gammaproteobacteria. This is in agreement with the two clades observed in the phylogenetic reconstructions of the cytoplasmic subunits (Fig. 22).

Overall, this separation of the membrane sequences into subgroups correlates with the clades shown in the joined phylogenetic reconstructions of both cytoplasmic subunits (see Section 3.7.2.), and can be further observed in the similarity networks annotated by the MCL cluster.

Analysis of the number of hemes per type is shown in Supplementary Fig. 7.9, where nodes are colored according to the number of heme-binding histidines. Both types A and B had four heme-binding histidines (in both $SdhC_A$ and $SdhD_A$), types C and D had two (one in SdhC and one in SdhD but not in all organisms due to the presence of partial metagenomic sequences), which is consistent with previous literature reports (Hägerhäll, Hederstedt 1996; Hägerhäll 1997; Cecchini *et al*. 2002). The similarity network annotation for $SdhC_F$ histidine content is set at zero because type F was excluded from multiple sequence alignment that was used to check histidine conservation since sequences of this type do not show homology to other canonical types.

The CD search annotation is fairly accurate at type prediction for SdhC sequences, and less accurate for SdhD sequences. Specifically, CD search annotated type D and type B subgroup I as fumarate reductases, and other clusters a succinate dehydrogenases. Altogether, type B CD search annotations split this structural type into 3 subgroups: a fumarate reductase (cluster 40), and two subgroups of succinate dehydrogenases (clusters 22+23 and a cluster 34). This supports the previously discussed clade separation of soluble subunit sequences of this type in the phylogenetic reconstructions (Section 3.7.2.).

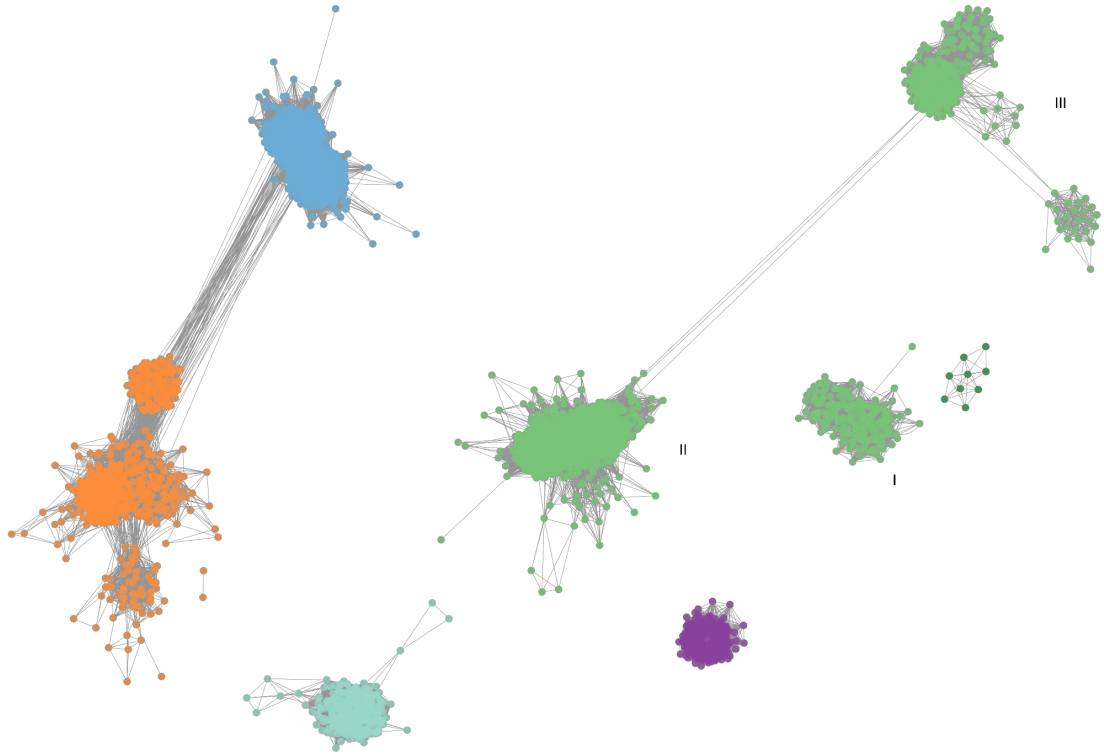


Fig. 24. Similarity network of SdhC sequences colored by structural type*: type A is in orange, type B is in green, type C is in blue, type D is in cyan, type F is in purple. Additionally, Chloroflexi fusion subunits were colored as type "B?" (dark green). Type B subgroups are marked with I, II, and III.*

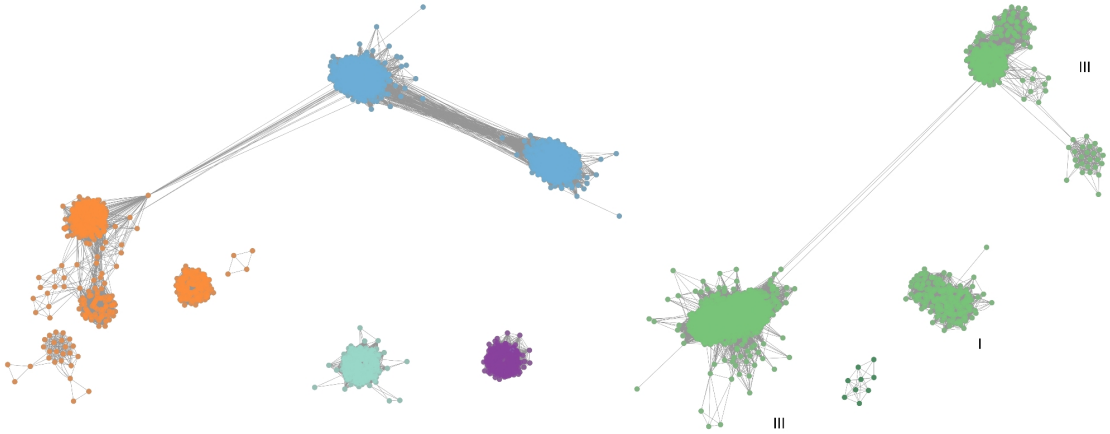


***Fig. 25. Similarity network of SdhD sequences colored by structural type**, as in Fig. 24.*

3.6.4. Membrane subunits (SdhE and SdhF)

The joined maximum likelihood phylogenetic reconstruction of SdhE and HdrB sequences together with the CCG domain parts of the homologous TfrB sequences (Schäfer *et al*. 2002) is shown in Fig. 26. It is worth noting that many clades of the phylogenetic reconstruction remained poorly resolved (with low bootstrap support), which makes the analysis of certain clades inconclusive, namely the HdrB clades (which are not a direct focus of this thesis), and the TfrB clade. The overall poor statistical support for the phylogenetic reconstruction of SdhE, HdrB, and TfrB (CCG domain) sequences was observed in the unreduced phylogenetic reconstruction of the SdhE/HdrB cluster as well. Although, while it appears that TfrB sequences are closer to the HdrB than SdhE, due to insufficient statistical support of this clade, further analysis is necessary to confirm whether or not this was the case. This potential proximity between TfrB and HdrB proteins would contradict the current view of TfrB evolution as a result of the fusion of $SdhB_E$ and SdhE (Schäfer *et al*. 2002). The same low support is also observed for the type E/E* clade, with an ultrafast bootstrap support of only 78. However, the separation between canonical type E clade and the type E* clades is statistically significant.The yellow HdrB clade within the type E/E* clades includes unclassified Euryarchaeota and Candidatus Poseidoniia sequences, whose genomes contain an additional complete SDH complex of type F and no additional cytoplasmic subunits or HdrA/C sequences identified. Therefore, it was not possible to conclude if these sequences are SdhEs or HdrBs, but rather they represent close homologs that *in vivo* fulfill some other function. Canonical type E and type E* (which includes the MFR complexes) form one clade and therefore there is no need to reclassify them into different types. It is however useful to use the “Type E vs. Type E*” distinction to indicate the lack of SdhF subunit in Type E* complexes. This phylogenetic reconstruction was helpful to differentiate between cyanobacterial HdrBs and SdhEs, as they are found in separate clades in the phylogenetic reconstruction, the HdrB clade and the SdhE clade, respectively.

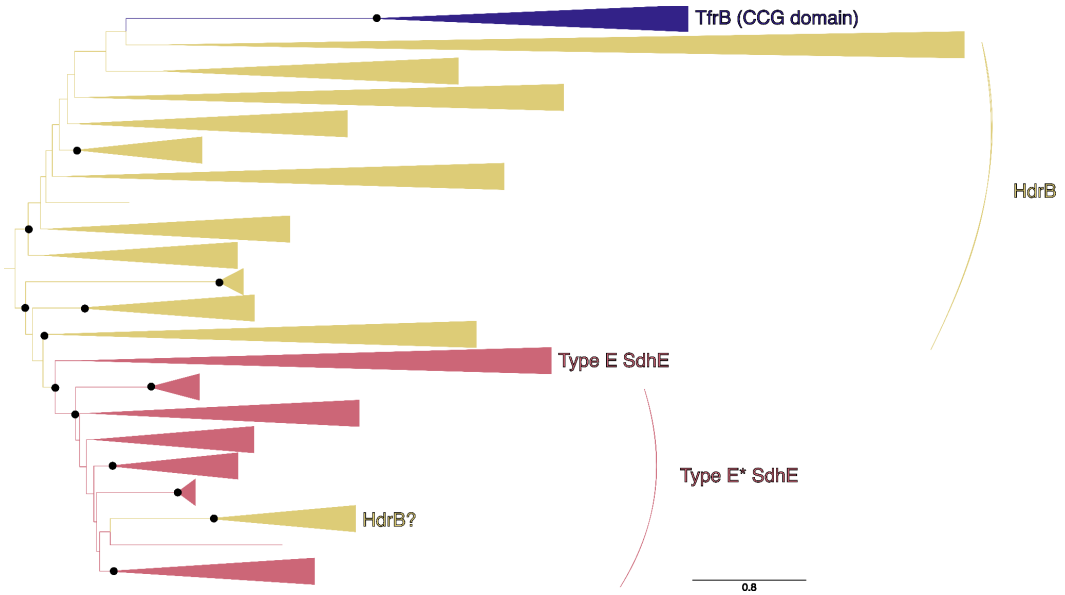


Fig. 26. Joined maximum likelihood phylogenetic reconstruction of SdhE and HdrB sequences as well as CCG domain part of TfrB sequences *(LG amino acid substitution matrix with four discrete gamma categories model; LG+G4). Rectangular view. HdrB clades are highlighted in yellow, type E/E* clades are in pink, and the CCG domain of TfrB clades are in dark blue. Significant ultrafast bootstrap values (>= 95) are marked with a black dot.*

Fig. 27 shows the maximum likelihood phylogenetic reconstruction of the SdhF subunit. This subunit is only present in the canonical type E complexes and with limited taxonomic distribution, being found only in some organisms of Thermoprotei (Sulfolobus, Sulfurisphaera, Sulfodiicoccus, Acidianus, Metallosphaera, *Candidatus* Aramenus) and Thermoplasmata (Picrophilus, Acidiplasma, Ferroplasma).

Both archaeal classes form separate clades on the phylogenetic reconstruction. Due to the scarcity of information about this subunit, it could only be identified through identity relationships to the known SdhF subunit sequences, clustering, and synteny analysis.

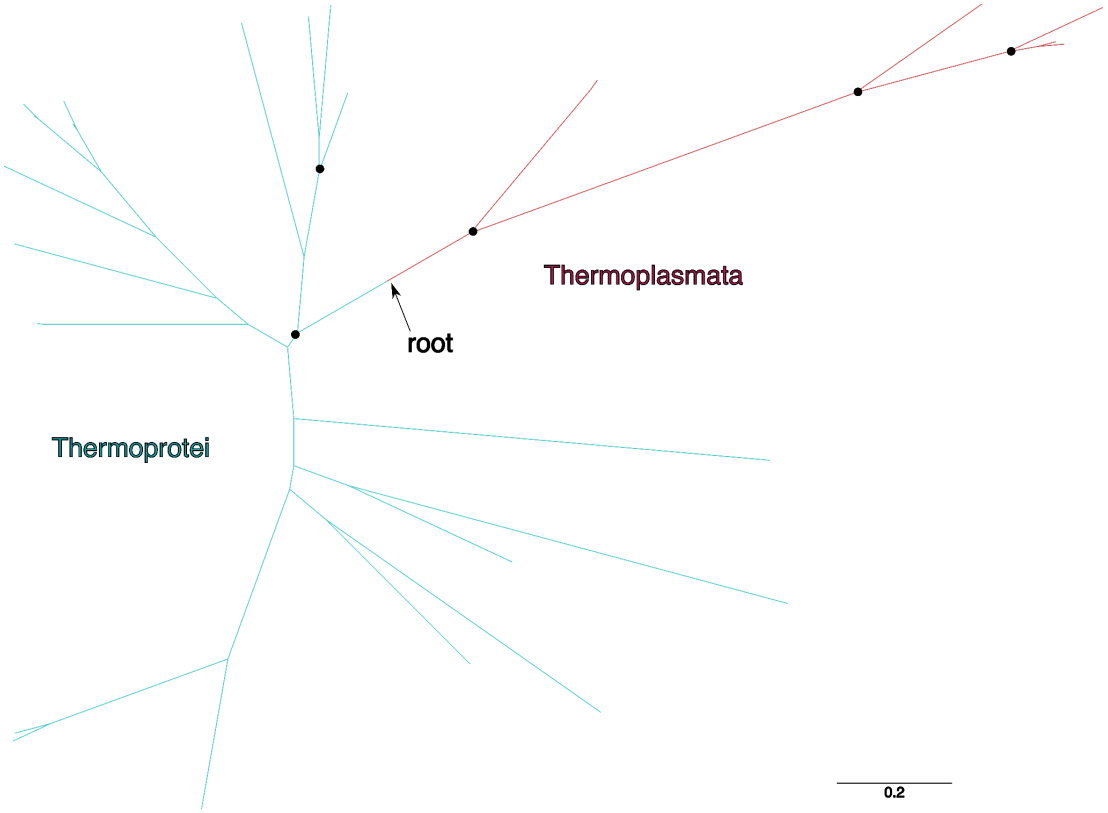


Fig. 27. Maximum likelihood phylogenetic reconstruction of SdhF subunit *(LG amino acid substitution matrix with invariable sites and four discrete gamma categories model; LG+I+G4). Radial view. Colored by taxonomy: Thermoprotei is in cyan, Thermoplasmata is in red. Significant ultrafast bootstrap values (>= 95) are marked with a black dot.*

Supplementary File 6.9 contains phylogenetic reconstructions in Newick format, per cluster as well as joined phylogenies for subunits SdhA, SdhB, and SdhE/HdrB/TfrB$_{CCG}$.

4. Discussion

Succinate:quinone oxidoreductases (including succinate dehydrogenases and fumarate reductases) are widespread among the three domains of life, and their function is highly conserved (Hederstedt, Rutberg 1981; Lemos *et al.* 2002; Jardim-Messeder *et al.* 2017). Moreover, these complexes have many homologous proteins with diverse functions (Mattevi *et al.* 1999; Cole *et al.* 1988; Jardim-Messeder *et al*. 2017; Heim *et al*. 1998; Pellicer *et al*. 1996). Such properties made complex II an attractive target for studying energy metabolism.

In this analysis, a large-scale comparative genomics analysis (~35000 genomes) of this enzyme was performed, which included the taxonomic distribution of the structural SDH/FRD types, as well as homology analysis and phylogenetic reconstruction of cytoplasmic subunits and similarity network analysis of membrane anchor subunits. Homology search helped identify new homologous proteins of the flavoprotein subunit, therefore the “Fumarate reductase superfamily” (proposed by Jardim-Messeder *et al*. 2017) could potentially be expanded.

Overall, the taxonomic distribution for each type shows that type A is the most diverse across both prokaryotic domains, while type B is the most diverse among Bacteria. Type C is restricted to Proteobacteria (according to the data analyzed in this thesis). Type F is not exclusive to Mycobacteria and not even to Actinobacteria, and more complexes of this type need to be characterized. We have found that thiol:fumarate reductases are also not exclusive to methanogens, and this complex also is in need for further biochemical characterization. It is also worth noting that since it was observed that the structural type could vary on a strain level, as it is in the case of *Rhodothermus marinus*, the possibility of the presence of additional types that were not reported in this analysis in the same lineage (due to absence of genomes) cannot be excluded. “No type” sequences may be either close homologous proteins of SDH or some type of soluble SDH that has not yet been characterized.

The syntenic arrangements of these complexes showed the absence of conservation in the neighbor sequences, except for very closely related species with similar metabolism (*e.g.* Enterobacteriaceae genomes have the same neighbors around SDH/FRD complexes).

There is a distinction in phylogenetic reconstruction between the $SdhA_B$ clades and the $SdhB_B$ clades, namely that while clades of this type are basal in both cases, in the reconstruction of SdhA these clades are separated by the root. One of the possible reasons for the root separation could have been a presence of very short partial metagenomic sequences from Deltaproteobacteria in the SdhA reconstruction, which might have skewed the rooting performed by MAD algorithm, since MAD calculates the rooting position by “(a) considering each branch separately as possible root position; (b) deriving the induced ancestor-descendant relationships of all the nodes in the phylogenetic reconstruction; (c) calculating the mean relative deviation from the molecular clock expectation that is associated with the root position on the branch. The branch that minimizes the relative deviations is the best candidate to contain the root node.” (Direct quote from Tria *et al*. 2017). However, this possibility was tested by removing this short sequence and redoing the phylogenetic reconstruction, which resulted in the same root position as before. Another possibility of such difference in the root positions between the soluble subunits of type B is the function these subunits perform in the complex. The function of the SdhB subunit is to transfer electrons, regardless of the direction of the reaction being catalyzed by the SdhA subunit. Therefore, it is possible that $SdhB_B$ sequences are more similar to each other than SdhA sequences of the same type, since some of them have evolved to become strict fumarate reductases, strict succinate dehydrogenases, or bifunctional enzymes. The $SdhA_B$ clade that is separated by the root from the others is likely to contain fumarate reductases, based on the characterized enzymes from this clade (*e.g. Wolinella succinogenes* FRD as described by Lancaster *et al*. 1999; Butler *et al.* 2006; Kurokawa, Sakamoto 2005). However, more enzymes need to be characterized biochemically to test this hypothesis.

