Biochimica et Biophysica Acta 1847 (2015) 451–457

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbabio

Evolution of *FOXRED1*, an FAD-dependent oxidoreductase necessary for NADH:ubiquinone oxidoreductase (Complex I) assembly

Bernard D. Lemire *

Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada, T6G2H7

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form 8 December 2014
Accepted 26 January 2015
Available online 10 February 2015

Keywords:
Leigh syndrome
Phylogeny
Horizontal gene transfer
NADH ubiquinone oxidoreductase
Mitochondrial disease

ABSTRACT

Complex I (NADH:ubiquinone oxidoreductase) is the major entry point for electrons into the respiratory chains of bacteria and mitochondria. Mammalian complex I is composed of 45 subunits and harbors FMN and iron–sulfur cluster cofactors. A heterogeneous disease profile is associated with complex I deficiency. In a large fraction of complex I deficiencies, the primary defect is not in any of the genes encoding a subunit. The proper assembly and function of complex I require the participation of at least 12 assembly factors or chaperones. *FOXRED1* encodes a complex I-specific assembly factor and mutations in this gene result in complex I deficiency, infantile onset encephalomyopathy and Leigh syndrome. The human FOXRED1 protein is a mitochondria-targeted 486-amino acid FAD-dependent oxidoreductase. It is most closely related to *N*-methyl amino acid dehydrogenases. *FOXRED1* orthologs are present in archaea, bacteria and eukaryotes. Fungal *FOXRED1* orthologs were likely acquired from alphaproteobacteria by horizontal gene transfer. The phylogenetic profile of *FOXRED1* orthologs does not parallel the phylogenetic profile of complex I, strongly suggesting that, at least in some organisms, *FOXRED1* has a function unrelated to complex I. The only large clade where all members investigated contain both *FOXRED1* and complex I is the metazoans. Some bacterial *FOXRED1* genes are present in metabolic operons related to amino acid metabolism. *FOXRED1* phylogenetic distribution and gene organization suggest a metabolic role for *FOXRED1* in complex I biogenesis should be considered.

1. Introduction

A powerful technique for assigning function to a protein sequence is the detection of orthologs, proteins derived from a single ancestral gene in the last common ancestor of the compared species [1]. Orthologs typically perform equivalent functions in their respective organisms [2]. Once orthologs are identified, phylogenetic profiles can be constructed to compare the presence or absence of proteins in complete genomes and to predict possible functional interactions between proteins of similar profiles. The underlying hypothesis is that functionally-interacting proteins co-evolve and will have orthologs in the same sets of organisms [3].

Complex I (NADH:ubiquinone oxidoreductase) is the major entry point for electrons into the respiratory chains of bacteria and mitochondria. It is an energy-conserving protein that contributes to the transmembrane proton electrochemical gradient that is used to drive ATP synthesis by the ATP synthase [4].

Complex I is the largest member of the electron transport chain, with 45 subunits in mammals [5,6]. In contrast, the prokaryotic enzyme normally consists of 11 to 14 highly conserved core subunits [7]. Seven core subunits are very hydrophobic and are encoded by the mitochondrial DNA in eukaryotes [7]. Mitochondrial and bacterial complex I form an L-shaped structure, with a large hydrophobic membrane arm and a hydrophilic peripheral arm protruding into the mitochondrial matrix or the bacterial cytoplasm [8–10]. Electron transport is mediated by a flavin mononucleotide cofactor at the site of NADH oxidation and eight to nine iron–sulfur centers that transport the electrons to the site of quinone reduction [4].

Complex I deficiencies can follow different inheritance patterns and etiologies. Mutations can occur in core subunit genes, in the ~30 so-called additional, accessory or supernumerary subunit genes, or in complex I-specific assembly factors [11,12]. The genetics, disease severity and disease progression are dependent on whether the mutated gene is in the mitochondrial DNA or in the nucleus [13]. The roles of most accessory subunits remain poorly defined. They have no direct role in catalysis but may be involved in enzyme assembly, stability and regulation. Some of the accessory subunits are crucial for function; mutations in these subunits lead to severe complex I deficiencies [14]. In at least half of patients with complex I deficiency, the primary defect cannot be ascribed to a subunit [13]. Rather, the deficiency arises from mutation of one of at least 12 assembly factors [15].

Abbreviations: DAO, D-amino acid oxidase; DDO, D-aspartate oxidase; DMGDH, dimethylglycine dehydrogenase; DMGO, dimethylglycine oxidase; FOXRED, FAD-dependent oxidoreductase; L2HGDH, L-2-hydroxyglutarate dehydrogenase; PDPR, pyruvate dehydrogenase phosphatase regulatory subunit; PIPOX, peroxisomal sarcosine oxidase; SARDH, sarcosine dehydrogenase

* Tel.: +1 780 492 4853.
E-mail address: bernard.lemire@ualberta.ca.

http://dx.doi.org/10.1016/j.bbabio.2015.01.014
0005-2728/

FOXRED1 encodes a complex I-specific assembly factor [16]. A homozygous R352W mutation in *FOXRED1* resulted in infantile onset encephalomyopathy and Leigh syndrome [16]. The patient's skeletal muscle had 7% residual complex I activity with a marked reduction in the amount of fully assembled holoenzyme [16]. Lentiviral-mediated expression of a *FOXRED1* transgene rescued complex I deficiency in patient fibroblasts. A second patient presenting with Leigh syndrome was determined to be a *FOXRED1* compound heterozygote with Q232X and N430S mutations; the transcript encoding the former, truncated protein was not detectable, indicating that it was degraded [17]. Fibroblasts from this patient retained 9% residual complex I activity and this deficit could also be rescued by *FOXRED1* expression.

The *FOXRED1* gene encodes a 486-amino acid FAD-dependent oxidoreductase domain containing protein. The protein has a predicted cleavable N-terminal mitochondrial targeting sequence and was shown to be localized to the mitochondrion. Recently, the FOXRED1 protein was found in direct association with complex I assembly intermediates [5].

Phylogenetic trees are graphic representations of multiple sequence alignments. In this study, I investigate the phylogenetic distribution of *FOXRED1* in an effort to understand how and when it became essential for complex I biogenesis. I present evidence that suggests that *FOXRED1* has a metabolic function in some organisms and that its role in complex I biogenesis is a more recent innovation.