Overall, some of the clades in the SdhB phylogeny received a lower statistical support than the clades of SdhA. This could be attributed to the changes in SdhB subunit that occurred to accommodate the modified (type D) or completely different membrane anchor (types E, E*, and F).

Type E/E* is most similar to type A, despite the significantly different membrane anchors. $SdhA_{E*}$ are intercalated while $SdhB_{E*}$ forms a single clade which could be hinting at differences in function for subclades of type E* (*e.g.* MFR vs SDH-functioning enzyme). It was not possible to differentiate between SDH/FRD complexes of type E* and MFR complexes in this analysis, as they have the exact same profile. Due to little information about MFR available, MFR was kept under the same type annotation as type E* complexes, despite being likely a homolog of SDH and not a true SDH (Juhnke *et al*. 2009; Weingarten *et al*. 2009). Interestingly, in the SdhA phylogeny, one type E* clade contains two subclades (one containing Aquificae sequences and the other containing Epsilonproteobacteria sequences, respectively) where, although by global identity analysis these sequences were classified as type E*, the SdhE subunit was not identified. For most SdhA sequences of the Epsilonproteobacteria clade, the respective SdhB subunit was not identified as well, which suggests that this subclade contains sequences of an enzyme homologous to a type E* SDH but not a true SDH. For Aquificae sequences, most of them had no KO annotation, or two KOs annotated with significance: one KO for the respective SDH subunit, and the other for NADH-dependent fumarate reductase. The latter was characterized in Aquificae (Muira *et al*. 2008), where the authors showed that subunits A and B of this enzyme are homologous to SdhA and SdhB, respectively. It is likely that this subclade contains the NADH-dependent fumarate reductase due to the presence of additional subunits from this enzyme in the respective genomes. Interestingly, this enzyme was detected only in a few genera of Aquificae, while in other Aquificae-affiliated genomes a complex of type E* was detected. Cases of Aquificae and Epsilonproteobacteria clades further support the hypothesis about the presence of several SDH subtypes or their close homologs in type E*.

The current evolutionary assumption about type B complexes is that it resulted through the fusion of type A membrane anchors with a loss of one transmembrane helix (Hederstedt, Hägerhäll 2996; Hägerhäll 1997; Lancaster 2013). However, the research presented in this thesis does not allow fully to corroborate this hypothesis. Not only type B sequences are placed at the root in both phylogenetic reconstructions for SdhA and SdhB, but they also share very low identity with type A for even soluble subunits. On the contrary, according to the phylogenetic reconstruction presented in this analysis, type A and type B soluble subunits resulted from an ancient duplication with a later attachment to the membrane (see below). Additional tests of redoing the phylogenetic reconstructions with different parameters, which have resulted in no changes in overall topology and root position for both SdhA and SdhB phylogenies, provide additional support for this topology.

Also, type B seems to potentially contain different subtypes which may vary functionally. This question will not be resolved by comparative genomics, it is urgently necessary to characterize more enzymes of this type, as well as other types. Taxonomically diverse biochemical and structural data for complex II is currently lacking.

Overall, the structural classification and phylogeny of SDH sequences are not an absolute match. Structural types A and B may contain subgroups/subtypes while type E/E* is understudied and is currently split into canonical type E, MFR, type E in Cyanobacteria, and so on. In this thesis, the canonical type E was kept as it was

originally classified (Lemos *et al*. 2001) and all other non-canonical cases were grouped under the name of “type E*”. Both types form a single monophyletic clade, but the canonical type has an additional membrane subunit SdhF, therefore the distinction was drawn. Other types (C, D, F) form monophyletic clades, and therefore there is no need for their re-classification.

4.1. Evolutionary scenario

Evolution of succinate dehydrogenases and their closest homologs (fumarate reductases) has implications for the evolution of prokaryotic diversity in terms of both energy and carbon metabolisms. Moreover, it cannot be dissociated from the evolution of the modular blocks that form each of its subunits and their respective cofactor content (Baymann *et al*. 2002). Based on the modular nature of the different types alone, a parsimony explanation for the evolution of these complexes would be an early separation of types A, B, C and D from TFR and type E/E*, with type F being a more recent innovation. However, we have observed trends in the evolution and diversity of the single subunits which suggest a more intricate evolution, with multiple events of membrane anchor replacements. According to the phylogenetic reconstructions of subunits A and B, there is a clear separation between type B enzymes and all of the remaining enzymatic complexes in here addressed.

One possible interpretation would be the ancestry of type B over the remaining types (Supplementary table 6.8). In this category are included bifunctional enzymes able to catalyze both reactions, as well ones specialized in one of the two reactions. However, the scarce taxonomy distribution of B type enzymes within Archaea argues against this scenario. A more parsimonious solution would be that the primordial complex was a soluble version of the enzyme, composed only of the flavin and iron-sulfur center subunits, with a later attachment to the membrane. This modular block organization is observed in many other homologous complexes, and would not require the pre-existence of hemes within biological complexes. In the last years, laboratory experiments have achieved the synthesis of several intermediates of the TCA cycle, where fumarate and succinate were present (Varma *et al*. 2018), 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 (Wimmer *et al*. 2021), both fumarate and succinate are part of the metabolic network, so is the reaction nowadays catalysed by the soluble version of NADH-fumarate reductase. In fact, the calculated DG of the reaction under alkaline vent conditions (-65.4 kJ.mol-1) (Wimmer *et al*. 2021), favors the conversion of fumarate to succinate in the reductive direction.

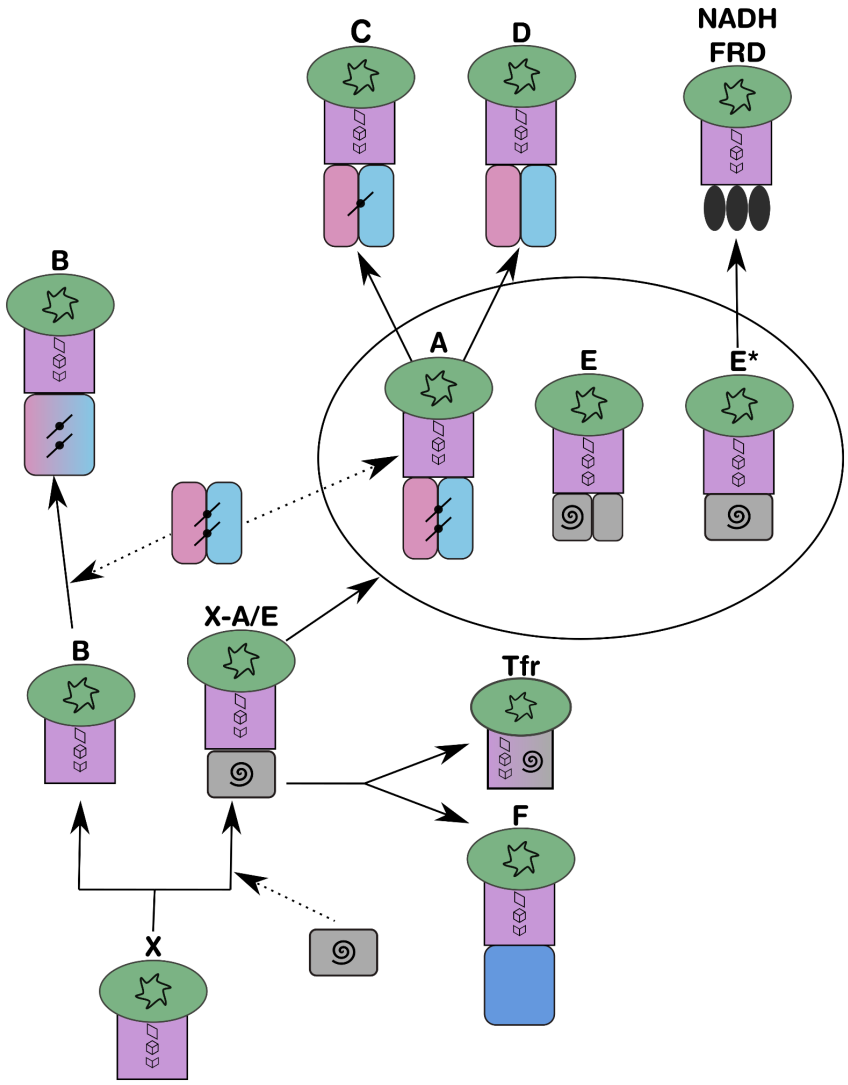


***Fig. 28. Schematic representation of potential evolutionary scenario of different SDH/FRD types**. For cofactor and subunit schematics of each type, please refer to Fig. 6.*

In this scenario, and considering the phylogenetic reconstructions topologies were also observed in other reconstructions (clades of types A and B in Lemos *et al*. 2002), type B SDH and FRD clade separation in (Kurokawa, Sakamoto 2005), also three subclades in (Butler *et al*. 2006), the first event would have been an ancient duplication of the FAD and iron-sulfur subunit, 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. 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 and in type A enzymes replaced by the transmembrane anchor. The membrane anchor recruited by type A enzymes would have been acquired within bacteria by type B enzymes as well. 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 (Kummerfeld, Teichmann 2005), and type B taxonomic distribution is narrowed than that of type A, Occam's razor favors the former over the latter hypothesis.

In Archaea, type E would have recruited the subunit F. 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 quinones (Hards *et al*. 2019). Type C and type D would have independently evolved from type A enzymes by losing one or two hemes respectively. 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 (Miura *et al*. 2008). 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 (Thauer *et al*. 1977), specialized SDHs only evolved at a later time, with the increase in oxygenation levels of the atmosphere (Holland 2006). 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 1996 paper, 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.

5. Conclusions

Phylogenetic analysis of cytoplasmic subunits of different types has shown that some structural types may be split into subtypes taxonomically (type E/E*, type A, type B) while others are monophyletic and possibly should not be separated into subtypes (types C, D, and F). This separation for type B has been observed in previous studies and may be due to the SDH/FRD function of the enzyme. In order to confirm that, more enzymes of this type have to be biochemically characterized.

Some types are taxonomically diverse while others are confined to a single domain (type B occurs only in Bacteria with very few metagenomic exceptions) or even a single phylum (type C is exclusive to Proteobacteria). Canonical SdhF-containing type E was found only in Thermoprotei and Thermoplasmata.

Unfortunately, it was not possible to perform a phylogenetic analysis of the membrane subunits of canonical SDH types (A, B, C, and D) due to the low global identity of the sequences of different types, therefore this question remains open for the future.

The functional annotations of SdhE sequences hint at a lack of models for these sequences since they were annotated as HdrBs or had no annotation by all functional annotation tools used in this project. In addition, the mapping of the different variations of the TCA cycle with the different types of SDH/FRDs could potentially aid in the functional and evolutionary distinction of between the different types. This is not planned to be covered by this project but remains an open question that needs to be addressed.

Overall, comparative analysis of this superfamily across 35000 genomic assemblies expanded the current taxonomic distribution of several types of the complex II in prokaryotes and allowed the potential identification of novel subtypes to be further experimentally characterized. Combined analysis of the phylogenetic reconstruction and similarity networks allowed the elaboration of a scenario, in which a primordial module, composed of the common cytoplasmatic subunits, undergone several independent events of membrane attachments, replacements, fusions and environmental adaptations to give raise to the current taxonomic distribution.

6. Supplementary Information Tables

Supplementary Table 6.1. Complex II and homologous queries

Characterized query complexes are not colored while complexes predicted from genomes are colored red. In SDH Type column, the color yellow indicates those complexes whose type was identified during this project. Blue color of some accessions shows that these accessions were retrieved from feature table based on their syntenic proximity to other query subunits of the same complex. For *Sulfolobus acidocaldarius*, SdhB and SdhE sequences were not found in the genome that the project dataset contains, SdhB is labeled as a pseudogene, and SdhE is possibly one of the pseudogenes labeled as “hypothetical protein”. The table is found as a supplementary file (https://ucloud.univie.ac.at/index.php/s/B7JWn7OsTifxQqz)

Supplementary Table 6.2. Genome dataset

Detailed table of metagenomic assemblies dataset is given as a supplementary file (https://ucloud.univie.ac.at/index.php/s/UYf6wesyOujfWLc)

Supplementary Table 6.3. Detailed MCL cluster functional annotation

Detailed cluster annotation table is given as a supplementary file (https://ucloud.univie.ac.at/index.php/s/TEi9L37o8EvGMrL)

Supplementary Table 6.4. Taxonomic distribution of SDH types matrix

The matrix with the values presented in the heatmap (Section 3.5) is given as a supplementary file (https://ucloud.univie.ac.at/index.php/s/EMH8VjrygSlcloy)

Supplementary Table 6.5. SDH types per each genome with their respective taxonomy

The table of genomes with their affiliated taxonomy and the types of SDH complexes found within each of them is given as a supplementary file (https://ucloud.univie.ac.at/index.php/s/GvX7TOVfWvxRvUM)

Supplementary Table 6.6. Cluster redundancy reduction

cluster	# sequences before reduction	# sequences after reduction
1	20582	4793
3	14339	3605
8	7421	2233
12	5968	1520
15	3964	1449
16	3688	1275
17	3649	2018
18	3396	1821
20	3013	1818
22	2910	1501
23	2829	1467
24	2733	846
26	2514	747
29	2047	656
30	2028	449
34	1700	859
36	1208	454
37	1067	383

Only clusters with more than 1000 sequences were used.

Supplementary Table 6.7. Type B subgroup taxonomy

Subgroup	Taxonomy (most abundant taxa are underlined)	Notes
Subgroup 1 (SdhC cluster 40)	Acidithiobacillia, Alphaproteobacteria, Betaproteobacteria (Azonexaceae, Chromobacteriaceae, Comamonadaceae, Suterellaceae, Zoogloeaceae, unclassified Burkholderiales), Gammaproteobacteria (Alteromonadaceae, Cellvibrionaceae, Ectothiorhodospiraceae, Ferrimonadaceae, Halieaceae, Moritellaceae, Pectobacteriaceae, Saccharospirillaceae, Shewanellaceae, Sinobacteriaceae, Thiotrichaceae), Epsilonproteobacteria (Campylobacteraceae, Helicobacteraceae), Deltaproteobacteria (Desulfarculaceae, Desulfobacteraceae, Desulfohalobiaceae, Desulfovibrionaceae, Syntrophaceae, Syntrophobacteraceae), Hydrogenophilalia, *Candidatus* Lambdaproteobacteria, Bacilli (Alicyclobacillaceae, Paenibacillaceae), Clostridia (Peptococcaceae), Dehalococcoidia, *Candidatus* Thorarchaeota, Candidatus Schekmanbacteria, unclassified Spirochaetes, unclassified Nitrospinae, unclassified Parcubacteria group	Possible fumarate reductases or bifunctional enzymes
Subgroup 2 (SdhC cluster 34)	Bacilli (Bacillaceae, Alicyclobacillaceae, Paenibacillaceae, Sporolactobacillaceae, Planococcaceae, Staphylococcaceae, Thermoactinomycetaceae), Clostridia (Lachnospiraceae, Ruminococcaceae, Clostridiaceae, Peptococcaceae, Heliobacteriaceae, Thermoanaerobacteraceae, Symbiobacteriaceae), Negativicutes, Methanobacteria, Coriobacteriia, Deltaproteobacteria (Syntrophaceae, unclassified Deltaproteobacteria), *Candidatus* Eisenbacteria, Planctomycetes, Candidate Division NC10, Acidobacteria, Elusimicrobia, *Candidatus* Omnitrophica, Oligoflexia, Blastocatellia, Armatimonadetes, Fimbriimonadia, *Candidatus* Sumerlaeota, Chloroflexi, Chlamydiae, Nitrospinae, Actinobacteria, Gammaproteobacteria (Chromatiaceae, unclassified Gammaproteobacteria), Synergistia.	Possible succinate dehydrogenases or bifunctional enzymes
Subgroup 3 (SdhC cluster 22/23)	Actinobacteria, Oligoflexia, Planctomycetes, Alphaproteobacteria, Cyanobacteria, Gammaproteobacteria (Thiotrichaceae, Piscirickettsiaceae, Ectothiorhodospiraceae, unclassified Gammaproteobacteria), Deltaproteobacteria (Bradymonadaceae, Polyangiaceae, Desulfobacteraceae, Desulfobulbaceae, Syntrophaceae, Sandaracinaceae, Myxococcaceae, Kofleriaceae, Nannocystaceae, Geobacteraceae, Desulforomonadaceae, unclassified Deltaproteobacteria), Clostridia (Thermoanaerobacteraceae), Acidobacteria, *Candidatus* Omnitrophica, Elusimicrobia, Ignavibacteriae, Candidate Division Zixibacteria, Candidate Division GN15, Verrucomicrobia, Nitrospirae, Chrysiogenetes, *Candidatus* Tectomicrobia, Spirochaetes, Gemmatimonadetes, Lentisphaerae, Chloroflexi, *Candidatus* Rokubacteria, Chlorobi, *Candidatus* Kapabacteria, *Candidatus* Hydrogenedentes, *Candidatus* Handelsmanbacteria, *Candidatus* Marinamargulisbacteria, Armatimonadetes, Kiritimatiellae, Fibrobacteres, Abditibacteriota, Vicinamibacteria, Calditrichaeota, *Candidatus* Cloacimonetes, *Candidatus* Marinimicrobia, *Candidatus* Blackallbacteria, *Candidatus* Melainabacteria, Bacteroidetes, Rhodothermaeota, Balneolaeota, unclassified Epsilonproteobacteria, unclassified Euryarchaeota.	Possible succinate dehydrogenases

Supplementary Table 6.8. SdhA and SdhB intertype global identity

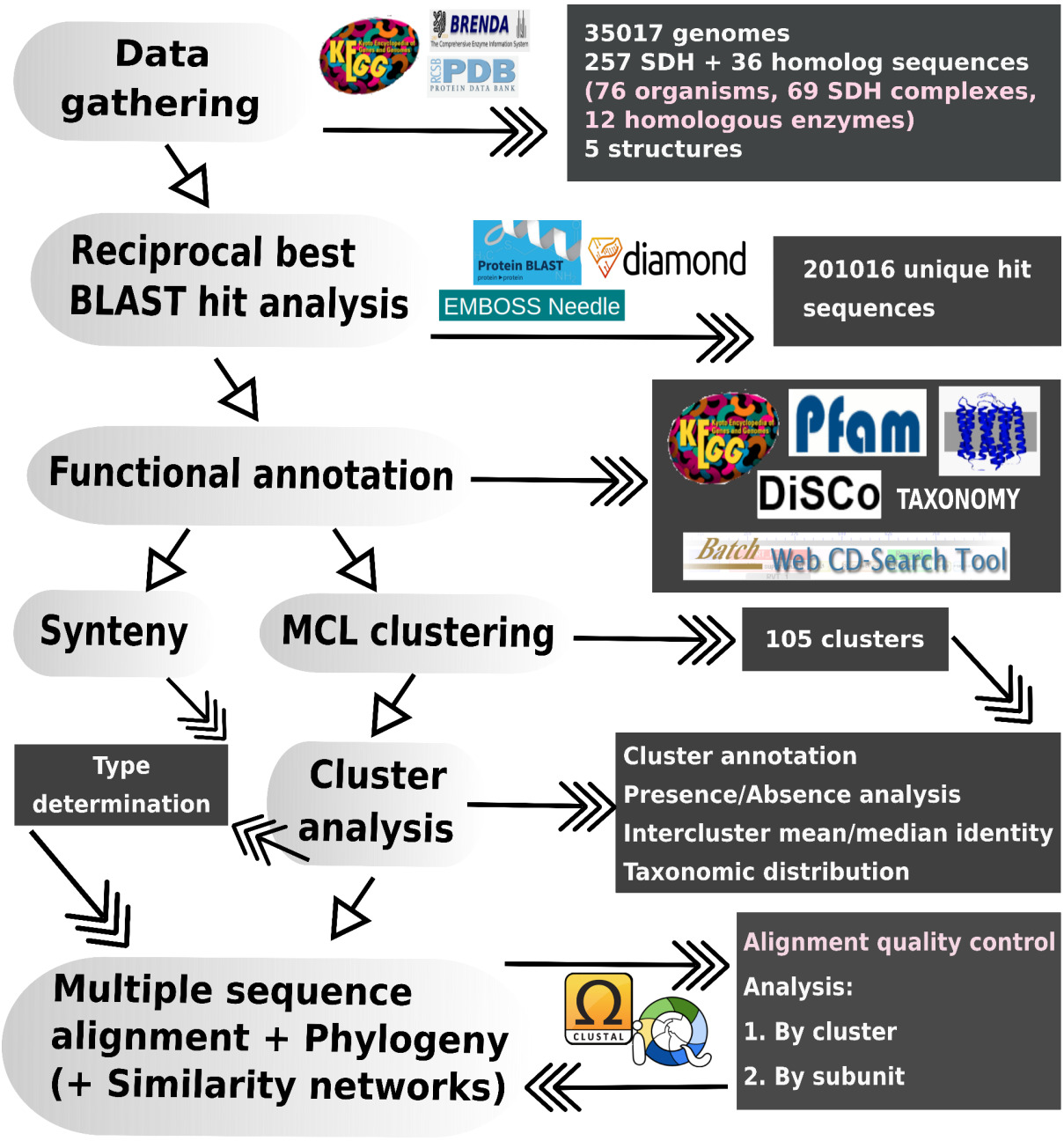
For intertype mean and median identity values for SdhA and SdhB subunits, see attached file (https://ucloud.univie.ac.at/index.php/s/iiBhn6B1AxADBja)

Supplementary File 6.9. Phylogenetic reconstructions in Newick format

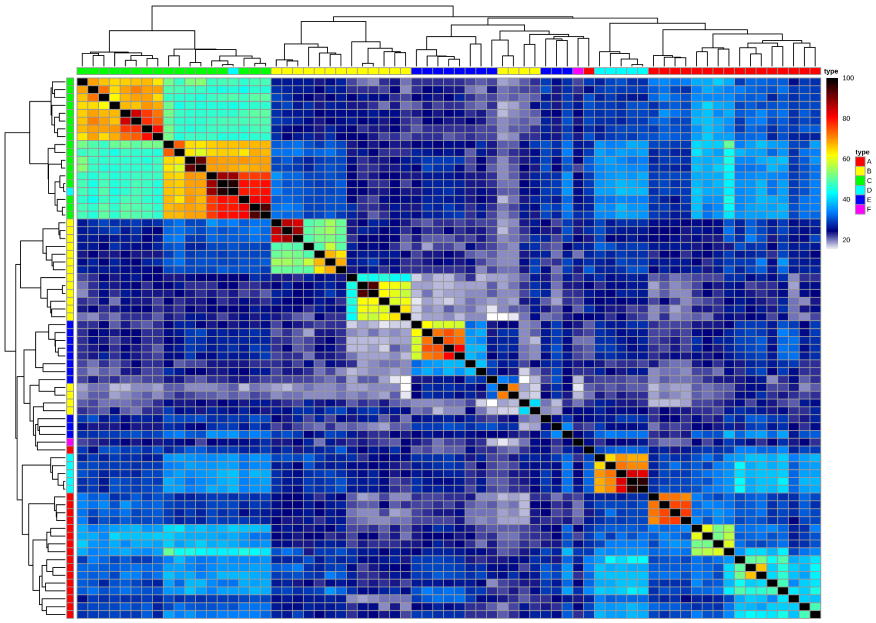
Supplementary file (https://ucloud.univie.ac.at/index.php/s/b68F4pdjbqHVLTL) contains phylogenetic reconstructions per cluster, as well as joined phylogenies for subunits A, B, and E. All phylogenies are given in Newick format, with their respective annotation tables in tabular format.

7. Supplementary Information Figures

Supplementary Fig. 7.1. Methodological Pipeline used in this analysis

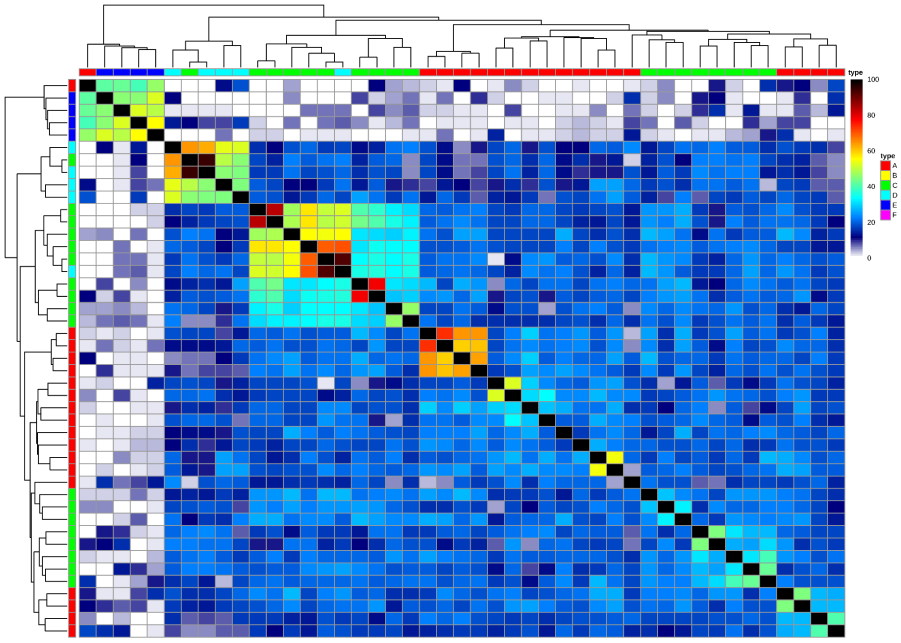


Supplementary Fig. 7.2. All vs all global identity of subunit B query sequences



Similarly to SdhA, there are two high identity clades within type C, one within type D, one within type E enzymes, and one high identity clade in type A. For type B, the identity seems to be lower, although not as low as for some SdhAs of type B.

Supplementary Fig. 7.3. All vs all global identity subunit D query sequences



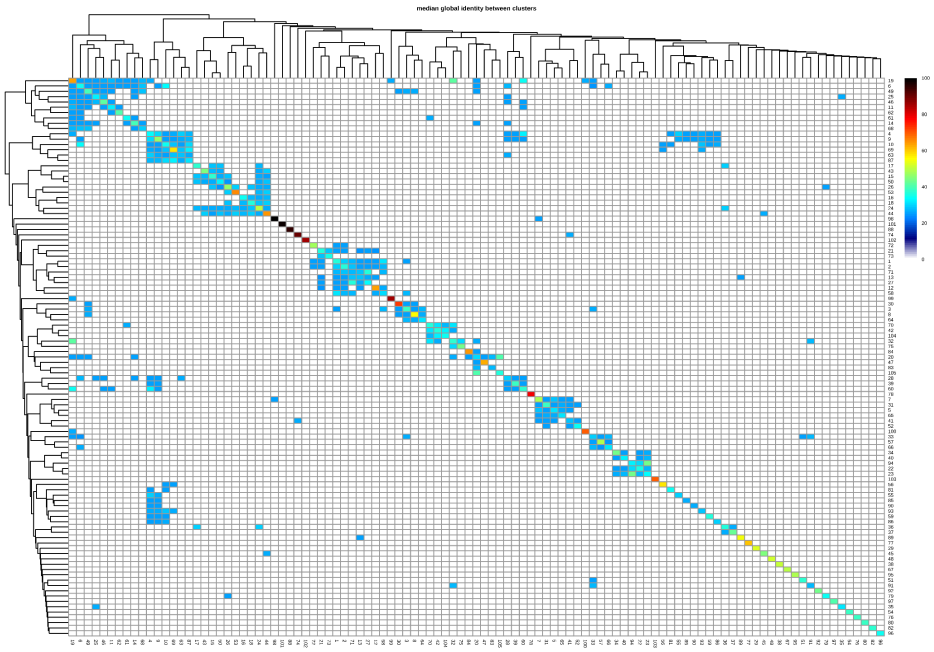
The sequences of this subunit have low homology to each other. The identity of SdhDs/SdhFs of type E is lower than that of SdhCs/SdhEs.

Supplementary Fig. 7.4. Taxonomic distribution of sequences per cluster

This is a preview of the full heatmap portraying the number of genomes that contain proteins from each cluster per class/order of organisms with their phylum in parentheses, sorted by the NCBI taxonomy The numbers in square brackets denote the number of genomes.

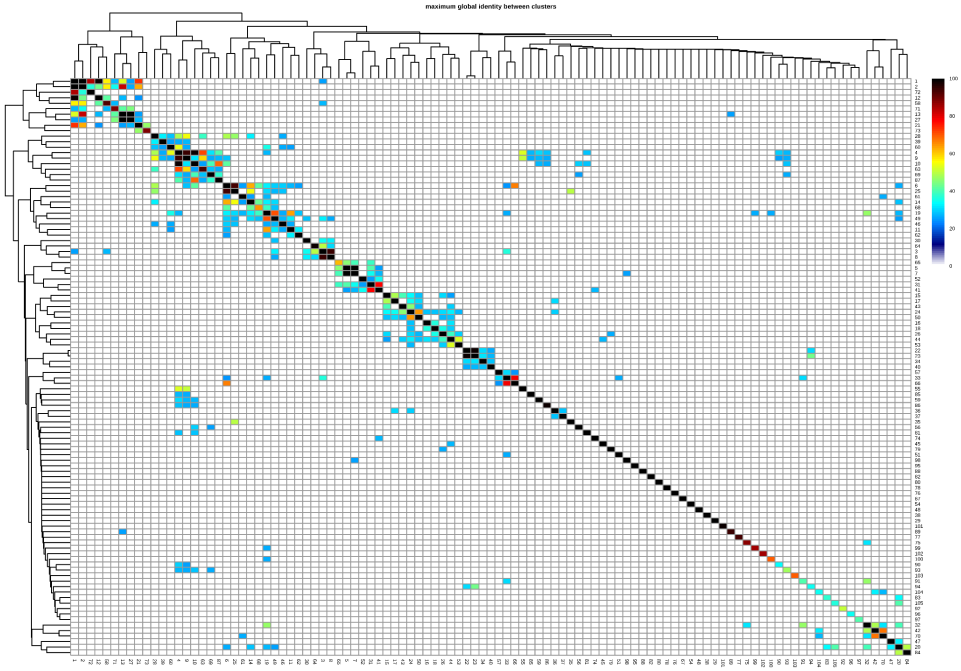
The order of the columns is per type of the most predominant protein in cluster: SdhA + TfrA, SdhA, FccA + other flavocytochromes c, SdhB, TfrB, SdhC sorted by type alphabetically (clusters of type A, type B, type C, type D, and type F), the cluster containing fusions of SdhC and SdhD (all are type A SDHs), SdhD sorted by type alphabetically (clusters of type A, type C, type D). Then it is SdhD/SdhF of type E, and SdhE/SdhC of type E together with HdrB in the same cluster. Following are all the clusters with non-SDH proteins: other HdrB clusters, HdrA, HdrC, and its fusions, HdrD, GlpC and their homologous FeS sequences and cluster with HdrE, AprA and AprB, QmoC, NadB, unidentified SdhA homologous proteins, GlcE, GlcF, and GlcD as well as fusions of GlcD and GlpC and GlpC separately, following is GlpC and QcrB fusions, and then GlpA and GlpD. The columns after Glp consist of LldE and LldF clusters, SoxB and ThiO clusters, DadA fused with GcvT as well as DadA separately, various FAD-linked oxidases, FAD-binding oxidoreductases, and FAD-binding proteins. The last columns include a variety of cytochrome c3 fusions, KstD, ThiG, CobZ/TcuA, and ferredoxins. The analysis of the taxonomic distribution of the homologous complexes, such as Hdr, Glc, and Glp lies beyond the scope of this thesis. Full resolution of the heatmap is given as a supplementary file (https://ucloud.univie.ac.at/index.php/s/A7tCsq6bXoajZsT)

Supplementary Fig. 7.5. Median intercluster global identity



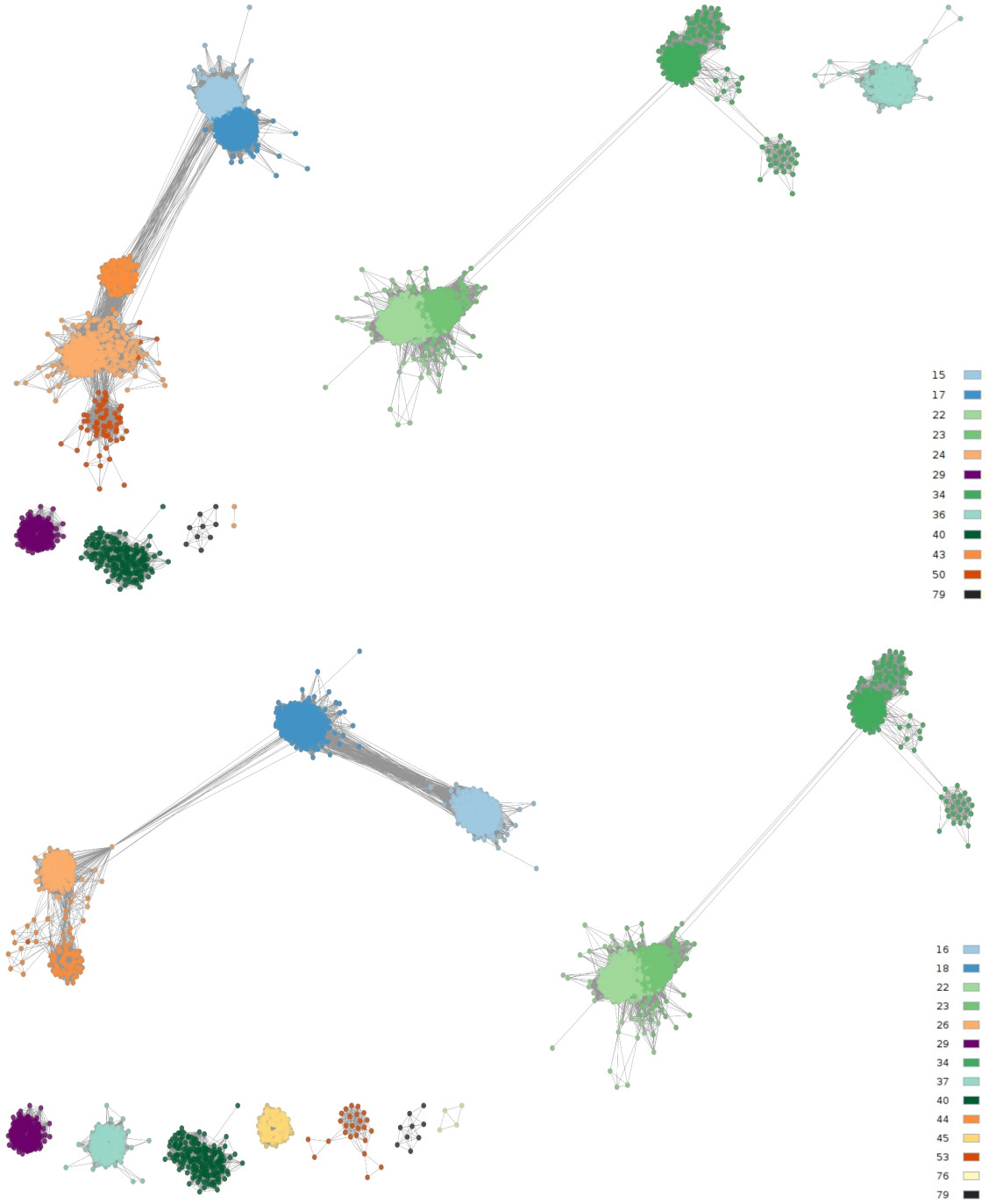
The matrix is hierarchically clustered.