2. Materials and methods

Presumptive *FOXRED1* orthologs were identified as symmetrical best Blastp hits with human *FOXRED1* [18]. Various databases, including National Center for Biotechnology Information [19], the Joint Genome Institute [20], Ensembl [21], the Fungal Genome Initiative [22] and genome project web sites were accessed. A list of sequences used in this study is available (supplementary materials). A combination of taxonomy-based sequence selection using the software CDSbank [23] and manual selection was used to identify protein sequences used for tree building. Tree building was carried out with MrBayes (version 3.2.2) [24], RAxML (version 8.0.24) [25] and PHYML (version 3.0) [26]. Sequence alignments were performed using MUSCLE (version 3.8.31) [27] as implemented in Jalview (version 14.0) [28] or on the GUIDANCE server [29]. Phylogenetic analyses were performed using a number of online servers including phylogeny.fr [30], T-REX [31] and the CIPRES Science Gateway (version 3.3) [32].

3. Results and discussion

3.1. FOXRED1: structure and relationship to other proteins.

FOXRED1 is an FAD-dependent oxidoreductase family member; these proteins contain a Pfam DAO (D-amino acid oxidase; PF01266) domain, and are part of the NADP_Rossmann clan CL0063 [33]. The human genome encodes nine DAO domain proteins (Table S1). Six of these, sarcosine dehydrogenase (SARDH), dimethylglycine dehydrogenase (DMGDH), peroxisomal sarcosine oxidase (PIPOX), L-2-hydroxyglutarate dehydrogenase (L2HGDH), D-amino acid oxidase (DAO) and D-aspartate oxidase (DDO) are oxidoreductases involved in amino acid metabolism (Table S2). Glycerol-3-phosphate dehydrogenase (GPD2) is involved in glycerophospholipid metabolism. Pyruvate dehydrogenase phosphatase regulatory subunit (PDPR) is a regulator of pyruvate dehydrogenase phosphatase [34].

FOXRED1, PIPOX, L2HGDH, DAO and DDO are single domain proteins. DMGDH, SARDH, PDPR and GPD2 have additional sequences at their C-termini (Fig. 1). SARDH, DMGDH and PDPR contain C-terminal bipartite glycine cleavage T-protein domains, which bind folate. The T-protein is part of a four protein glycine decarboxylase system involved in glycine metabolism [35]. The structure of the *Arthrobacter globiformis* dimethylglycine oxidase (DMGO; PDB: 1PJ5) has been solved to 1.60 Å

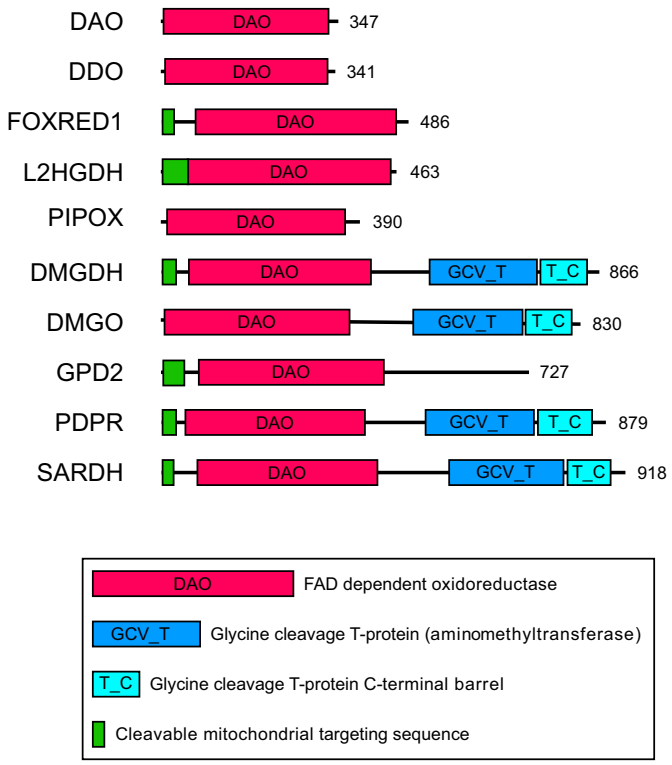

Fig. 1. Domain structure of FOXRED1 homologs. The diagrams represent the domain structures of the human proteins, except for DMGO, which is from *Arthrobacter globiformis*. The DAO, GCV_T, and GCV_T_C domains are Pfam domains PF01266, PF01571 and PF08669, respectively [33]. The number of amino acids in each protein is indicated.

[36]. DMGO is part of a betaine catabolism pathway that converts betaine to sarcosine and glycine.

FOXRED1 is more closely related to human DMGDH, PDPR, PIPOX and SARDH (Blastp evalues: 2×10^{-9}, 6×10^{-6}, 9×10^{-8} and 6×10^{-9}, respectively) than it is to the other human DAO proteins, DAO, DDO, GPD2 and L2HGDH (evalues: >1) (Fig. S1, Table S1). I have limited my analyses to the former proteins. The single domain proteins FOXRED1 and PIPOX and the two domain proteins DMGDH, PDPR and SARDH form two highly supported clades (Fig. 2A, B). FOXREDs form a highly supported sister clade to the PIPOXes, and both of these clades contain bacterial and eukaryotic members. The ancestral FOXRED was likely present in the last common ancestor of bacteria and eukaryotes. Archaeal FOXREDs are not robustly embedded in the FOXRED clade (Fig. S2). However, an ancestral DAO domain was likely present in the last universal common ancestor.

Three SARDH-like PDPRs were identified; these proteins from *Thalassiosira pseudonana*, *Emiliana huxleyi* and *Thecamonas trahens* are symmetrical best hits with human SARDH, but consistently tree with the PDPRs rather than the SARDH (support values 1/79/78; Fig. 2). In trees using full-length protein sequences, the three SARDH-like PDPRs also form a highly supported sister clade (1/100/100) with PDPRs (Fig. S4). This suggests that PDPRs are likely catalytically active in many organisms, unlike the report for the human protein [34].

Unlike SARDH and DMGDH, FOXRED1 does not have a covalently attached FAD cofactor (Fig. S1). Although FOXRED1 is annotated as having a possible transmembrane domain between residues 62 and 82, it is a soluble, carbonate-extractable protein [16]. FOXRED1 does have a mitochondrial targeting sequence and is localized to mitochondria [16]. Additional discussion of the FAD cofactor, the transmembrane segment and the mitochondrial targeting sequence is provided (supplementary materials).