Supplementary Fig. 7.6. Maximum intercluster global identity



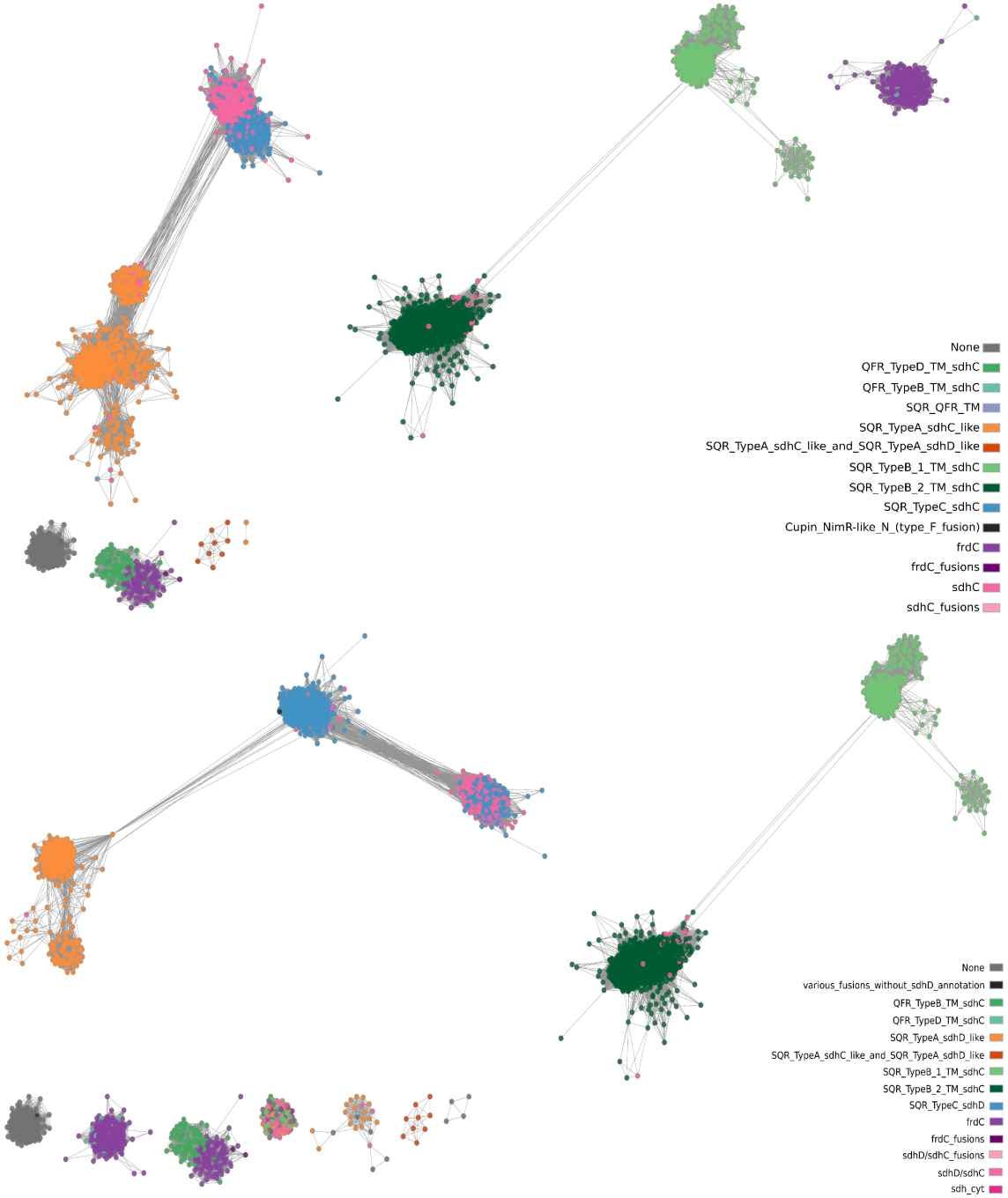
The matrix is hierarchically clustered.

Supplementary Fig. 7.7. Similarity networks for SdhC and SdhD



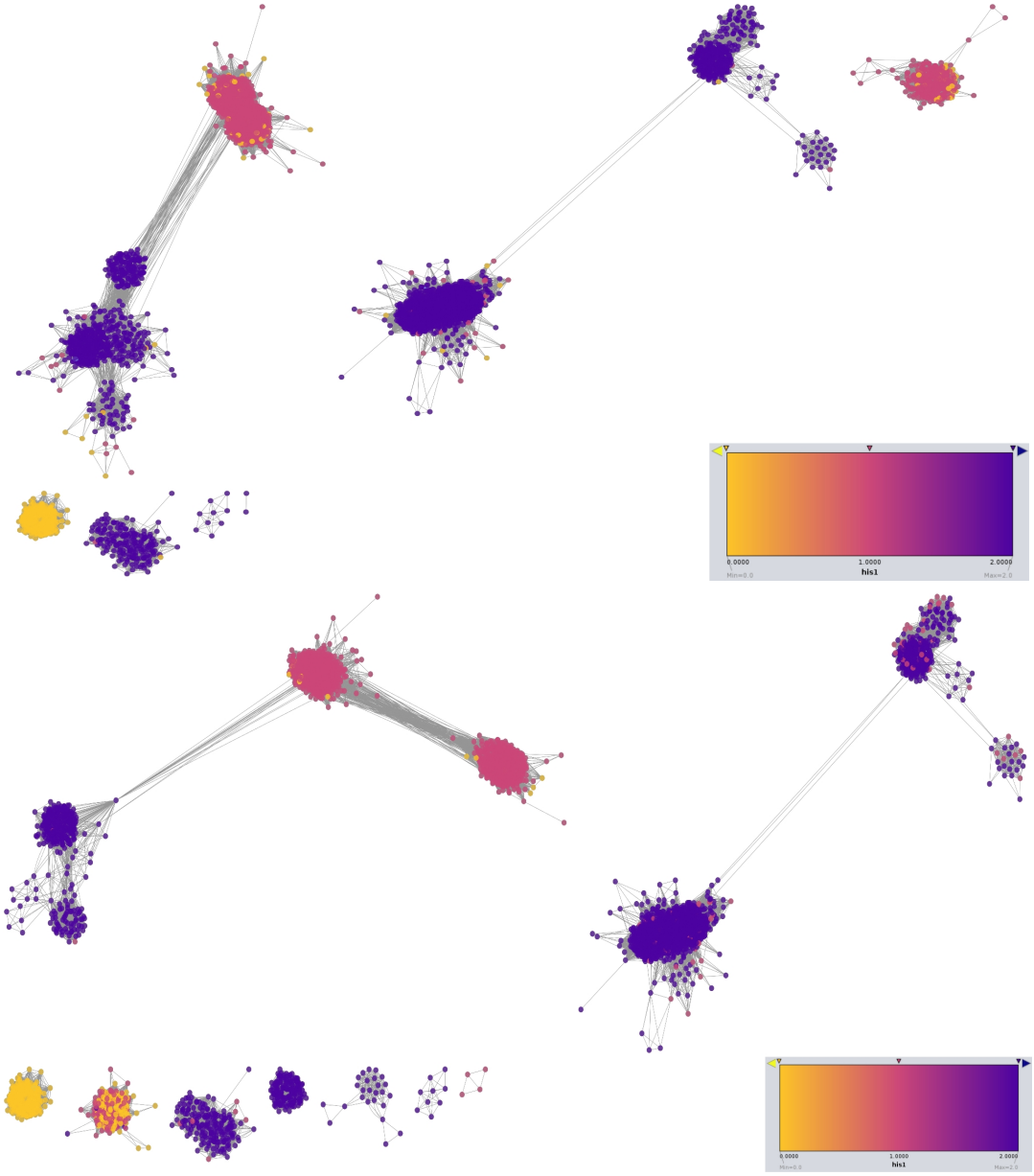
Colored by the MCL cluster number.

Supplementary Fig. 7.8. Similarity networks for SdhC and SdhD



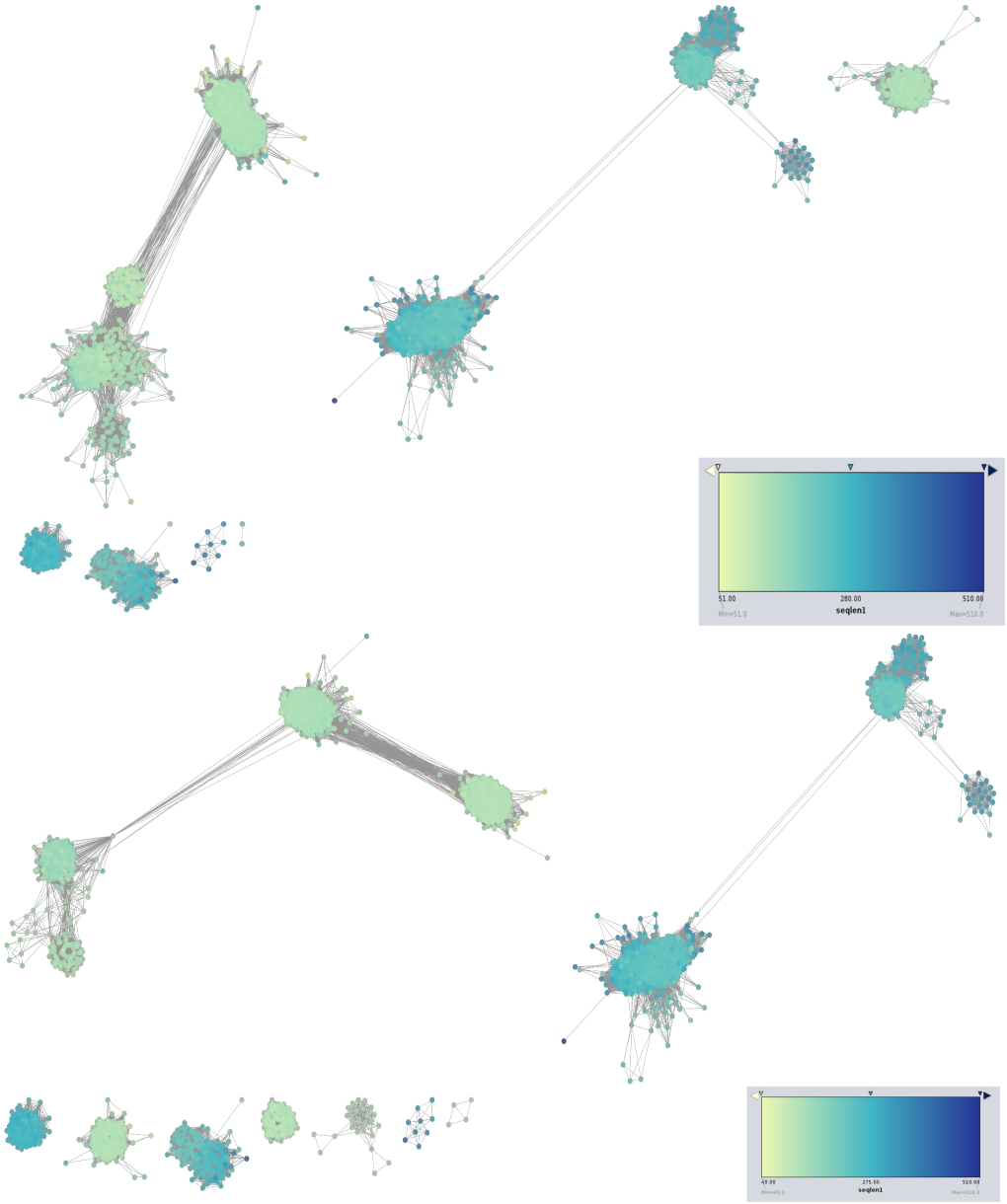
Colored by CD search annotations. Gray colors indicates lack of CD SEARCH annotations.

Supplementary Fig. 7.9. Similarity networks for SdhC and SdhD



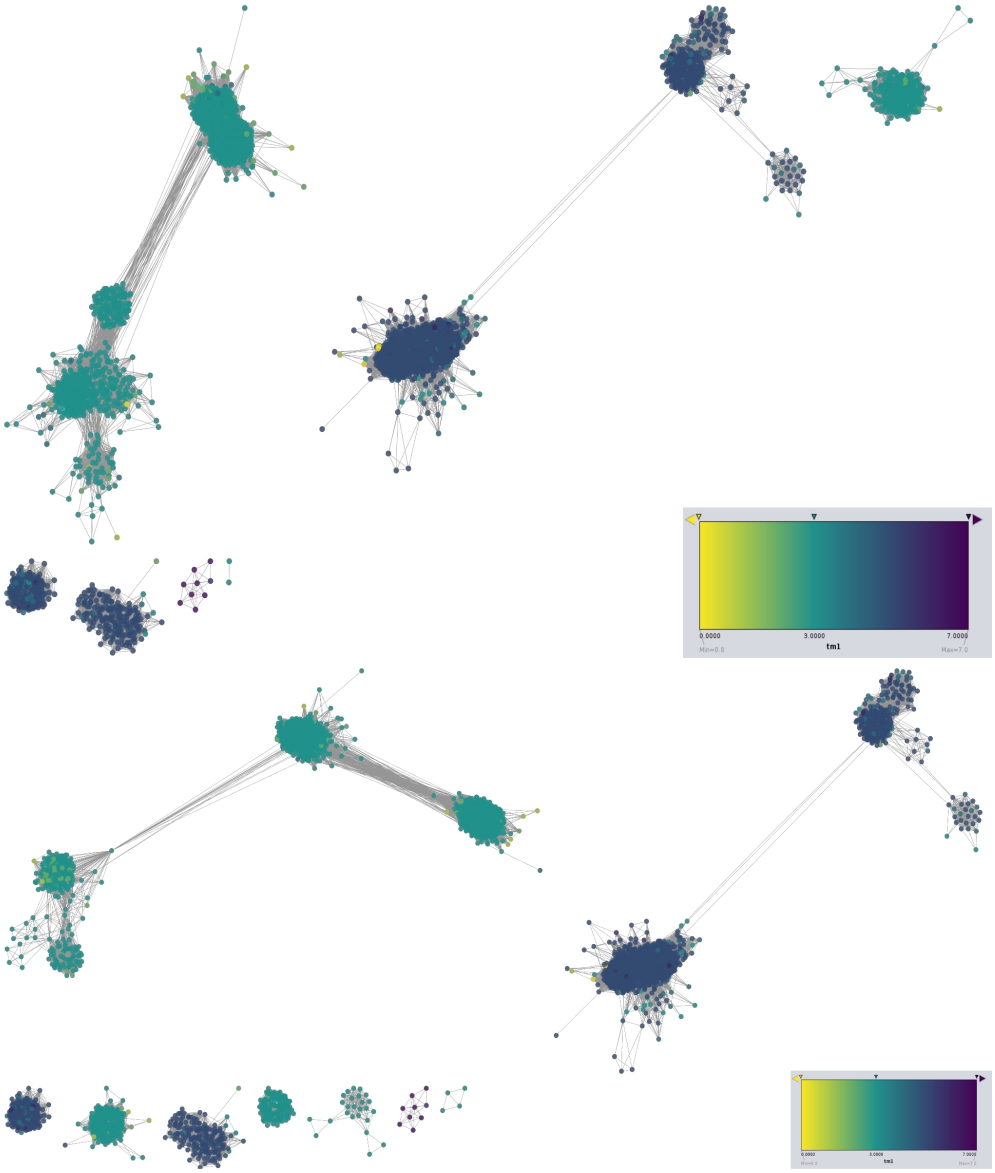
Colored by the number of heme-binding histidines. Type B is colored by the number of histidines per split half of the sequence that was added to the multiple sequence alignment of SdhC, or multiple sequence alignment of SdhD. Purple indicates two histidines, pink is one histidine, and yellow indicates no heme-binding histidines.

Supplementary Fig. 7.10. Similarity networks for SdhC and SdhD



Colored by sequence length. The minimum sequence length for SdhC is 51, for SdhD is 40. These are partial sequences. The maximum sequence length is 510 for both, which is a fusion of type B.

Supplementary Fig. 7.11. Similarity networks for SdhC and SdhD



Colored by the number of transmembrane helices as predicted by TMHMM. The minimum number of helices predicted is zero (yellow), the maximum is 7 (Chloroflexi fusions). Zero helices cases are likely to be a failure of software used (TMHMM, Krogh, *et al*. 2001) to predict helices, in those cases the helix prediction was double-checked using TMPred (Hofmann, Stoffel 1993), which resulted in the prediction of two or three helices for cases of types A, C, and D, and four to six helices for cases of types B and F.

8. References

Aklujkar, Muktak, Julia Krushkal, Genevieve Dibartolo, Alla Lapidus, Miriam L. Land, and Derek R. Lovley. 2009. “The Genome Sequence of Geobacter Metallireducens: Features of Metabolism, Physiology and Regulation Common and Dissimilar to Geobacter Sulfurreducens.” *BMC Microbiology* 9: 1–22. https://doi.org/10.1186/1471-2180-9-109.

Aliabadi, Zarrintaj, Herbert H Winkler, and David O Wood. 1993. “Isolation and Characterization of the Rickettsia Prowazekii Gene Encoding the Flavoprotein Subunit of Succinate Dehydrogenase.” *Gene* 133: 135–40. https://doi.org/https://doi.org/10.1016/0378-1119(93)90238-X.

Allen, J F, J A Raven, Frauke Baymann, Evelyne Lebrun, Myriam Brugna, Barbara Schoepp–Cothenet, Marie–Thérèse Giudici–Orticoni, and Wolfgang Nitschke. 2003. “The Redox Protein Construction Kit: Pre-Last Universal Common Ancestor Evolution of Energy-Conserving Enzymes.” *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358 (1429): 267–74. https://doi.org/10.1098/rstb.2002.1184.

Anderson, Iain J., Priya DasSarma, Susan Lucas, Alex Copeland, Alla Lapidus, Tijana Glavina Del Rio, Hope Tice, et al. 2016. “Complete Genome Sequence of the Antarctic Halorubrum Lacusprofundi Type Strain ACAM 34.” *Standards in Genomic Sciences* 11 (1): 2–7. https://doi.org/10.1186/s40793-016-0194-2.

Anderson, Iain, Markus Göker, Matt Nolan, Susan Lucas, Nancy Hammon, Shweta Deshpande, Jan Fang Cheng, et al. 2011. “Complete Genome Sequence of the Hyperthermophilic Chemolithoautotroph Pyrolobus Fumarii Type Strain (1A T).” *Standards in Genomic Sciences* 4 (3): 381–92. https://doi.org/10.4056/sigs.2014648.

Anderson, Iain, Carla Risso, Dawn Holmes, Susan Lucas, Alex Copeland, Alla Lapidus, Jan Fang Cheng, et al. 2011. “Complete Genome Sequence of Ferroglobus Placidus AEDII12DO.” *Standards in Genomic Sciences* 5 (1): 50–60. https://doi.org/10.4056/sigs.2225018.

Andersson, Siv G. E., Alireza Zomorodipour, Jan O. Andersson, Thomas Sicheritz-Pontén, U. Cecilia M. Alsmark, Raf M. Podowski, A. Kristina Näslund, Ann Sofie Eriksson, Herbert H. Winkler, and Charles G. Kurland. 1998. “The Genome Sequence of Rickettsia Prowazekii and the Origin of Mitochondria.” *Nature* 396 (6707): 133–40. https://doi.org/10.1038/24094.

Anemüller, Stefan, Thomas Hettmann, Ralf Moll, Miguel Teixeira, and Günter Schäfer. 1995. “EPR Characterization of an Archaeal Succinate Dehydrogenase in the Membrane-Bound State.” *European Journal of Biochemistry* 232 (2): 563–68. https://doi.org/10.1111/j.1432-1033.1995.tb20845.x.

Angelov, Angel, and Wolfgang Liebl. 2006. “Insights into Extreme Thermoacidophily Based on Genome Analysis of Picrophilus Torridus and Other Thermoacidophilic Archaea.” *Journal of Biotechnology* 126 (1): 3–10. https://doi.org/10.1016/j.jbiotec.2006.02.017.

Armougom, Fabrice, Sébastien Moretti, Olivier Poirot, Stéphane Audic, Pierre Dumas, Basile Schaeli, Vladimir Keduas, and Cedric Notredame. 2006. “Expresso: Automatic Incorporation of Structural Information in Multiple Sequence Alignments Using 3D-Coffee.” *Nucleic Acids Research* 34 (WEB. SERV. ISS.): 604–8. https://doi.org/10.1093/nar/gkl092.

Auernik, Kathryne S., and Robert M. Kelly. 2008. “Identification of Components of Electron Transport Chains in the Extremely Thermoacidophilic Crenarchaeon Metallosphaera Sedula through Iron and Sulfur Compound Oxidation Transcriptomes.” *Applied and Environmental Microbiology* 74 (24): 7723–32. https://doi.org/10.1128/AEM.01545-08.

Baar, Claudia, Mark Eppinger, Guenter Raddatz, Jörg Simon, Christa Lanz, Oliver Klimmek, Ramkumar Nandakumar, et al. 2003. “Complete Genome Sequence and Analysis of Wolinella Succinogenes.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (20): 11690–95. https://doi.org/10.1073/pnas.1932838100.

Bapteste, Eric, Leo van Iersel, Axel Janke, Scot Kelchner, Steven Kelk, James O McInerney, David A Morrison, et al. 2013. “Networks: Expanding Evolutionary Thinking.” *Trends in Genetics* 29 (8): 439–41. https://doi.org/https://doi.org/10.1016/j.tig.2013.05.007.

Barassi, Carlos A., Robert G. Kranz, and Robert B. Gennis. 1985. “Succinate Dehydrogenase in Rhodopseudomonas Sphaeroides: Subunit Composition and Immunocross-Reactivity with Other Related Bacteria.” *Journal of Bacteriology* 163 (2): 778–82. https://doi.org/10.1128/jb.163.2.778-782.1985.

Baughn, Anthony D., and Michael H. Malamy. 2003. “The Essential Role of Fumarate Reductase in Haem-Dependent Growth Stimulation of Bacteroides Fragilis.” *Microbiology* 149 (6): 1551–58. https://doi.org/10.1099/mic.0.26247-0.

Berman, Helen M., John D. Westbrook, Zukang Feng, Gary Gilliland, T .N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. “The Protein Data Bank.” *Nucleic Acids Research* 28 (1): 235–42. https://doi.org/10.1093/nar/28.1.235.

Bezawork-Geleta, Ayenachew, Jakub Rohlena, Lanfeng Dong, Karel Pacak, and Jiri Neuzil. 2017. “Mitochondrial Complex II: At the Crossroads.” *Trends in Biochemical Sciences* 42 (4): 312–25. https://doi.org/https://doi.org/10.1016/j.tibs.2017.01.003.

Blattner, Frederick R., Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, et al. 1997. “The Complete Genome Sequence of Escherichia Coli K-12.” *Science* 277 (5331): 1453–62. https://doi.org/10.1126/science.277.5331.1453.

Bogachev, Alexander V, Yulia V Bertsova, Dmitry A Bloch, and Michael I Verkhovsky. 2012. “Urocanate Reductase: Identification of a Novel Anaerobic Respiratory Pathway in Shewanella Oneidensis MR-1.” *Molecular Microbiology* 86 (6): 1452–63. https://doi.org/https://doi.org/10.1111/mmi.12067.

Bott, Michael, and Axel Niebisch. 2003. “The Respiratory Chain of Corynebacterium Glutamicum.” *Journal of Biotechnology* 104 (1–3): 129–53. https://doi.org/10.1016/S0168-1656(03)00144-5.

Bowman, John P., Sharee A. McCammon, David S. Nichols, Jenny H. Skerratt, Suzanne M. Rea, Peter D. Nichols, and Tom A. McMeekin. 1997. “Shewanella Gelidimarina Sp. Nov. and Shewanella Frigidimarina Sp. Nov., Novel Antarctic Species with the Ability to Produce Eicosapentaenoic Acid (20:5ω3) and Grow Anaerobically by Dissimilatory Fe(III) Reduction.” *International Journal of Systematic Bacteriology* 47 (4): 1040–47. https://doi.org/10.1099/00207713-47-4-1040.

Buan, Nicole R., and William W. Metcalf. 2010. “Methanogenesis by Methanosarcina Acetivorans Involves Two Structurally and Functionally Distinct Classes of Heterodisulfide Reductase.” *Molecular Microbiology* 75 (4): 843–53. https://doi.org/10.1111/j.1365-2958.2009.06990.x.

Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. “Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND.” *Nature Methods* 18 (4): 366–68. https://doi.org/10.1038/s41592-021-01101-x.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson. 2015. “Fast and Sensitive Protein Alignment Using DIAMOND.” *Nature Methods* 12 (1): 59–60. https://doi.org/10.1038/nmeth.3176.

Buckel, Wolfgang, and Rudolf K. Thauer. 2013. “Energy Conservation via Electron Bifurcating Ferredoxin Reduction and Proton/Na+ Translocating Ferredoxin Oxidation.” *Biochimica et Biophysica Acta - Bioenergetics* 1827 (2): 94–113. https://doi.org/10.1016/j.bbabio.2012.07.002.

Bult, Carol J., Owen White, Gary J. Olsen, Lixin Zhou, Robert D. Fleischmann, Granger G. Sutton, Judith A. Blake, et al. 1996. “Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus Jannaschii.” *Science (New York, N.Y.)* 273 (5278): 1058–73. https://doi.org/10.1126/science.273.5278.1058.

Burans, James P., Melvyn Lynn, and Morris R. Solotorovsky. 1983. “Induction of Active Immunity with Membrane Fractions from Haemophilus Influenzae Type B.” *Infection and Immunity* 41 (1): 285–93. https://doi.org/10.1128/iai.41.1.285-293.1983.

Butler, Jessica E., Richard H. Glaven, Abraham Esteve-Núñez, Cinthia Núñez, Evgenya S. Shelobolina, Daniel R. Bond, and Derek R. Lovley. 2006. “Genetic Characterization of a Single Bifunctional Enzyme for Fumarate Reduction and Succinate Oxidation in Geobacter Sulfurreducens and Engineering of Fumarate Reduction in Geobacter Metallireducens.” *Journal of Bacteriology* 188 (2): 450–55. https://doi.org/10.1128/JB.188.2.450-455.2006.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10: 1–9. https://doi.org/10.1186/1471-2105-10-421.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. “TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses.” *Bioinformatics* 25 (15): 1972–73. https://doi.org/10.1093/bioinformatics/btp348.

Carlson, John H., Stephen F. Porcella, Grant McClarty, and Harlan D. Caldwell. 2005. “Comparative Genomic Analysis of Chlamydia Trachomatis Oculotropic and Genitotropic Strains.” *Infection and Immunity* 73 (10): 6407–18. https://doi.org/10.1128/IAI.73.10.6407-6418.2005.

Caspi, Ron, Richard Billington, Ingrid M. Keseler, Anamika Kothari, Markus Krummenacker, Peter E. Midford, Wai Kit Ong, Suzanne Paley, Pallavi Subhraveti, and Peter D. Karp. 2020. “The MetaCyc Database of Metabolic Pathways and Enzymes-a 2019 Update.” *Nucleic Acids Research* 48 (D1): D455–D453. https://doi.org/10.1093/nar/gkz862.

Cecchini, Gary, Imke Schröder, Robert P. Gunsalus, and Elena Maklashina. 2002. “Succinate Dehydrogenase and Fumarate Reductase from Escherichia Coli.” *Biochimica et Biophysica Acta - Bioenergetics* 1553 (1–2): 140–57. https://doi.org/10.1016/S0005-2728(01)00238-9.

Chang, Antje, Lisa Jeske, Sandra Ulbrich, Julia Hofmann, Julia Koblitz, Ida Schomburg, Meina Neumann-Schaal, Dieter Jahn, and Dietmar Schomburg. 2021. “BRENDA, the ELIXIR Core Data Resource in 2021: New Developments and Updates.” *Nucleic Acids Research* 49 (D1): D498–508. https://doi.org/10.1093/nar/gkaa1025.

Chen, Lanming, Brügger Kim, Skovgaard Marie, Redder Peter, She Qunxin, Torarinsson Elfar, Greve Bo, et al. 2005. “The Genome of Sulfolobus Acidocaldarius, a Model Organism of the Crenarchaeota.” *Journal of Bacteriology* 187 (14): 4992–99. https://doi.org/10.1128/JB.187.14.4992.

Chern, Charng-Kuang, Akikazu Ando, Iwao Kusaka, and Sakuzo Fukui. 1976. “A Succinate Dehydrogenase-Deficient Mutant of Agrobacterium Tumefaciens.” *Agricultural and Biological Chemistry* 40 (4): 779–84. https://doi.org/10.1080/00021369.1976.10862116.

Cole, Stewart T., Karin Eiglmeier, Sohail Ahmed, Nadine Honore, Lynne Elmes, Wayne F. Anderson, and Joel H. Weiner. 1988. “Nucleotide Sequence and Gene-Polypeptide Relationships of the GlpABC Operon Encoding the Anaerobic Sn-Glycerol-3-Phosphate Dehydrogenase of Escherichia Coli K-12.” *Journal of Bacteriology* 170 (6): 2448–56. https://doi.org/10.1128/jb.170.6.2448-2456.1988.

Condon, Caro, Richard Cammack, Daulat S. Patil, and Peter Owen. 1985. “The Succinate Dehydrogenase of Escherichia Coli. Immunochemical Resolution and Biophysical Characterization of a 4-Subunit Enzyme Complex.” *Journal of Biological Chemistry* 260 (16): 9427–34. https://doi.org/10.1016/s0021-9258(17)39384-5.

Consortium, The Uniprot. 2019. “UniProt: A Worldwide Hub of Protein Knowledge.” *Nucleic Acids Research* 47 (D1): D506–15. https://doi.org/10.1093/nar/gky1049.

Cooley, Jason W., Crispin A. Howitt, and Wim F.J. Vermaas. 2000. “Succinate:Quinol Oxidoreductases in the Cyanobacterium Synechocystis Sp. Strain PCC 6803: Presence and Function in Metabolism and Electron Transport.” *Journal of Bacteriology* 182 (3): 714–22. https://doi.org/10.1128/JB.182.3.714-722.2000.

Crowe, Brian A., Peter Owen, Daulat S. Patil, and Richard Cammack. 1983. “Characterization of Succinate Dehydrogenase from Micrococcus Luteus (Lysodeikticus) by Electron-spin-resonance Spectroscopy.” *European Journal of Biochemistry* 137 (1–2): 191–96. https://doi.org/10.1111/j.1432-1033.1983.tb07814.x.

Darriba, Diego, Guillermo L Taboada, Ramón Doallo, and David Posada. 2012. “JModelTest 2: More Models, New Heuristics and High-Performance Computing Europe PMC Funders Group.” *Nature Methods* 9 (8): 772. https://doi.org/10.1038/nmeth.2109.jModelTest.

Davis, K. A., Youssef Hatefi, Irving P. Crawford, and Herrick Baltscheffsky. 1977. “Purification, Molecular Properties, and Amino Acid Composition of the Subunits of Rhodospirillum Rubrum Succinate Dehydrogenase.” *Archives of Biochemistry and Biophysics* 180 (2): 459–64. https://doi.org/10.1016/0003-9861(77)90060-1.

Deckert, Gerard, Patrick V. Warren, Terry Gaasterland, William G. Young, Anna L. Lenox, David E. Graham, Ross Overbeek, et al. 1998. “The Complete Genome of the Hyperthermophilic Bacterium Aquifex Aeolicus.” *Nature* 392 (6674): 353–58. https://doi.org/10.1038/32831.

Dombrowski, Nina, Jun-Hoe Lee, Tom A Williams, Pierre Offre, and Anja Spang. 2019. “Genomic Diversity, Lifestyles and Evolutionary Origins of DPANN Archaea.” *FEMS Microbiology Letters* 366 (2): fnz008. https://doi.org/10.1093/femsle/fnz008.

Dubler, Robert E., William A. Toscano, and Richard A. Hartline. 1974. “Transport of Succinate by Pseudomonas Putida.” *Archives of Biochemistry and Biophysics* 160 (2): 422–29. https://doi.org/10.1016/0003-9861(74)90416-0.

Dunning Hotopp, Julie C., Mingqun Lin, Ramana Madupu, Jonathan Crabtree, Samuel V. Angiuoli, Jonathan Eisen, Rekha Seshadri, et al. 2006. “Comparative Genomics of Emerging Human Ehrlichiosis Agents.” *PLoS Genetics* 2 (2): 208–23. https://doi.org/10.1371/journal.pgen.0020021.

Enright, Anton J., Stijn Van Dongen, and Christos A. Ouzounis. 2002. “An Efficient Algorithm for Large-Scale Detection of Protein Families.” *Nucleic Acids Research* 30 (7): 1575–84. https://doi.org/10.1093/nar/30.7.1575.

Falb, Michaela, Friedhelm Pfeiffer, Peter Palm, Karin Rodewald, Volker Hickmann, Jörg Tittor, and Dieter Oesterhelt. 2005. “Living with Two Extremes: Conclusions from the Genome Sequence of Natronomonas Pharaonis.” *Genome Research* 15 (10): 1336–43. https://doi.org/10.1101/gr.3952905.

Feng, Yutian, Uri Neri, Sean Gosselin, Artemis S Louyakis, R Thane Papke, Uri Gophna, and Johann Peter Gogarten. 2021. “The Evolutionary Origins of Extreme Halophilic Archaeal Lineages.” *Genome Biology and Evolution* 13 (8): evab166. https://doi.org/10.1093/gbe/evab166.

Fernandes, Andreia S., Alexander A. Konstantinov, Miguel Teixeira, and Manuela M. Pereira. 2005. “Quinone Reduction by Rhodothermus Marinus Succinate:Menaquinone Oxidoreductase Is Not Stimulated by the Membrane Potential.” *Biochemical and Biophysical Research Communications* 330 (2): 565–70. https://doi.org/10.1016/j.bbrc.2005.03.015.

Fernandes, Andreia S., Manuela M. Pereira, and Miguel Teixeira. 2001. “The Succinate Dehydrogenase from the Thermohalophilic Bacterium Rhodothermus Marinus: Redox-Bohr Effect on Heme BL1.” *Journal of Bioenergetics and Biomembranes* 33 (4): 343–52. https://doi.org/10.1023/A:1010663424846.

Fitz-Gibbon, Sorel T., Heidi Ladner, Ung Jin Kim, Karl O. Stetter, Melvin I. Simon, and Jeffrey H. Miller. 2002. “Genome Sequence of the Hyperthermophilic Crenarchaeon Pyrobaculum Aerophilum.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (2): 984–89. https://doi.org/10.1073/pnas.241636498.