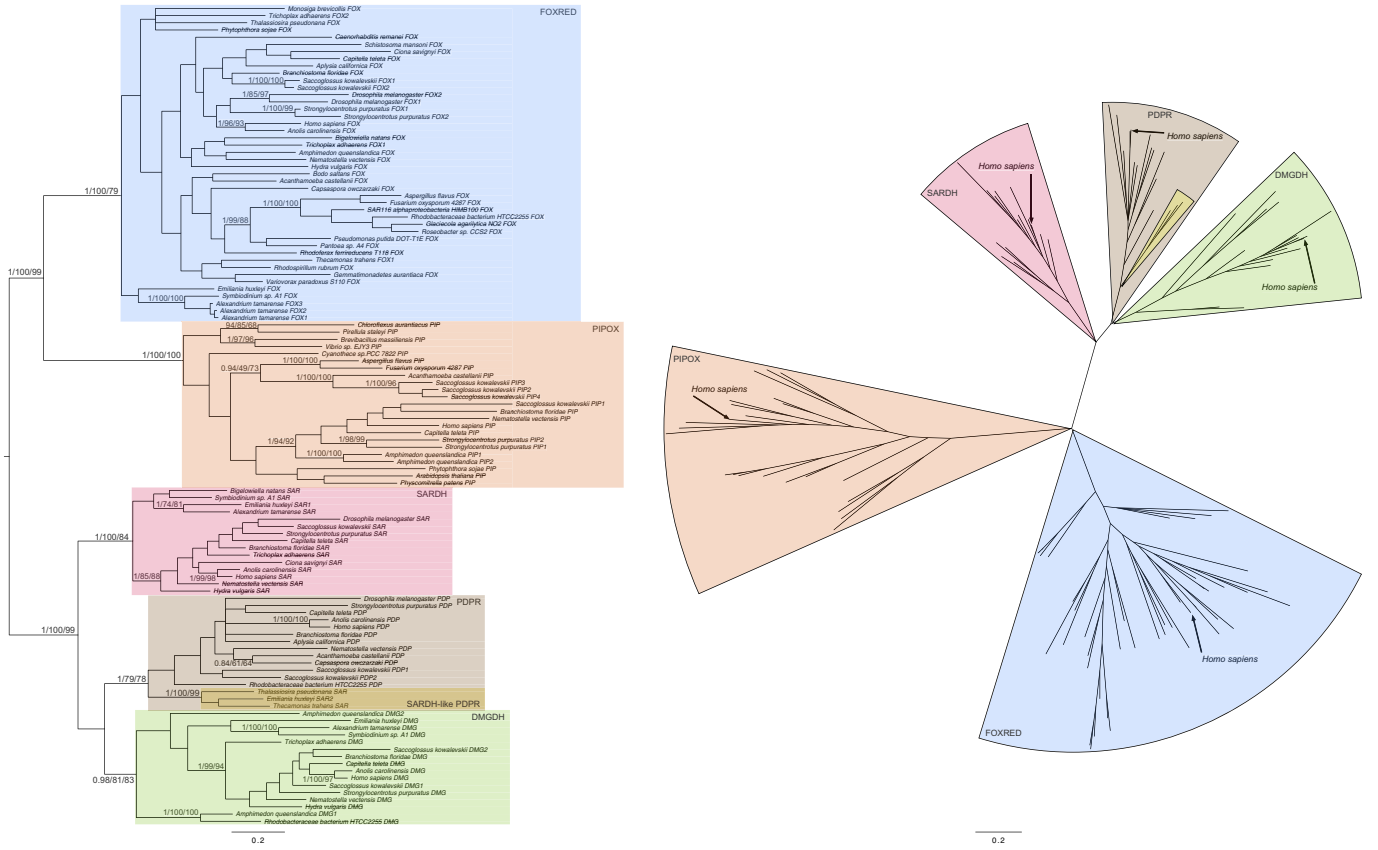

Fig. 2. A) Phylogeny of FOXRED1 and its DMGDH, PDPR, PIPOX and SARDH paralogs. Only DAO domain sequences were aligned using MUSCLE. Support values for nodes are derived from MrBayes/PHYML/RAxML. B) The unrooted tree from MrBayes analysis.

3.2. Phylogenetic distribution of FOXRED1 homologs

I investigated the phylogenetic distribution of the FOXRED1 protein and its four most closely related relatives: SARDH, DMGDH, PIPOX and PDPR. I used the amino terminal DAO domain-containing regions of the five human proteins as query sequences in database searches. For SARDH, PDPR and DMGDH, the C-terminal glycine cleavage T-protein domains were removed (Table 1). Hits were considered putative orthologs if they were symmetrical best hits (Blastp) with the human query sequence. The simplest of organisms with at least one copy of each of the five proteins are *Saccoglossus kowalevskii* (acorn worm), *Strongylocentrotus purpuratus* (purple sea urchin) and *Nematostella vectensis* (sea anemone), all eumetazoans (Table 1).

FOXRED1 distribution does not correlate well with complex I distribution. *FOXRED1* is present in archaea, bacteria and eukaryotes, as is complex I (Table 1, Fig. 3). Many organisms have both a *FOXRED1* ortholog and a complex I. All metazoans investigated, from the simplest, such as *Trichoplax adhaerens*, to the mammals, have both *FOXRED1* and complex I. The metazoans form the only large clade of organisms with both proteins. However, many organisms have a complex I but do not have a *FOXRED1*. Significant amongst these is the kingdom viridiplantae, which includes chlorophyta (green algae) and streptophyta (which includes all land plants), in which I could not identify any FOXREDs. Finally, a small number of organisms have a *FOXRED1* but no complex I; these include bacteria, such as the firmicute, *Bacillus coagulans 36D1*, the gammaproteobacterium, *Glaciecola agarilytica NO2*, and eukaryotes, such as the two alveolates, *Alexandrium tamarense* and *Symbiodinium sp. A1* (Table 1). Several organisms contain more than one FOXRED1; *A. tamarense* has three FOXREDs and many species of *Drosophila* have two.

The recent identification of *FOXRED1* as a disease-causing gene and the paucity of information about its role in the biogenesis of complex I are likely because of its absence in many model system organisms, where mutational analysis can be exploited. Complex I has been most intensively investigated with molecular genetic approaches in the eukaryotes *Neurospora crassa*, *Yarrowia lipolytica, Chlamydomonas reinhardtii* and *Arabidopsis thaliana*, and in the prokaryotes *Escherichia coli* and *Paracoccus denitrificans* (Table 1). None of these organisms has a *FOXRED1*. *FOXRED1* orthologs are present in flies, worms and zebrafish, but have not been extensively investigated (see supplementary materials).

Few fungi have FOXREDs. I could only identify *FOXRED1* orthologs in one phylum, Ascomycota, and only in Sordariomycetes (*Fusarium*) and Eurotiomycetes (*Aspergillus*). The fungal FOXREDS were likely acquired from alphaproteobacteria by horizontal gene transfer (Fig. S5).

FOXRED1 orthologs in *Aspergillus* and *Fusarium* species do not have predicted amino-terminal mitochondrial targeting sequences (Fig. S6). The absence of targeting sequences in fungal FOXREDs is consistent with a metabolic function for the protein; a small molecule product of an enzymatic reaction can easily be transported into mitochondria. It is more difficult to reconcile a role as a chaperone in complex I assembly if the fungal FOXREDs are not localized in the mitochondrial matrix.

FOXRED1 was not detected in currently available genome sequences of amitochondriate organisms (Table 1). Eukaryotes with mitochondria-related organelles are taxonomically diverse but usually occupy anaerobic or microaerophilic environments [37]. Interestingly, both *Nyctotherus ovalis* (a ciliate) and *Blastocystis hominis* (a stramenopile) possess a complex I [38,39]. In both organisms, complex I may be involved in fumarate respiration [38,40].

The co-expression of *FOXRED1* orthologs and metabolic enzymes is strongly suggestive of a metabolic role for *FOXRED1* in bacteria. *FOXRED1* genes are present in known operons in *Sinorhizobium fredii* and *Pseudomonas putida* and likely in a variety of other bacteria [41]

Table 1

The human DAO domain-containing protein sequences FOXRED1 (NP_060017.1, 1–486), PDPR (AAI50252.1, 1–435), SARDH (NP_001128179.1, 1–462), DMGDH (NP_037523.2, 1–447) and PIPOX (NP_057602.2, 1–390) were used as query sequences to identify the most significant hit (accession number) and its Blastp evalue. Best hits that are symmetrical best hits and presumptive orthologs are highlighted in green. The presence of complex I was assessed by searching using human complex I subunit sequences as queries: NDUFV1 (NP_009034), NDUFS7 (NP_077718), ND1 (YP_003024026) and ND5 (YP_003024036).