Florin, C, T Köhler, M Grandguillot, and P Plesiat. 1996. “Comamonas Testosteroni 3-Ketosteroid-Delta 4(5 Alpha)-Dehydrogenase: Gene and Protein Characterization.” *Journal of Bacteriology* 178 (11): 3322–30. https://doi.org/10.1128/jb.178.11.3322-3330.1996.

Frank, Yulia A., Vitaly V. Kadnikov, Anastasia P. Lukina, David Banks, Alexey V. Beletsky, Andrey V. Mardanov, Elena I. Sen’kina, Marat R. Avakyan, Olga V. Karnachuk, and Nikolai V. Ravin. 2016. “Characterization and Genome Analysis of the First Facultatively Alkaliphilic Thermodesulfovibrio Isolated from the Deep Terrestrial Subsurface.” *Frontiers in Microbiology* 7 (DEC). https://doi.org/10.3389/fmicb.2016.02000.

Fuchs, Tanja, Harald Huber, Kristina Teiner, Siegfried Burggraf, and Karl O. Stetter. 1995. “Metallosphaera Prunae, Sp. Nov., a Novel Metal-Mobilizing, Thermoacidophilic Archaeum, Isolated from a Uranium Mine in Germany.” *Systematic and Applied Microbiology* 18 (4): 560–66. https://doi.org/10.1016/S0723-2020(11)80416-9.

Fütterer, Ole, Angel Angelov, Heiko Liesegang, Gerhard Gottschalk, Christa Schleper, Bernd Schepers, Christiane Dock, Garabed Antranikian, and Wolfgang Liebl. 2004. “Genome Sequence of Picrophilus Torridus and Its Implications for Life around PH 0.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (24): 9091–96. https://doi.org/10.1073/pnas.0401356101.

García, L. M., Martha L. Contreras-Zentella, Ruben Jaramillo, M. C. Benito-Mercadé, G. Mendoza-Hernández, I. P. Del Arenal, J. Membrillo-Hernández, and J. Edgardo Escamilla. 2008. “The Succinate:Menaquinone Reductase of Bacillus Cereus - Characterization of the Membrane-Bound and Purified Enzyme.” *Canadian Journal of Microbiology* 54 (6): 456–66. https://doi.org/10.1139/W08-037.

Gomes, Cláudio M., Rita S. Lemos, Miguel Teixeira, Arnulf Kletzin, Harald Huber, Karl O. Stetter, Günter Schäfer, and Stefan Anemüller. 1999. “The Unusual Iron Sulfur Composition of the Acidianus Ambivalens Succinate Dehydrogenase Complex.” *Biochimica et Biophysica Acta - Bioenergetics* 1411 (1): 134–41. https://doi.org/10.1016/S0005-2728(99)00046-8.

Gong, Hongri, Yan Gao, Xiaoting Zhou, Yu Xiao, Weiwei Wang, Yanting Tang, Shan Zhou, et al. 2020. “Cryo-EM Structure of Trimeric Mycobacterium Smegmatis Succinate Dehydrogenase with a Membrane-Anchor SdhF.” *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-18011-9.

Goodner, Brad, Gregory Hinkle, Stacie Gattung, Nancy Miller, Mary Blanchard, Barbara Qurollo, Barry S. Goldman, et al. 2001. “Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium Tumefaciens C58 Brad.” *Science* 294 (14): 2323–28. doi: 10.1126/science.1066803.

Gradin, Christina H., Lars Hederstedt, and Herrick Baltscheffsky. 1985. “Soluble Succinate Dehydrogenase from the Halophilic Archaebacterium, Halobacterium Halobium.” *Archives of Biochemistry and Biophysics* 239 (1): 200–205. https://doi.org/10.1016/0003-9861(85)90827-6.

Guan, Hong Hsiang, Yin Cheng Hsieh, Pei Ju Lin, Yen Chieh Huang, Masato Yoshimura, Li Ying Chen, Shao Kang Chen, et al. 2018. “Structural Insights into the Electron/Proton Transfer Pathways in the Quinol:Fumarate Reductase from Desulfovibrio Gigas.” *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-33193-5.

Guccione, Edward, Andrew Hitchcock, Stephen J. Hall, Francis Mulholland, Neil Shearer, Arnoud H.M. van Vliet, and David J. Kelly. 2010. “Reduction of Fumarate, Mesaconate and Crotonate by Mfr, a Novel Oxygen-Regulated Periplasmic Reductase in Campylobacter Jejuni.” *Environmental Microbiology* 12 (3): 576–91. https://doi.org/10.1111/j.1462-2920.2009.02096.x.

Guy, Lionel, Jens Roat Kultima, Siv G.E. Andersson, and John Quackenbush. 2011. “GenoPlotR: Comparative Gene and Genome Visualization in R.” *Bioinformatics* 27 (13): 2334–35. https://doi.org/10.1093/bioinformatics/btq413.

Hägerhäll, Cecilia. 1997. “Succinate: Quinone Oxidoreductases.” *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1320 (2): 107–41. https://doi.org/10.1016/s0005-2728(97)00019-4.

Hägerhäll, Cecilia, and Lars Hederstedt. 1996. “A Structural Model for the Membrane-Integral Domain of Succinate:Quinone Oxidoreductases.” *FEBS Letters* 389 (1): 25–31. https://doi.org/10.1016/0014-5793(96)00529-7.

Hamann, Nils, Eckhard Bill, Jacob E. Shokes, Robert A. Scott, Marina Bennati, and Reiner Hedderich. 2009. “The CCG-Domain-Containing Subunit SdhE of Succinate:Quinone Oxidoreductase from Sulfolobus Solfataricus P2 Binds a [4Fe-4S] Cluster.” *Journal of Biological Inorganic Chemistry* 14 (3): 457–70. https://doi.org/10.1007/s00775-008-0462-8.

Hards, Kiel, Salome Molano Rodriguez, Charlotte Cairns, and Gregory M. Cook. 2019. “Alternate Quinone Coupling in a New Class of Succinate Dehydrogenase May Potentiate Mycobacterial Respiratory Control.” *FEBS Letters* 593 (5): 475–86. https://doi.org/10.1002/1873-3468.13330.

Harrison, Alistair, David W. Dyer, Allison Gillaspy, William C. Ray, Rachna Mungur, Matthew B. Carson, Huachun Zhong, et al. 2005. “Genomic Sequence of an Otitis Media Isolate of Nontypeable Haemophilus Influenzae: Comparative Study with H. Influenzae Serotype d, Strain KW20.” *Journal of Bacteriology* 187 (13): 4627–36.

Hartman, Amber L., Cédric Norais, Jonathan H. Badger, Stéphane Delmas, Sam Haldenby, Ramana Madupu, Jeffrey Robinson, et al. 2010. “The Complete Genome Sequence of Haloferax Volcanii DS2, a Model Archaeon.” *PLoS ONE* 5 (3). https://doi.org/10.1371/journal.pone.0009605.

Hedderich, Reiner, Jürgen Koch, Dietmar Linder, and Rudolf K. Thauer. 1994. “The Heterodisulfide Reductase from Methanobacterium Thermoautotrophicum Contains Sequence Motifs Characteristic of Pyridine-Nucleotide-Dependent Thioredoxin Reductases.” *European Journal of Biochemistry* 225 (1): 253–61. https://doi.org/10.1111/j.1432-1033.1994.00253.x.

Hederstedt, Lars. 1999. “Respiration Without O2.” *Science* 284 (5422): 1941–42. https://doi.org/10.1126/science.284.5422.1941.

Hederstedt, Lars. 2002. “Succinate:Quinone Oxidoreductase in the Bacteria Paracoccus Denitrificans and Bacillus Subtilis.” *Biochimica et Biophysica Acta - Bioenergetics* 1553 (1–2): 74–83. https://doi.org/10.1016/S0005-2728(01)00231-6.

Hederstedt, Lars. 2003. “Complex II Is Complex Too.” *Science* 299: 671–72. https://doi.org/10.1126/science.1081821.

Hederstedt, Lars, and Lars Rutberg. 1981. “Succinate Dehydrogenase - a Comparative Review.” *Microbiological Reviews* 45 (4): 542–55. https://doi.org/10.1128/mmbr.45.4.542-555.1981.

Heidelberg, John F., Jonathan A. Elsen, William C. Nelson, Rebecca A. Clayton, Michelle L. Gwinn, Robert J. Dodson, Daniel H. Haft, et al. 2000. “DNA Sequence of Both Chromosomes of the Cholera Pathogen Vibrio Cholerae.” *Nature* 406 (6795): 477–83. https://doi.org/10.1038/35020000.

Heidelberg, John F., Ian T. Paulsen, Karen E. Nelson, Eric J. Gaidos, William C. Nelson, Timothy D. Read, Jonathan A. Eisen, et al. 2002. “Genome Sequence of the Dissimilatory Metal Ion-Reducing Bacterium Shewanella Oneidensis.” *Nature Biotechnology* 20 (11): 1118–23. https://doi.org/10.1038/nbt749.

Heim, Steffen, Andreas Künkel, Rudolf K. Thauer, and Reiner Hedderich. 1998. “Thiol:Fumarate Reductase (Tfr) from Methanobacterium Thermoautotrophicum. Identification of the Catalytic Sites for Fumarate Reduction and Thiol Oxidation.” *European Journal of Biochemistry* 253 (1): 292–99. https://doi.org/10.1046/j.1432-1327.1998.2530292.x.

Henne, Anke, Holger Brüggemann, Carsten Raasch, Arnim Wiezer, Thomas Hartsch, Heiko Liesegang, Andre Johann, et al. 2004. “The Genome Sequence of the Extreme Thermophile Thermus Thermophilus.” *Nature Biotechnology* 22 (5): 547–53. https://doi.org/10.1038/nbt956.

Hoang, Diep Thi, Olga Chernomor, Arndt Von Haeseler, Bui Quang Minh, and Le Sy Vinh. 2018. “UFBoot2: Improving the Ultrafast Bootstrap Approximation.” *Molecular Biology and Evolution* 35 (2): 518–22. https://doi.org/10.1093/molbev/msx281.

Hofmann, Kay, and W. Stoffel. 1993. “Tmbase-A Database of Membrane Spanning Protein Segments.” *Biological Chemistry* 374: 166.

Holland, Heinrich D. 2006. “The Oxygenation of the Atmosphere and Oceans.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1470): 903–15. https://doi.org/10.1098/rstb.2006.1838.

Honoré, Nadine, and Stewart T. Cole. 1985. “Cloning and Molecular Genetic Analysis of the Frd Operon Encoding the Fumarate Reductase of Proteus Vulgaris.” *FEMS Microbiology Letters* 30 (3): 393–97. https://doi.org/10.1111/j.1574-6968.1985.tb01117.x.

Huber, Harald, Michael J. Hohn, Karl O. Stetter, and Reinhard Rachel. 2003. “The Phylum Nanoarchaeota: Present Knowledge and Future Perspectives of a Unique Form of Life.” *Research in Microbiology* 154 (3): 165–71. https://doi.org/10.1016/S0923-2508(03)00035-4.

Ikeda, Masata, and Satoshi Nakagawa. 2003. “The Corynebacterium Glutamicum Genome: Features and Impacts on Biotechnological Processes.” *Applied Microbiology and Biotechnology* 62 (2–3): 99–109. https://doi.org/10.1007/s00253-003-1328-1.

Ivanova, Natalia, Alexei Sorokin, Iain Anderson, Nathalie Galleron, Benjamin Candelon, Vinayak Kapatral, Anamitra Bhattacharyya, et al. 2003. “Genome Sequence of Bacillus Cereus and Comparative Analysis with Bacillus Anthracis.” *Nature* 423 (6935): 87–91. https://doi.org/10.1038/nature01582.

Iverson, Tina M. 2013. “Catalytic Mechanisms of Complex II Enzymes: A Structural Perspective.” *Biochimica et Biophysica Acta - Bioenergetics* 1827 (5): 648–57. https://doi.org/10.1016/j.bbabio.2012.09.008.

Iwasaki, Toshio, Takayoshi Wakagi, and Tairo Oshima. 1995. “Resolution of the Aerobic Respiratory System of the Thermoacidophilic Archaeon, Sulfolobus Sp. Strain 7: III. The Archaeal Novel Respiratory Complex II (Succinate:Caldariellaquinone Oxidoreductase Complex) Inherently Lacks Heme Group.” *Journal of Biological Chemistry* 270 (52): 30902–8. https://doi.org/10.1074/jbc.270.52.30902.

Jahn, Ulrike, Harald Huber, Wolfgang Eisenreich, Michael Hügler, and Georg Fuchs. 2007. “Insights into the Autotrophic CO2 Fixation Pathway of the Archaeon Ignicoccus Hospitalis: Comprehensive Analysis of the Central Carbon Metabolism.” *Journal of Bacteriology* 189 (11): 4108–19. https://doi.org/10.1128/JB.00047-07.

Janssen, Silke, Günter Schäfer, Stefan Anemüller, and Ralf Moll. 1997. “A Succinate Dehydrogenase with Novel Structure and Properties from the Hyperthermophilic Archaeon Sulfolobus Acidocaldarius: Genetic and Biophysical Characterization.” *Journal of Bacteriology* 179 (17): 5560–69. https://doi.org/10.1128/jb.179.17.5560-5569.1997.

Jardim-Messeder, Douglas, Caroline Cabreira-Cagliari, Rafael Rauber, Andreia Carina Turchetto-Zolet, Rogério Margis, and Márcia Margis-Pinheiro. 2017. “Fumarate Reductase Superfamily: A Diverse Group of Enzymes Whose Evolution Is Correlated to the Establishment of Different Metabolic Pathways.” *Mitochondrion* 34 (May): 56–66. https://doi.org/10.1016/j.mito.2017.01.002.

Jiang, Chengjian, Yu Liu, Can Meng, Lanlan Wu, Jie Huang, Jie Deng, Jinyi Wang, Peihong Shen, and Bo Wu. 2013. “Expression of a Metagenome-Derived Fumarate Reductase from Marine Microorganisms and Its Characterization.” *Folia Microbiologica* 58 (6): 663–71. https://doi.org/10.1007/s12223-013-0256-4.

Juhnke, Hanno D., Heiko Hiltscher, Hamid R. Nasiri, Harald Schwalbe, and C. Roy D. Lancaster. 2009. “Production, Characterization and Determination of the Real Catalytic Properties of the Putative ‘succinate Dehydrogenase’ from Wolinella Succinogenes.” *Molecular Microbiology* 71 (5): 1088–1101. https://doi.org/10.1111/j.1365-2958.2008.06581.x.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K.F. Wong, Arndt Von Haeseler, and Lars S. Jermiin. 2017. “ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates.” *Nature Methods* 14 (6): 587–89. https://doi.org/10.1038/nmeth.4285.

Kanehisa, Minoru, and Susumu Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1): 27–30. https://doi.org/10.1093/nar/28.1.27.

Kaneko, Takakazu, Shusei Sato, Hirokazu Kotani, Ayako Tanaka, Erika Asamizu, Yasukazu Nakamura, Nobuyuki Miyajima, et al. 1996. “Sequence Analysis of the Genome of the Unicellular Cyanobacterium Synechocystis Sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-Coding Regions.” *DNA Research* 3 (3): 185–209. https://doi.org/10.1093/dnares/3.3.185.

Kawarabayasi, Yutaka, Yumi Hino, Hiroshi Horikawa, Koji Jin-No, Mikio Takahashi, Mitsuo Sekine, Sin Ichi Baba, et al. 2001. “Complete Genome Sequence of an Aerobic Thermoacidophilic Crenarchaeon, Sulfolobus Tokodaii Strain.” *DNA Research* 8 (4): 123–40. https://doi.org/10.1093/dnares/8.4.123.

Kawarabayasi, Yutaka, Yumi Hino, Hiroshi Horikawa, Syuji Yamazaki, Yuji Haikawa, Koji Jin-No, Mikio Takahashi, et al. 1999. “Complete Genome Sequence of an Aerobic Hyper-Thermophilic Crenarchaeon, Aeropyrum Pernix K1.” *DNA Research* 6 (2): 83–101. https://doi.org/10.1093/dnares/6.2.83.

Kawarabayasi, Yutaka, Mituhiro Sawada, Hiroshi Horikawa, Yuji Haikawa, Yumi Hino, Saori Yamamoto, Mitsuo Sekine, et al. 1998. “Complete Sequence and Gene Organization of the Genome of a Hyper-Thermophilic Archaebacterium, Pyrococcus Horikoshii OT3.” *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes* 5 (2): 147–55. https://doi.org/10.1093/dnares/5.2.147.

Kerou, Melina, Pierre Offre, Luis Valledor, Sophie S. Abby, Michael Melcher, Matthias Nagler, Wolfram Weckwerth, and Christa Schleper. 2016. “Proteomics and Comparative Genomics of Nitrososphaera Viennensis Reveal the Core Genome and Adaptations of Archaeal Ammonia Oxidizers.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (49): E7937–46. https://doi.org/10.1073/pnas.1601212113.

Klenk, Hans Peter, Rebecca A. Clayton, Jean Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, et al. 1997. “The Complete Genome Sequence of the Hyperthermophilic, Sulphate-Reducing Archaeon Archaeoglobus Fulgidus.” *Nature* 390 (6658): 364–70. https://doi.org/10.1038/37052.

Kolaj-Robin, Olga, Mohamed R. Noor, Sarah R. O’Kane, Frauke Baymann, and Tewfik Soulimane. 2013. “Atypical Features of Thermus Thermophilus Succinate:Quinone Reductase.” *PLoS ONE* 8 (1). https://doi.org/10.1371/journal.pone.0053559.