	Taxonomy	FOXRED present	Number of homologs	Accession top hit	Evalue top hit	Accession top hit	Evalue top hit	Accession top hit	Evalue top hit	Accession top hit	Evalue top hit	Accession top hit	Evalue top hit	Complex I present
Halophilic archaeon DL31	archaea	+	2	YP_004807583	2.E−37	YP_004807838	2.E−20	YP_004807838	3.E−19	YP_004807583	4.E−28	YP_004807838	4.E−26	Yes
Haloferax larsenii	archaea	+	1	WP_007543834	9.E−42	WP_007543834	3.E−23	WP_007543834	6.E−22	WP_007543834	4.E−36	WP_007543834	1.E−16	Yes
Thermococcus litoralis	archaea		1	YP_008428251	2.E−34	YP_008428251	3.E−44	YP_008428251	6.E−47	YP_008428251	7.E−51	YP_008428251	2.E−17	Yes
Bacillus coagulans 36D1	firmicute	+	1	YP_004860295	2.E−66	YP_004860295	9.E−19	YP_004860295	7.E−25	YP_004860295	1.E−19	YP_004860296	1.E−19	No
Chloroflexus aurantiacus J−10−fl	chloroflexi	+	2	YP_001634061	3.E−52	YP_001634061	8.E−42	YP_001634061	2.E−51	YP_001634061	3.E−38	YP_001637283	8.E−54	Yes
Cyanothece sp. PCC 7822	cyanobacteria		2	YP_003890603.1	9.E−05	YP_003890603.1	8.E−10	YP_003889066.1	6.E−08	YP_003889066.1 YP_003890603.1	1.E−11 9.E−11	YP_003889066.1	5.E−71	Yes
Desulfobacterium autotrophicum HRM2	deltaproteobacteria	+	1	YP_002605070	2.E−57	YP_002604158	1.E−16	YP_002604158	5.E−19	YP_002604158	3.E−23	YP_002605070	2.E−05	Yes
Escherichia coli K12	gammaproteobacteria		1	NP_415577.1	5.E−03	NP_415707.1	3.E−11	NP_416744.1	4.E−09	NP_415817.1	8.E−07	NP_415577.1	2.E−45	Yes
Fervidobacterium pennivorans DSM 9078	thermotogae	+	1	YP_005471104	2.E−46	YP_005471104	2.E−39	YP_005471104	6.E−35	YP_005471104	2.E−38	YP_005471104	2.E−21	Yes
Gemmatimonas aurantiaca T−27	gemmatimonadetes	+	1	YP_002763323	2.E−70	YP_002762669	1.E−05	YP_002763323	2.E−05	YP_002762669	6.E−08	YP_002761832	3.E−16	Yes
Glaciecola agarilytica NO2	gammaproteobacteria	+	1	GAC07454.1	2.E−80	BAEK01000084	7.E−04	BAEK01000084	7.E−13	BAEK01000084	2.E−10	BAEK01000084	3.E−05	No
Pantoea sp.A4	gammaproteobacteria	+	2	WP_017345971.1	7.E−83	ALXE01000020	2.E−18	ALXE01000019	7.E−21	AKIT01000029	3.E−19	ALXE01000028.1	4.E−44	Yes
Paracoccus denitrificans	alphaproteobacteria		2	YP_914320	1.E−12	YP_918672	2.E−95	YP_918672	2.E−85	YP_915729	4.E−73	YP_918650	3.E−15	Yes
Pseudomonas putida DOT−T1E	gammaproteobacteria	+	1	YP_006535485	5.E−94	YP_006532686	1.E−21	YP_006535485	1.E−18	YP_006535485	1.E−16	YP_006535239	4.E−08	Yes
Pseudomonas sp. GM33	gammaproteobacteria	+	3	WP_007972670.1	3.E−14	WP_007978235.1	4.E−95	WP_007978235.1	1.E−88	WP_007978213.1	7.E−114	WP_007979503.1	2.E−06	Yes
Rhodobacterales bacterium HTCC2255	alphaproteobacteria	+	3	WP_008034433	1.E−78	WP_008035279	2.E−80	WP_008035279	1.E−78	WP_008035491	5.E−103	WP_008035491	6.E−10	Yes
Rhodoferax ferrireducens T118	betaproteobacteria	+	1	YP_524469	3.E−102	YP_522163	6.E−22	YP_522163	9.E−24	YP_524469	4.E−19	YP_524469	1.E−09	Yes
Rhodospirillum rubrum ATCC 11170	alphaproteobacteria	+	1	YP_427157	2.E−96	YP_426633	9.E−06	YP_426633	3.E−15	YP_427958	8.E−13	YP_427157	2.E−08	Yes
Sinorhizobium fredii HH103	alphaproteobacteria	+ +	6	YP_005193057.1 YP_005189847.1	5.E−88 2.E−83	YP_005189397.1	6.E−96	YP_005193188.1	1.E−25	YP_005189398 YP_005189747.1	5.E−100 4.E−37	YP_005193188.1	7.E−17	Yes
Thermus thermophilus HB8	deinococcus−thermus	+	1	YP_144802.1	7.E−19	YP_144802.1	6.E−07	YP_144802.1	1.E−13	YP_144802.1	7.E−16	YP_144804.1	6.E−07	Yes
Variovorax paradoxus S110	betaproteobacteria	+	2	WP_020722958.1	3.E−92	YP_002942232	5.E−25	YP_002942232	6.E−25	WP_018905950	2.E−17	WP_019652886	5.E−41	Yes
Acanthamoeba castellanii	amoebozoa	+	3	XP_004356483	5.E−93	XP_004353662	5.E−107	XP_004353662	1.E−89	XP_004353662	1.E−66	XP_004352808	3.E−42	Yes
Alexandrium tamarense	alveolata	+ + +	15	GAJG01047488.1 GAJB01023065.1 GAIT01017835.1	3.E−98 7.E−97 7.E−96	GAJB01003346.1	1.E−72	GAJG01065609.1 GAIT01038316.1 GAJQ01013888.1 GAJB01002800.1 GAJB01003346.1 GAJG01084105.1 GAJB01002451.1	4.E−146 5.E−145 6.E−145 3.E−127 6.E−89 3.E−86 3.E−43	GAIQ01044360.1 GAJB01002837.1 GAIT01027724.1 GAJG01019075.1	2.E−101 2.E−101 2.E−100 3.E−100	GAIT01050926.1	1.E−15	No
Amphimedon queenslandica	metazoa	+	5	XP_003387896	4.E−117	XP_003390674.1	2.E−57	XP_003390674.1	7.E−60	XP_003390674.1 XP_003391032.1	3E−85 4.E−64	XP_003387790.1 XP_003387789.1	6.E−94 1.E−87	Yes
Anolis carolinensis	metazoa	+	4	XP_003225770	0.E+00	XP_003229415	0.E+00	XP_003228894	0.E+00	XP_003216348	0.E+00	XP_003229415	7.E−11	Yes
Aplysia californica	metazoa	+	3	XP_005109016	6.E−143	XP_005096338.1	1.E−113	XP_005096338.1	4.E−80	XP_005096338.1	1.E−49	XP_005111516.1	7.E−84	Yes
Arabidopsis thaliana	viridiplantae		1	NP_201530	8.E−05	NP_201530	3.E−07	NP_199655	4.E−04	NP_201530	6.E−04	NP_180034	2.E−69	Yes
Aspergillus flavus	fungi	+	3	jgi|Aspfl1|32985	4.E−61	jgi|Aspfl1|32600	4.E−34	jgi|Aspfl1|32600	1.E−36	jgi|Aspfl1|32600	2.E−39	jgi|Aspfl1|37720	3.E−54	Yes
Bigellowiella natans	rhizaria	+	3	jgi|Bigna1|38746	5.E−93	jgi|Bigna1|143381	2.E−48	jgi|Bigna1|143381	4.E−116	jgi|Bigna1|143381	2.E−45	jgi|Bigna1|71202	1.E−19	Yes
Bodo saltans	euglenozoa	+	1	BS70775.1	1.E−22	BS45215.1	3.E−02	BS52240.1	1.E−04	BS70775.1	1.E−05	BS32415.1	3.E−01	Yes
Branchiostoma floridae	metazoa	+	5	XP_002596418	4.E−164	XP_002612437	3.E−137	XP_002594859	0.E+00	XP_002609761	0.E+00	XP_002611829	1.E−127	Yes
Caenorhabditis elegans	metazoa	+	4	M04B2.4	2.E−99	Y106G6H.5	2.E−47	Y37E3.17a	1.E−35	Y37E3.17a	3.E−71	C15B12.1	2.E−64	Yes
Capsaspora owczarzaki	ichthyosporea	+	3	XP_004345258.1	9.E−90	XP_004343467	2.E−117	XP_004343467	2.E−78	XP_004343467	7.E−52	XP_004349776	4.E−61	Yes
Chlamydomonas reinhardtii CC−1373	viridiplantae		0	XP_001692123.1	2.E−07	XP_001692123.1	2.E−02	XP_001692442.1	2.E−03	XP_001692442.1	3.E−04	XP_001700567.1	2.E−32	Yes
Ciona savignyi	metazoa	+	3	ENSCSAVP00000011427	6.E−78	ENSCSAVP00000009849	8.E−81	ENSCSAVP00000009849	4.E−194	ENSCSAVP00000009849	5.E−74	ENSCSAVP00000003434	3.E−46	Yes
Drosophila melanogaster	metazoa	+ +	4	NP_610228 NP_536791	2.E−127 5.E−102	NP_572162	4.E−83	NP_611263	3.E−156	NP_611263	7.E−61	NP_611263	2.E−09	Yes
Emiliania huxleyi CCMP1516	haptophyceae	+	5	XP_005760525.1	2.E−46	XP_005785242.1	1.E−74	XP_005791899.1 XP_005785242.1	2.E−99 4.E−87	XP_005779402.1	1.E−78	XP_005778512.1	2.E−36	Yes
Fusarium oxysporum 4287	fungi	+	3	FOXG_09711T0	0.E+00	FOXG_12340T0	2.E−34	FOXG_12340T0	9.E−37	FOXG_12340T0	1.E−41	FOXG_01881T0	1.E−52	Yes
Hydra vulgaris	metazoa	+	3	XP_002155683	3.E−72	XP_002158472	1.E−65	XP_002158472	0.E+00	XP_002156170	2.E−172	XP_002156170	6.E−08	Yes
Monosiga brevicollis MX1	choanoflagellida	+	1	XP_001742063	5.E−65	XP_001742226	3.E−07	XP_001747302	5.E−07	XP_001750081	4.E−02	XP_001747302	2.E−10	Yes
Nematostella vectensis	metazoa	+	6	XP_001619878	3.E−116	XP_001638535 XP_001633663	4E−120 1E−100	XP_001624293	0.E+00	XP_001632395	0.E+00	XP_001636208	2.E−118	Yes
Neurospora crassa OR74a	fungi		0	XP_960462.1	2.E−01	XP_965711.1	1.E−02	XP_960462.1	1.E−05	XP_960462.1	4.E−07	XP_960220.1	1.E−08	Yes
Physcomitrella patens	viridiplantae		1	XP_001776959	3.E−06	XP_001776959	2.E−03	XP_001761729	4.E−05	XP_001776959	3.E−10	XP_001776959	3.E−81	Yes
Phytophthora sojae	stramenopile	+	2	AAQY02000045.1	5.E−106	AAQY02000045	9.E−07	AAQY02000045.1	2.E−07	AAQY02000099	2.E−05	AAQY02000099	2.E−50	Yes
Saccoglossus kowalevskii	metazoa	+ +	11	XP_002734363 XP_002734362.1	2.E−160 5.E−153	XP_002740757.1 XP_002737967.1	3.E−116 6.E−99	NP_001171860.1	0.E+00	XP_002737399.1 XP_002737400.1	0.E+00 4.E−165	XP_002732031.1 XP_002741300.1 XP_002737047.1 XP_002737054.1	1.E−56 2.E−28 7.E−27 1.E−26	Yes
Salpingoeca rosetta	choanoflagellida	+	1	XP_004999027	4.E−23	XP_004999027	2.E−02	XP_004999027	4.E−12	XP_004999027	6.E−05	XP_004999027	2.E−12	Yes
Schistosoma mansoni	metazoa	+	1	XP_002580007	3.E−96	XP_002580007	2.E−04	XP_002571715.1	3.E−03	XP_002580007	2.E−03	XP_002580007	8.E−03	Yes
Strongylocentrotus purpuratus	metazoa	+ +	7	XP_784019 XP_792444.3	2.E−153 4.E−37	XP_786380.2	6.E−115	XP_003727333.1	0.E+00	XP_792268.3	0.E+00	XP_797352.2 XP_796417.3	3.E−107 2.E−120	Yes
Symbiodinium sp. A1	alveolata	+	4	GAKY01118127.1	1.E−92	GAKY01036249.1	1.E−50	GAKY01036249.1	1.E−144	GAKY01085942.1	1.E−88	GAKY01056787.1	2.E−15	No
Tetranychus urticae	metazoa	+ +	2	tetur14g02620 tetur03g07760	1.E−93 1.E−90	tetur14g02620	8.E−07	tetur14g02620	5.E−08	tetur14g02620	4.E−07	tetur14g02620	2.E−04	Yes
Thalassiosira pseudonana	stramenopile	+	3	XP_002294365	2.E−90	XP_002290388	1.E−57	XP_002290388	2.E−83	XP_002290388	4.E−55	XP_002297081	5.E−30	Yes
Thecamonas trahens ATCC 50062	apusozoa	+ +	4	AMSG_02503T0 AMSG_03783T0	4.E−80 2.E−08	AMSG_02498T0	1.E−60	AMSG_02498T0	2.E−64	AMSG_02498T0	2.E−58	AMSG_03839T0	5.E−39	Yes
Trichoplax adhaerens	metazoa	+ +	4	XP_002108741 XP_002114542	7.E−132 2.E−96	XP_002110013	5.E−67	XP_002110013	0.E+00	XP_002110817	3.E−177	XP_002118575	6.E−20	Yes
Yarrowia lipolytica	fungi		0	XP_500543	1.E−03	XP_502399	3.E−08	XP_500862	8.E−04	XP_500862	1.E−02	XP_503994	3.E−01	Yes
Blastocystis hominis, Singapore isolate B	stramenopile		0	CABX01000147.1	1.E+00	CABX01000021.1	1.E+00	CABX01000023.1	3.E+00	CABX01000023.1	3.E−01	CABX01000089.1	9.E−01	Yes
Cryptosporidium parvum	alveolata		0	XP_627944.1	3.E−01	XP_625499.1	5.E−01	XP_628285.1	2.E−02	XP_625499.1	5.E−02	XP_627167.1	2.E−03	No
Encephalitozoon cuniculi	fungi		0	NP_586202.1	2.E−01	NP_586202.1	3.E−01	NP_586250.1	5.E−01	NP_586202.1	1.E+00	NP_597140.1	1.E−01	No
Entamoeba histolytica HM−1:IMSS	amoebozoa		0	XP_653096.1	2.E−01	XP_001913743.1	2.E−11	XP_649611.2	1.E−07	XP_001913743.1	4.E−09	XP_654048.1	5.E−01	No
Giardia lamblia	fornicata		0	XP_001705832	4.E−01	XP_001707988	5.E−14	XP_001707418	4.E−01	XP_001707988	3.E−05	XP_001704370	2.E−01	No
Mastigamoeba balamuthi	amoebozoa		0	EC698218.1	3.E+00	EC710463.1	1.E+01	EC703826.1	3.E+00	EC694392.1	5.E+00	EC708958.1	2.E+00	No
Mikrocytos mackini a4303	rhizaria		0	GAHX01000842.1	5.E+00	GAHX01000776.1	7.E+00	GAHX01003908.1	5.E−01	GAHX01004810.1	7.E+00	GAHX01001339.1	5.E−01	No
Nosema ceranae BRL01	fungi		0	XP_002995067.1	7.E−01	XP_002996277.1	9.E−03	XP_002996277.1	5.E−03	XP_002996277.1	2.E+00	XP_002996285.1	1.E+00	No
Nyctotherus ovalis	alveolata		0	AM896437.1	6.E+00	AM899062.1	1.E+01	AM897036.1	5.E+00	AM896888.1	2.E+00	AM898669.1	6.E+00	Yes
Piromyces sp. E2	fungi		0	GT914446.1	2.E+00	GT914247.1	4.E−03	GT904155.1	6.E−01	GT904929.1	2.E+00	GT900195.1	3.E−01	No
Sawyeria marylandensis	heterolobosea		0	EC819336.1	2.E+00	EC822702.1	6.E−01	EC825062.1	1.E+00	EC822693.1	4.E−12	EC822689.1	5.E−01	No
Trichomonas vaginalis	parabasilia		0	XP_001319204	5.E+00	XP_001305657	2.E+00	XP_001311329	1.E+00	XP_001315398	2.E+00	XP_001308456	1.E+01	No