Kolaj-Robin, Olga, Sarah R. O’Kane, Wolfgang Nitschke, Christophe Léger, Frauke Baymann, and Tewfik Soulimane. 2011. “Biochemical and Biophysical Characterization of Succinate: Quinone Reductase from Thermus Thermophilus.” *Biochimica et Biophysica Acta - Bioenergetics* 1807 (1): 68–79. https://doi.org/10.1016/j.bbabio.2010.10.009.

Kolde, Raivo. 2019. “Pheatmap: Pretty Heatmaps.” R Package Version 1.0.12. 2019. https://cran.r-project.org/package=pheatmap.

Krishna Murti, C. R. 1960. “Action of Streptomycin on Vibrio Cholerae.” *Biochemical Journal* 76: 362–69. https://doi.org/10.1042/bj0760362.

Kröger, Achim, Simone Biel, Jörg Simon, Roland Gross, Gottfried Unden, and C. Roy D. Lancaster. 2002. “Fumarate Respiration of Wolinella Succinogenes: Enzymology, Energetics and Coupling Mechanism.” *Biochimica et Biophysica Acta - Bioenergetics* 1553 (1–2): 23–38. https://doi.org/10.1016/S0005-2728(01)00234-1.

Krogh, Anders, Björn Larsson, Gunnar Von Heijne, and Erik L.L. Sonnhammer. 2001. “Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes.” *Journal of Molecular Biology* 305 (3): 567–80. https://doi.org/10.1006/jmbi.2000.4315.

Kummerfeld, Sarah K, and Sarah A Teichmann. 2005. “Relative Rates of Gene Fusion and Fission in Multi-Domain Proteins.” *Trends in Genetics* 21 (1): 25–30. https://doi.org/https://doi.org/10.1016/j.tig.2004.11.007.

Kurokawa, Tatsuki, and Junshi Sakamoto. 2005. “Purification and Characterization of Succinate:Menaquinone Oxidoreductase from Corynebacterium Glutamicum.” *Archives of Microbiology* 183 (5): 317–24. https://doi.org/10.1007/s00203-005-0775-8.

Lancaster, C. Roy D. 2002. “Wolinella Succinogenes Quinol:Fumarate Reductase—2.2-Å Resolution Crystal Structure and the E-Pathway Hypothesis of Coupled Transmembrane Proton and Electron Transfer.” *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1565 (2): 215–31. https://doi.org/https://doi.org/10.1016/S0005-2736(02)00571-0.

Lancaster, C. Roy D. 2011. “The Superfamily of Succinate:Quinone Oxidoreductases and Its Implications for the Cyanobacterial Enzymes.” In *Bioenergetic Processes of Cyanobacteria: From Evolutionary Singularity to Ecological Diversity*, edited by Guenter A Peschek, Christian Obinger, and Gernot Renger, 469–511. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-0388-9_17.

Lancaster, C. Roy D. 2013. “The Di-Heme Family of Respiratory Complex II Enzymes.” *Biochimica et Biophysica Acta - Bioenergetics* 1827 (5): 679–87. https://doi.org/10.1016/j.bbabio.2013.02.012.

Lancaster, C. Roy D. 2001. “Succinate:Quinone Oxidoreductases - What Can We Learn from Wolinella Succinogenes Quinol:Fumarate Reductase?” *FEBS Letters* 504 (3): 133–41. https://doi.org/10.1016/S0014-5793(01)02706-5.

Lancaster, C. Roy D. 2003. “Wolinella Succinogenes Quinol:Fumarate Reductase and Its Comparison to E. Coli Succinate:Quinone Reductase.” *FEBS Letters* 555 (1): 21–28. https://doi.org/10.1016/S0014-5793(03)01100-1.

Lancaster, C. Roy D. 2011. “Succinate:Quinone Oxidoreductases.” In *Encyclopedia of Inorganic and Bioinorganic Chemistry*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119951438.eibc0555.

Lancaster, C. Roy D. 2002. “Succinate:Quinone Oxidoreductases: An Overview.” *Biochimica et Biophysica Acta (BBA)* 1553: 1–6. https://doi.org/10.1016/s0005-2728(01)00240-7.

Lancaster, C. Roy D., Achim Kröger, Manfred Auer, and Hartmut Michel. 1999. “Structure of Fumarate Reductase from Wolinella Succinogenes at 2.2 Å Resolution.” *Nature* 402 (6760): 377–85. https://doi.org/10.1038/46483.

Lemos, Rita S., Andreia S. Fernandes, Manuela M. Pereira, Cláudio M. Gomes, and Miguel Teixeira. 2002. “Quinol:Fumarate Oxidoreductases and Succinate:Quinone Oxidoreductases: Phylogenetic Relationships, Metal Centres and Membrane Attachment.” *Biochimica et Biophysica Acta - Bioenergetics* 1553 (1–2): 158–70. https://doi.org/10.1016/S0005-2728(01)00239-0.

Lemos, Rita S., Cláudio M. Gomes, and Miguel Teixeira. 2001. “Acidianus Ambivalens Complex Ii Typifies a Novel Family of Succinate Dehydrogenases.” *Biochemical and Biophysical Research Communications* 281 (1): 141–50. https://doi.org/10.1006/bbrc.2001.4317.

Lewis, Jeffrey A., and Jorge C. Escalante-Semerena. 2006. “The FAD-Dependent Tricarballylate Dehydrogenase (TcuA) Enzyme of Salmonella Enterica Converts Tricarballylate into Cis-Aconitate.” *Journal of Bacteriology* 188 (15): 5479–86. https://doi.org/10.1128/JB.00514-06.

Leys, David, Alexandre S. Tsapin, Kenneth H. Nealson, Terrance E. Meyer, Michael A. Cusanovich, and Jozef J. Van Beeumen. 1999. “Structure and Mechanism of the Flavocytochrome c Fumarate Reductase of Shewanella Putrefaciens MR-1.” *Nature Structural Biology* 6 (12): 1113–17. https://doi.org/10.1038/70051.

Lim, Si Kyu, Sang Jun Kim, Sun Ho Cha, You Kwan Oh, Hae Jin Rhee, Mi Sun Kim, and Jeong K. Lee. 2009. “Complete Genome Sequence of Rhodobacter Sphaeroides KD131.” *Journal of Bacteriology* 191 (3): 1118–19. https://doi.org/10.1128/JB.01565-08.

Lu, Shennan, Jiyao Wang, Farideh Chitsaz, Myra K. Derbyshire, Renata C. Geer, Noreen R. Gonzales, Marc Gwadz, et al. 2020. “CDD/SPARCLE: The Conserved Domain Database in 2020.” *Nucleic Acids Research* 48 (D1): D265–68. https://doi.org/10.1093/nar/gkz991.

Lu, Zheng, and James A. Imlay. 2017. “The Fumarate Reductase of Bacteroides Thetaiotaomicron, Unlike That of Escherichia Coli, Is Configured so That It Does Not Generate Reactive Oxygen Species.” *MBio* 8 (1). https://doi.org/10.1128/mBio.01873-16.

Madej, M. Gregor, Hamid R. Nasiri, Nicole S. Hilgendorff, Harald Schwalbe, and C. Roy D. Lancaster. 2006. “Evidence for Transmembrane Proton Transfer in a Dihaem-Containing Membrane Protein Complex.” *EMBO Journal* 25 (20): 4963–70. https://doi.org/10.1038/sj.emboj.7601361.

Maier, Tamara M., Judith M. Myers, and Charles R. Myers. 2004. “Erratum: Identification of the Gene Encoding the Sole Physiological Fumarate Reductase in Shewanella Oneidensis MR-1 (Journal of Basic Microbiology (2003) 43:4 (312-327)).” *Journal of Basic Microbiology* 44 (1): 75. https://doi.org/10.1002/jobm.200490000.

Marinoni, Ilaria, Simona Nonnis, Carmine Monteferrante, Peter Heathcote, Elisabeth Härtig, Lars H. Böttger, Alfred X. Trautwein, Armando Negri, Alessandra M. Albertini, and Gabriella Tedeschi. 2008. “Characterization of L-Aspartate Oxidase and Quinolinate Synthase from Bacillus Subtilis.” *FEBS Journal* 275 (20): 5090–5107. https://doi.org/10.1111/j.1742-4658.2008.06641.x.

Massung, Robert F., Shannon L. Hiratzka, Kelly A. Brayton, Guy H. Palmer, and Kemba N. Lee. 2008. “Succinate Dehydrogenase Gene Arrangement and Expression in Anaplasma Phagocytophilum.” *Gene* 414 (1–2): 41–48. https://doi.org/10.1016/j.gene.2008.02.005.

Matsson, Mikael, Daria Tolstoy, Roland Aasa, and Lars Hederstedt. 2000. “The Distal Heme Center in Bacillus Subtilis Succinate:Quinone Reductase Is Crucial for Electron Transfer to Menaquinone.” *Biochemistry* 39 (29): 8617–24. https://doi.org/10.1021/bi000271m.

Matsushita, Kazunobu, Mamoru Yamada, Emiko Shinagawa, Osao Adachi, and Minoru Ameyama. 1980. “Membrane-Bound Respiratory Chain of Pseudomonas Aeruginosa Grown Aerobically.” *Journal of Bacteriology* 141 (1): 389–92. https://doi.org/10.1128/jb.141.1.389-392.1980.

Mattevi, Andrea, Gabriella Tedeschi, Luca Bacchella, Alessandro Coda, Armando Negri, and Severino Ronchi. 1999. “Structure of L-Aspartate Oxidase: Implications for the Succinate Dehydrogenase/Fumarate Reductase Oxidoreductase Family.” *Structure* 7 (7): 745–56. https://doi.org/10.1016/S0969-2126(99)80099-9.

Methé, B. A., K. E. Nelson, J. A. Eisen, I. T. Paulsen, W. Nelson, J. F. Heidelberg, D. Wu, et al. 2003. “Genome of Geobacter Sulfurreducens: Metal Reduction in Subsurface Environments.” *Science* 302 (5652): 1967–69. https://doi.org/10.1126/science.1088727.

Metruccio, Matteo M.E., Laura Fantappiè, Davide Serruto, Alessandro Muzzi, Davide Roncarati, Claudio Donati, Vineenzo Scarlato, and Isabel Delany. 2009. “The Hfq-Dependent Small Noncoding RNA NrrF Directly Mediates Fur-Dependent Positive Regulation of Succinate Dehydrogenase in Neisseria Meningitidis.” *Journal of Bacteriology* 191 (4): 1330–42. https://doi.org/10.1128/JB.00849-08.

Meyer, Birte, and Jan Kuevert. 2007. “Phylogeny of the Alpha and Beta Subunits of the Dissimilatory Adenosine-5′ -Phosphosulfate (APS) Reductase from Sulfate-Reducing Prokaryotes - Origin and Evolution of the Dissimilatory Sulfate-Reduction Pathway.” *Microbiology* 153 (7): 2026–44. https://doi.org/10.1099/mic.0.2006/003152-0.

Mileni, Mauro, Fraser MacMillan, Christos Tziatzios, Klaus Zwicker, Alexander H. Haas, Werner Mäntele, Jörg Simon, and C. Roy D. Lancaster. 2006. “Heterologous Production in Wolinella Succinogenes and Characterization of the Quinol:Fumarate Reductase Enzymes from Helicobacter Pylori and Campylobacter Jejuni.” *Biochemical Journal* 395 (1): 191–201. https://doi.org/10.1042/BJ20051675.

Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt Von Haeseler, Robert Lanfear, and Emma Teeling. 2020. “IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.” *Molecular Biology and Evolution* 37 (5): 1530–34. https://doi.org/10.1093/molbev/msaa015.

Minogue, T. D., H. E. Daligault, K. W. Davenport, K. A. Bishop-Lilly, D. C. Bruce, P. S. Chain, S. R. Coyne, et al. 2014. “Draft Genome Assemblies of Proteus Mirabilis ATCC 7002 and Proteus Vulgaris ATCC 49132.” *Genome Announcements* 2 (5): 1064–65. https://doi.org/10.1128/genomeA.01064-14.

Mistry, Jaina, Alex Bateman, and Robert D. Finn. 2007. “Predicting Active Site Residue Annotations in the Pfam Database.” *BMC Bioinformatics* 8 (August). https://doi.org/10.1186/1471-2105-8-298.

Miura, Akane, Masafumi Kameya, Hiroyuki Arai, Masaharu Ishii, and Yasuo Igarashi. 2008. “A Soluble NADH-Dependent Fumarate Reductase in the Reductive Tricarboxylic Acid Cycle of Hydrogenobacter Thermophilus TK-6.” *Journal of Bacteriology* 190 (21): 7170–77. https://doi.org/10.1128/JB.00747-08.

Moll, Ralf, and Günter Schäfer. 1991. “Purification and Characterisation of an Archaebacterial Succinate Dehydrogenase Complex from the Plasma Membrane of the Thermoacidophile Sulfolobus Acidocaldarius.” *European Journal of Biochemistry* 201 (3): 593–600. https://doi.org/10.1111/j.1432-1033.1991.tb16319.x.

Moosavi, Behrooz, Edward A. Berry, Xiao Lei Zhu, Wen Chao Yang, and Guang Fu Yang. 2019. “The Assembly of Succinate Dehydrogenase: A Key Enzyme in Bioenergetics.” *Cellular and Molecular Life Sciences* 76 (20): 4023–42. https://doi.org/10.1007/s00018-019-03200-7.

Morais-Silva, Fabio O., Antonio Mauro Rezende, Catarina Pimentel, Catia I. Santos, Carla Clemente, Ana Varela-Raposo, Daniela M. Resende, et al. 2014. “Genome Sequence of the Model Sulfate Reducer Desulfovibrio Gigas: A Comparative Analysis within the Desulfovibrio Genus.” *MicrobiologyOpen* 3 (4): 513–30. https://doi.org/10.1002/mbo3.184.

Munk, A. Christine, Alex Copeland, Susan Lucas, Alla Lapidus, Tijana Glavina del Rio, Kerrie Barry, John C. Detter, et al. 2011. “Complete Genome Sequence of Rhodospirillum Rubrum Type Strain (S1 T).” *Standards in Genomic Sciences* 4 (3): 293–302. https://doi.org/10.4056/sigs.1804360.

Needleman, Saul B., and Christian D. Wunsch. 1970. “A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins.” *Journal of Molecular Biology* 48 (3): 443–53. https://doi.org/https://doi.org/10.1016/0022-2836(70)90057-4.

Nelson-Sathi, Shijulal, Dagan, Giddy Landan, Arnold Janssen, Mike Steel, James McInerney, Uwe Deppenmeier, and William Martin. 2012. “Acquisition of 1,000 Eubacterial Genes Physiologically Transformed a Methanogen at the Origin of Haloarchaea.” *Proceedings of the National Academy of Sciences* 109 (50): 20537–42. https://doi.org/10.1073/pnas.1209119109.

Nelson-Sathi, Shijulal, Filipa L Sousa, Mayo Roettger, Nabor Lozada-Chávez, Thorsten Thiergart, Arnold Janssen, David Bryant, et al. 2015. “Origins of Major Archaeal Clades Correspond to Gene Acquisitions from Bacteria.” *Nature* 517 (7532): 77–80. https://doi.org/10.1038/nature13805.

Nelson, K. E., C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A.P. Martins Dos Santos, D. E. Fouts, et al. 2003. “Erratum: Complete Genome Sequence and Comparative Analysis of the Metabolically Versatile Pseudomonas Putida KT2440 (Environmental Microbiology 4:12 (799-808)).” *Environmental Microbiology* 5 (7): 630. https://doi.org/10.1046/j.1462-2920.2003.00480.x.

Neukirchen, Sinje, and Filipa L. Sousa. 2021. “DiSCo: A Sequence-Based Type-Specific Predictor of Dsr-Dependent Dissimilatory Sulphur Metabolism in Microbial Data.” *Microbial Genomics* 7 (7). https://doi.org/10.1099/MGEN.0.000603.

Notredame, Cédric, Desmond G. Higgins, and Jaap Heringa. 2000. “T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment.” *Journal of Molecular Biology* 302 (1): 205–17. https://doi.org/10.1006/jmbi.2000.4042.

Pankhurst, Katherine L., Christopher G. Mowat, Emma L. Rothery, Janette M. Hudson, Anne K. Jones, Caroline S. Miles, Malcolm D. Walkinshaw, Fraser A. Armstrong, Graeme A. Reid, and Stephen K. Chapman. 2006. “A Proton Delivery Pathway in the Soluble Fumarate Reductase from Shewanella Frigidimarina.” *Journal of Biological Chemistry* 281 (29): 20589–97. https://doi.org/10.1074/jbc.M603077200.

Park, Soon-Jung -J, Ching-Ping -P Tseng, and Robert P. Gunsalus. 1995. “Regulation of Succinate Dehydrogenase SdhCDAB Operon Expression in Escherichia Coli in Response to Carbon Supply and Anaerobiosis: Role of ArcA and Fnr.” *Molecular Microbiology* 15 (3): 473–82. https://doi.org/10.1111/j.1365-2958.1995.tb02261.x.

Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, et al. 2000. “The Genome Sequence of the Food-Borne Pathogen Campylobacter Jejuni Reveals Hypervariable Sequences.” *Nature* 403 (6770): 665–68. https://doi.org/10.1038/35001088.

Pearson, Melanie M., Mohammed Sebaihia, Carol Churcher, Michael A. Quail, Aswin S. Seshasayee, Nicholas M. Luscombe, Zahra Abdellah, et al. 2008. “Complete Genome Sequence of Uropathogenic Proteus Mirabilis, a Master of Both Adherence and Motility.” *Journal of Bacteriology* 190 (11): 4027–37. https://doi.org/10.1128/JB.01981-07.