Organisms in bold contain a FOXRED1

Accession is most similar to query in this column

Archaea

Bacteria

Eukaryote

FOXRED present, complex I absent

FOXRED absent, complex I present

Amitochondriate organism

(Fig. 4). These include alphaproteobacteria, such as *Roseobacter*, betaproteobacteria, such as *Burkholderia*, and gammaproteobacteria, such as *Pantoea* and *Glaciecola* and *Acidomicrobidae*. In all of these gene arrangements, the *FOXRED1* gene is downstream of the *N*-methylhydantoinase A and B genes. *N*-methylhydantoinase (EC 3.5.2.14) catalyzes the ATP-dependent hydrolysis of *N*-methylhydantoin

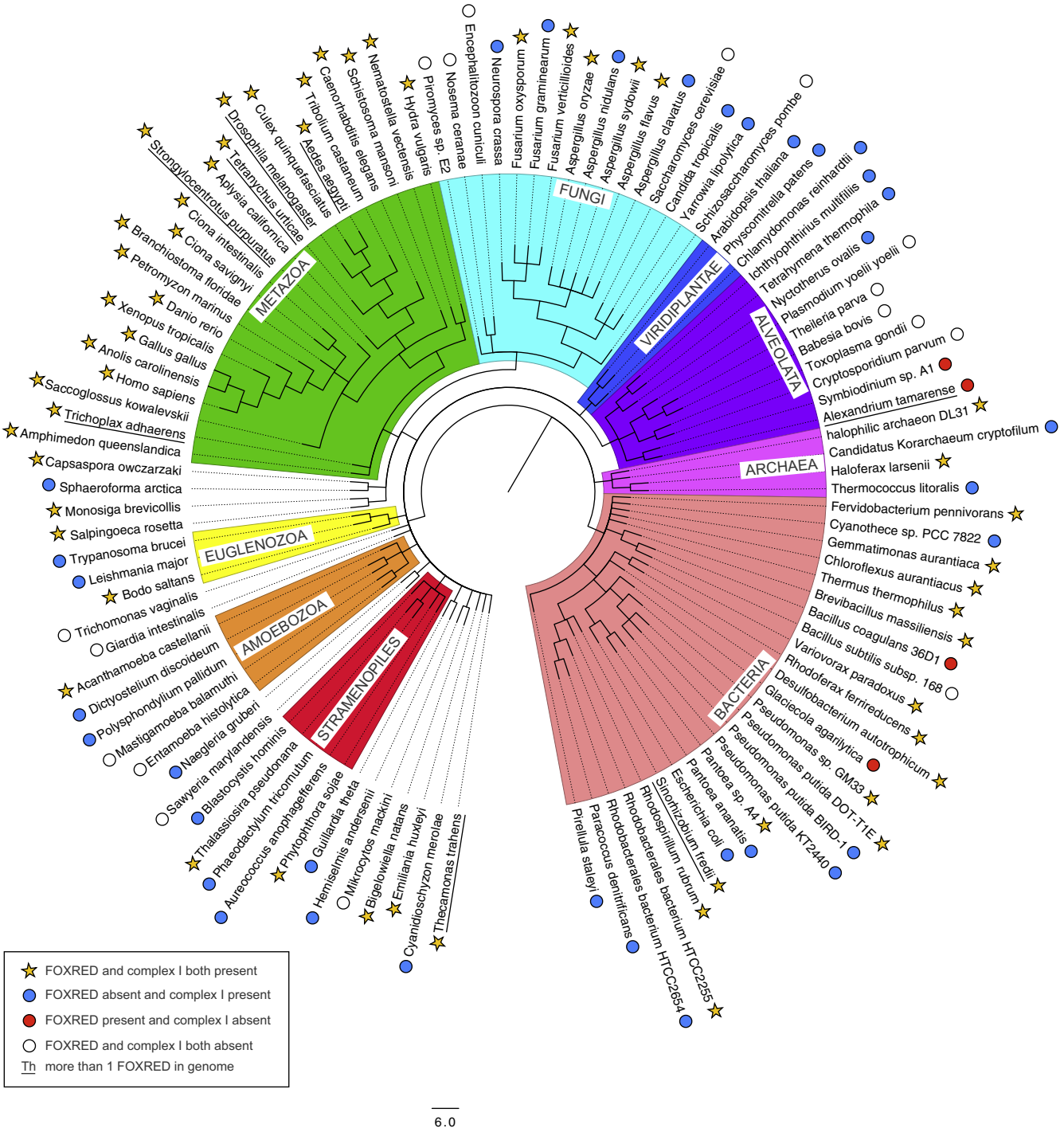

Fig. 3. Phylogenetic distribution of selected organisms and the presence of a FOXRED1 ortholog and of complex I. NCBI taxonomic identification codes and NCBI Common Tree were used to produce a tree for the organisms in Table 1. Stars, organisms with both a FOXRED1 ortholog and complex I; blue circle, organisms without a FOXRED1 ortholog but with complex I; red circle, organisms with a FOXRED1 ortholog but no complex I; open circles, organisms with neither FOXRED1 ortholog or complex I. Organisms with more than one FOXRED1 ortholog are underlined.

to *N*-carbamoylsarcosine in the creatinine degradation pathway (Fig. 5). The *N*-carbamoylsarcosine can be converted to sarcosine, a substrate for SARDH or DMGDH. The presence of *N*-methylhydantoinase allows organisms to grow on creatinine and creatine as sole carbon or nitrogen sources [42]. *N*-methylhydantoinase also functions in the metabolism of arginine and proline; it is closely related to human 5-oxoprolinase (EC 3.5.2.9), which catalyzes the ATP-dependent hydrolysis of 5-oxoproline to L-glutamate in glutathione synthesis. Interestingly, mammalian PIPOX has wide substrate specificity can utilize larger molecules such as L-proline as a substrate [43]. The presence of *FOXRED1* orthologs in similar gene arrangements in diverse bacterial species suggests that horizontal gene transfer has occurred. In addition, the extremely limited distribution of *FOXRED1* orthologs amongst proteobacteria and in particular, *Enterobacteriaceae* is also suggestive of horizontal gene transfer (Fig. S7).