Pecsi, Ildiko, Kiel Hards, Nandula Ekanayaka, Michael Berney, Travis Hartman, William R. Jacobs, and Gregory M. Cook. 2014. “Essentiality of Succinate Dehydrogenase in Mycobacterium Smegmatis and Its Role in the Generation of the Membrane Potential under Hypoxia.” *MBio* 5 (4). https://doi.org/10.1128/mBio.01093-14.

Pellicer, Maria Teresa, Josefa Badía, Juan Aguilar, and Laura Baldomà. 1996. “Glc Locus of Escherichia Coli: Characterization of Genes Encoding the Subunits of Glycolate Oxidase and the Glc Regulator Protein.” *Journal of Bacteriology* 178 (7): 2051–59. https://doi.org/10.1128/jb.178.7.2051-2059.1996.

Pennoyer, Jeffrey D., Tomoko Ohnishi, and Bernard L. Trumpower. 1988. “Purification and Properties of Succinate-Ubiquinone Oxidoreductase Complex from Paracoccus Denitrificans.” *Biochimica et Biophysica Acta* 935 (2): 195–207. https://doi.org/10.1016/0005-2728(88)90216-2.

Pereira, Inês A.Cardoso, Ana Raquel Ramos, Fabian Grein, Marta Coimbra Marques, Sofia Marques da Silva, and Sofia Santos Venceslau. 2011. “A Comparative Genomic Analysis of Energy Metabolism in Sulfate Reducing Bacteria and Archaea.” *Frontiers in Microbiology* 2 (APR). https://doi.org/10.3389/fmicb.2011.00069.

Pereira, Manuela M., Tiago M. Bandeiras, Andreia S. Fernandes, Rita S. Lemos, Ana M.P. Melo, and Miguel Teixeira. 2004. “Respiratory Chains from Aerobic Thermophilic Prokaryotes.” *Journal of Bioenergetics and Biomembranes* 36 (1): 93–105. https://doi.org/10.1023/B:JOBB.0000019601.74394.67.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. “UCSF Chimera—A Visualization System for Exploratory Research and Analysis.” *Journal of Computational Chemistry* 25 (13): 1605–12. https://doi.org/https://doi.org/10.1002/jcc.20084.

Pfeiffer, F., S. C. Schuster, A. Broicher, M. Falb, P. Palm, K. Rodewald, A. Ruepp, J. Soppa, J. Tittor, and D. Oesterhelt. 2008. “Evolution in the Laboratory: The Genome of Halobacterium Salinarum Strain R1 Compared to That of Strain NRC-1.” *Genomics* 91 (4): 335–46. https://doi.org/10.1016/j.ygeno.2008.01.001.

Pinchuk, Grigory, Dmitry Rodionov, Chen Yang, Xiaoqing Li, Andrei Osterman, Etienne Dervyn, Oleg Geydebrekht, et al. 2009. “Genomic Reconstruction of Shewanella Oneidensis MR-1 Metabolism Reveals a Previously Uncharacterized Machinery for Lactate Utilization.” *Proceedings of the National Academy of Sciences* 106 (8): 2874–79. https://doi.org/10.1073/pnas.0806798106.

Plesiat, P, M Grandguillot, S Harayama, S Vragar, and Y Michel-Briand. 1991. “Cloning, Sequencing, and Expression of the Pseudomonas Testosteroni Gene Encoding 3-Oxosteroid Delta 1-Dehydrogenase.” *Journal of Bacteriology* 173 (22): 7219–27. https://doi.org/10.1128/jb.173.22.7219-7227.1991.

Podar, Mircea, Iain Anderson, Kira S. Makarova, James G. Elkins, Natalia Ivanova, Mark A. Wall, Athanasios Lykidis, et al. 2008. “A Genomic Analysis of the Archaeal System Ignicoccus Hospitalis-Nanoarchaeum Equitans.” *Genome Biology* 9 (11). https://doi.org/10.1186/gb-2008-9-11-r158.

Ramos, Ana Raquel, Fabian Grein, Gonçalo P. Oliveira, Sofia S. Venceslau, Kimberly L. Keller, Judy D. Wall, and Inês A.C. Pereira. 2015. “The FlxABCD-HdrABC Proteins Correspond to a Novel NADH Dehydrogenase/Heterodisulfide Reductase Widespread in Anaerobic Bacteria and Involved in Ethanol Metabolism in Desulfovibrio Vulgaris Hildenborough.” *Environmental Microbiology* 17 (7): 2288–2305. https://doi.org/10.1111/1462-2920.12689.

Ren, Shuang-Xi, Gang Fu, Xiu-Gao Jiang, Rong Zeng, You-Gang Miao, Hai Xu, Yi-Xuan Zhang, et al. 2003. “Unique Physiological and Pathogenic Features of Leptospira Interrogans Revealed by Whole-Genome Sequencing.” *Nature* 422 (6934): 888–93. https://doi.org/10.1038/nature01597.

Rinke, Christian, Patrick Schwientek, Alexander Sczyrba, Natalia N Ivanova, Iain J Anderson, Jan-Fang Cheng, Aaron Darling, et al. 2013. “Insights into the Phylogeny and Coding Potential of Microbial Dark Matter.” *Nature* 499 (7459): 431–37. https://doi.org/10.1038/nature12352.

Ruepp, Andreas, Werner Graml, Martha Leticia Santos-Martinez, Kristin K. Koretke, Cralg Volker, H. Werner Mewes, Dmitrij Frishman, Susanne Stocker, Andrei N. Lupas, and Wolfgang Baumeister. 2000. “The Genome Sequence of the Thermoacidophilic Scavenger Thermoplasma Acidophilum.” *Nature* 407 (6803): 508–13. https://doi.org/10.1038/35035069.

Ruprecht, Jonathan, Victoria Yankovskaya, Elena Maklashina, So Iwata, and Gary Cecchini. 2009. “Structure of Escherichia Coli Succinate: Quinone Oxidoreductase with an Occupied and Empty Quinone-Binding Site.” *Journal of Biological Chemistry* 284 (43): 29836–46. https://doi.org/10.1074/jbc.M109.010058.

Sakuraba, Haruhiko, Takenori Satomura, Ryushi Kawakami, Sanae Yamamoto, Yutaka Kawarabayasi, Hisasi Kikuchi, and Toshihisa Ohshima. 2002. “L-Aspartate Oxidase Is Present in the Anaerobic Hyperthermophilic Archaeon Pyrococcus Horikoshii OT-3: Characteristics and Role in the de Novo Biosynthesis of Nicotinamide Adenine Dinucleotide Proposed by Genome Sequencing.” *Extremophiles* 6 (4): 275–81. https://doi.org/10.1007/s00792-001-0254-3.

Sakuraba, Haruhiko, Kazunari Yoneda, Issaku Asai, Hideaki Tsuge, Nobuhiko Katunuma, and Toshihisa Ohshima. 2008. “Structure of L-Aspartate Oxidase from the Hyperthermophilic Archaeon Sulfolobus Tokodaii.” *Biochimica et Biophysica Acta - Proteins and Proteomics* 1784 (3): 563–71. https://doi.org/10.1016/j.bbapap.2007.12.012.

Salanoubat, M., S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, et al. 2002. “Genome Sequence of the Plant Pathogen Ralstonia Solanacearum.” *Nature* 415 (6871): 497–502. https://doi.org/10.1038/415497a.

Schäfer, Günter. 1996. “Bioenergetics of the Archaebacterium Sulfolobus.” *Biochimica et Biophysica Acta - Bioenergetics* 1277 (3): 163–200. https://doi.org/10.1016/S0005-2728(96)00104-1.

Schäfer, Günter, Stefan Anemüller, and Ralf Moll. 2002. “Archaeal Complex II: ‘Classical’ and ‘Non-Classical’ Succinate:Quinone Reductases with Unusual Features.” *Biochimica et Biophysica Acta - Bioenergetics* 1553 (1–2): 57–73. https://doi.org/10.1016/S0005-2728(01)00232-8.

Scharf, Birgit, Roswitha Wittenberg, and Martin Engelhard. 1997. “Electron Transfer Proteins from the Haloalkaliphilic Archaeon Natronobacterium Pharaonis: Possible Components of the Respiratory Chain Include Cytochrome Bc and a Terminal Oxidase Cytochrome Ba3.” *Biochemistry* 36 (15): 4471–79. https://doi.org/10.1021/bi962312d.

Schomburg, I., L. Jeske, M. Ulbrich, S. Placzek, A. Chang, and D. Schomburg. 2017. “The BRENDA Enzyme Information System–From a Database to an Expert System.” *Journal of Biotechnology* 261 (November): 194–206. https://doi.org/10.1016/j.jbiotec.2017.04.020.

Shabbiri, Khadija, Waqar Ahmad, Quratulain Syed, and Ahmad Adnan. 2010. “Isolation and Purification of Complex II from Proteus Mirabilis Strain ATCC 29245.” *Brazilian Journal of Microbiology* 41: 796–804. https://doi.org/10.1590/S1517-83822010005000032.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models.” *Genome Research* 13 (22): 426. https://doi.org/10.1101/gr.1239303.metabolite.

She, Qunxin, Rama K. Singh, Fabrice Confalonieri, Yvan Zivanovic, Ghislaine Allard, Mariana J. Awayez, Christina C.Y. Chan-Weiher, et al. 2001. “The Complete Genome of the Crenarchaeon Sulfolobus Solfataricus P2.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (14): 7835–40. https://doi.org/10.1073/pnas.141222098.

Shih, Patrick M., Dongying Wu, Amel Latifi, Seth D. Axen, David P. Fewer, Emmanuel Talla, Alexandra Calteau, et al. 2013. “Improving the Coverage of the Cyanobacterial Phylum Using Diversity-Driven Genome Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (3): 1053–58. https://doi.org/10.1073/pnas.1217107110.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. “Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega.” *Molecular Systems Biology* 7 (539). https://doi.org/10.1038/msb.2011.75.

Sirand-Pugnet, Pascal, Christine Citti, Aurélien Barré, and Alain Blanchard. 2007. “Evolution of Mollicutes: Down a Bumpy Road with Twists and Turns.” *Research in Microbiology* 158 (10): 754–66. https://doi.org/10.1016/j.resmic.2007.09.007.

Smith, Deborah K., and Herbert H. Winkler. 1979. “Separation of Inner and Outer Membranes of Rickettsia Prowazekii and Characterization of Their Polypeptide Compositions.” *Journal of Bacteriology* 137 (2): 963–71. https://doi.org/10.1128/jb.137.2.963-971.1979.

Stieglmeier, Michaela, Andreas Klingl, Ricardo J.E. Alves, Simon K.M.R. Rittmann, Michael Melcher, Nikolaus Leisch, and Christa Schleper. 2014. “Nitrososphaera Viennensis Gen. Nov., Sp. Nov., an Aerobic and Mesophilic, Ammonia-Oxidizing Archaeon from Soil and a Member of the Archaeal Phylum Thaumarchaeota.” *International Journal of Systematic and Evolutionary Microbiology* 64 (PART 8): 2738–52. https://doi.org/10.1099/ijs.0.063172-0.

Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S.L. Brinkman, et al. 2000. “Complete Genome Sequence of Pseudomonas Aeruginosa PAO1, an Opportunistic Pathogen.” *Nature* 406 (6799): 959–64. https://doi.org/10.1038/35023079.

Taguchi, H, and T Ohta. 1991. “D-Lactate Dehydrogenase Is a Member of the D-Isomer-Specific 2-Hydroxyacid Dehydrogenase Family. Cloning, Sequencing, and Expression in Escherichia Coli of the D-Lactate

Dehydrogenase Gene of Lactobacillus Plantarum." *Journal of Biological Chemistry* 266 (19): 12588–94. https://doi.org/10.1016/S0021-9258(18)98939-8.

Team, R Core. 2020. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org.

Tettelin, Hervé, Nigel J. Saunders, John Heidelberg, Alex C. Jeffries, Karen E. Nelson, Jonathan A. Eisen, Karen A. Ketchum, et al. 2000. "Complete Genome Sequence of Neisseria Meningitidis Serogroup B Strain MC58." *Science* 287 (5459): 1809–15. https://doi.org/10.1126/science.287.5459.1809.

Thauer, Rudolf K, Kurt Jungermann, and Karl Decker. 1977. "Energy Conservation in Chemotrophic Anaerobic Bacteria." *Bacteriological Reviews* 41 (1): 100–180.

Tomasiak, Thomas M., Tara L. Archuleta, Juni Andréll, César Luna-Chávez, Tyler A. Davis, Maruf Sarwar, Amy J. Ham, et al. 2011. "Geometric Restraint Drives On- and off-Pathway Catalysis by the Escherichia Coli Menaquinol:Fumarate Reductase." *Journal of Biological Chemistry* 286 (4): 3047–56. https://doi.org/10.1074/jbc.M110.192849.

Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Flelschmann, K. A. Ketchum, et al. 1997. "Erratum: The Complete Genome Sequence of the Gastric Pathogen Helicobacter Pylori (Nature (1997) 388 (539-547))." *Nature* 389 (6649): 412. https://doi.org/10.1038/38792.

Tria, Fernando Domingues Kümmel, Giddy Landan, and Tal Dagan. 2017. "Phylogenetic Rooting Using Minimal Ancestor Deviation." *Nature Ecology & Evolution* 1 (June): 193. https://doi.org/10.1038/s41559-017-0193.

Dongen, Stijn Van. 2000. "Graph Clustering by Flow Simulation." University of Utrecht.

Varma, Sreejith J, Kamila B Muchowska, Paul Chatelain, and Joseph Moran. 2018. "Native Iron Reduces CO2 to Intermediates and End-Products of the Acetyl-CoA Pathway." *Nature Ecology & Evolution* 2 (6): 1019–24. https://doi.org/10.1038/s41559-018-0542-2.

Walker, C. B., J. R. De La Torre, M. G. Klotz, H. Urakawa, N. Pinel, D. J. Arp, C. Brochier-Armanet, et al. 2010. "Nitrosopumilus Maritimus Genome Reveals Unique Mechanisms for Nitrification and Autotrophy in Globally Distributed Marine Crenarchaea." *Proceedings of the National Academy of Sciences of the United States of America* 107 (19): 8818–23. https://doi.org/10.1073/pnas.0913533107.

Weingarten, Rebecca A., Michael E. Taveirne, and Jonathan W. Olson. 2009. "The Dual-Functioning Fumarate Reductase Is the Sole Succinate:Quinone Reductase in Campylobacter Jejuni and Is Required for Full Host Colonization." *Journal of Bacteriology* 191 (16): 5293–5300. https://doi.org/10.1128/JB.00166-09.

Westenberg, David J., and Mary Lou Guerinot. 1999. "Succinate Dehydrogenase (Sdh) from Bradyrhizobium Japonicum Is Closely Related to Mitochondrial Sdh." *Journal of Bacteriology* 181 (15): 4676–79. https://doi.org/10.1128/jb.181.15.4676-4679.1999.

White, Owen, Jonathan A. Eisen, John F. Heidelberg, Erin K. Hickey, Jeremy D. Peterson, Robert J. Dodson, Daniel H. Haft, et al. 1999. "Genome Sequence of the Radioresistant Bacterium Deinococcus Radiodurans R1." *Science* 286 (5444): 1571–77. https://doi.org/10.1126/science.286.5444.1571.

Wimmer, Jessica L E, Joana C Xavier, Andrey d. N Vieira, Delfina P H Pereira, Jacqueline Leidner, Filipa L Sousa, Karl Kleinermanns, Martina Preiner, and William F Martin. 2021. "Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last Universal Common Ancestor (LUCA) ." *Frontiers in Microbiology* . https://www.frontiersin.org/article/10.3389/fmicb.2021.793664.

Wolf, Yuri I., and Eugene V. Koonin. 2012. "A Tight Link between Orthologs and Bidirectional Best Hits in Bacterial and Archaeal Genomes." *Genome Biology and Evolution* 4 (12): 1286–94. https://doi.org/10.1093/gbe/evs100.

Wu, Martin, Ling V. Sun, Jessica Vamathevan, Markus Riegler, Robert Deboy, Jeremy C. Brownlie, Elizabeth A. McGraw, et al. 2004. "Phylogenomics of the Reproductive Parasite Wolbachia Pipientis WMel: A Streamlined Genome Overrun by Mobile Genetic Elements." *PLoS Biology* 2 (3): 327–41. https://doi.org/10.1371/journal.pbio.0020069.

Xu, Jian, Magnus K. Bjursell, Jason Himrod, Su Deng, Lynn K. Carmichael, Herbert C. Chiang, Lora V. Hooper, and Jeffrey I. Gordon. 2003. "A Genomic View of the Human-Bacteroides Thetaiotaomicron Symbiosis." *Science* 299 (5615): 2074–76. https://doi.org/10.1126/science.1080029.

Yankovskaya, Victoria, Rob Horsefield, Susanna Törnroth, Cesar Luna-Chavez, Hideto Miyoshi, Christophe Léger, Bernadette Byrne, Gary Cecchini, and So Iwata. 2003. “Architecture of Succinate Dehydrogenase and Reactive Oxygen Species Generation.” *Science* 299 (5607): 700–704. https://doi.org/10.1126/science.1079605.

Young, Michael, Vladislav Artsatbanov, Harry R. Beller, Govind Chandra, Keith F. Chater, Lynn G. Dover, Ee Been Goh, et al. 2010. “Genome Sequence of the Fleming Strain of Micrococcus Luteus, a Simple Free-Living Actinobacterium.” *Journal of Bacteriology* 192 (3): 841–60. https://doi.org/10.1128/JB.01254-09.

Zomer, Anna W M, Paul A M Michels, and Fred R Opperdoes. 1999. “Molecular Characterisation of Trypanosoma Brucei Alkyl Dihydroxyacetone-Phosphate Synthase.” *Molecular and Biochemical Parasitology* 104 (1): 55–66. https://doi.org/https://doi.org/10.1016/S0166-6851(99)00141-3.