4. Conclusions

In this study, I investigated the phylogenetic and structural relationships of FOXRED1. If FOXRED1 and complex I had been functionally-interacting proteins throughout evolution, they would have co-evolved

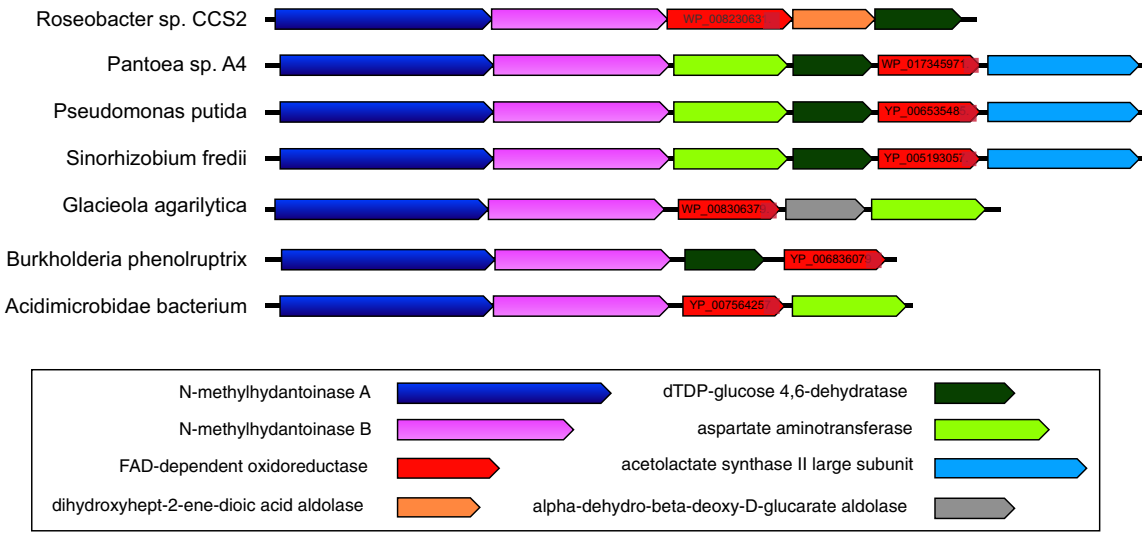

Fig. 4. Diagram of *FOXRED1* ortholog-containing operons. The accession numbers of the nucleotide sequences used to create this diagram are: *Roseobacter* (AAYB01000001.1), *Pantoea* (NZ_ALXE01000017.1), *Pseudomonas* (CP003734.1), *Sinorhizobium* (HE616899.1), *Glaciecola* (BAEK01000084.1), *Burkholderia* (NC_018696.1), *Acidimicrobidae* (NC_020520.1). The accession numbers of the respective FOXRED1 proteins are indicated in the red boxes.

and would have orthologs in the same sets of organisms. This is not the case (Fig. 3). I suggest that *FOXRED1* does not have a complex I-related function in some organisms, such as *Glaciecola agarilytica NO2*. Phylogenetic analysis robustly places the *Glaciecola* FOXRED1 amongst organisms with a complex I, indicating that it shares the greatest sequence conservation with the FOXREDs of those organisms. (Fig. 2, S5B). *Roseobacter sp. CCS2* has a FOXRED1 that is very closely related to the *Glaciecola* protein, but other species of *Roseobacter*, such as *Roseobacter sp. AzwK-3b* do not have a FOXRED; yet both *Roseobacter* species have a complex I. Similarly, *Rhodobacterales bacterium HTCC2255* and *Glaciecola* have closely related FOXREDs; *Rhodobacterales bacterium HTCC2255* has a complex I, but not *Rhodobacterales bacterium HTCC2654* (Fig. 3). The most parsimonious explanation for these observations is that the *Roseobacter* and *Rhodobacterales* FOXREDs are not needed for complex I assembly, as assembly occurs in the absence of a *FOXRED1* ortholog in closely related species.

The firmicutes *Bacillus coagulans 36D1* and *Brevibacillus massiliensis* as well as the archaea *Haloferax larsenii* are organisms with FOXREDs with closely related species that do not. However, the argument that their FOXREDs are not involved in complex I assembly is less compelling because these proteins are phylogenetically more distinct (Figs. S2, S3).

If complex I assembly can occur without FOXRED1 in some organisms, then FOXRED1's complex I-related function is a more recently acquired one. That a *FOXRED1* appears to be present in all metazoan genomes suggests that it plays an essential role in this kingdom; I hypothesize that this essential role is complex I assembly.

FOXRED1 is a member of the D-amino acid oxidases, most closely related to *N*-methyl amino acid dehydrogenases. *FOXRED1* sequence conservation and gene context suggest that a role in amino acid metabolism should be considered. In some proteobacteria, *FOXRED1* is located in an operon, downstream of *N*-methylhydantoinase genes (Fig. 4). *N*-methylhydantoinase functions in creatine and creatinine degradation, producing glycine, much like human SARDH and DMGDH function in choline degradation to produce glycine. Glycine is not considered a dietary essential amino acid in humans because pathways for its synthesis (for example from choline) exist. However, glycine needs are high, as

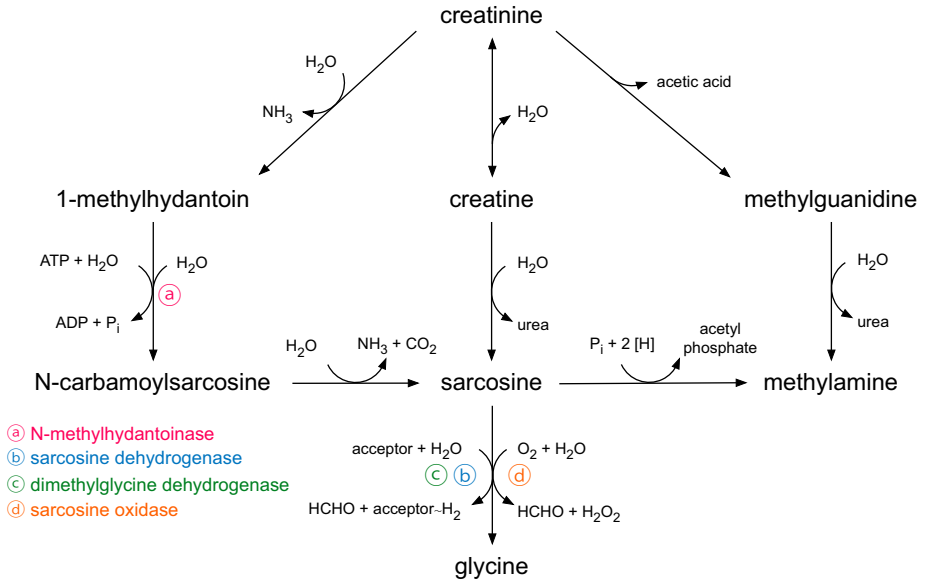

Fig. 5. Creatinine and creatine degradation. *N*-methylhydantoinase is composed of 2 subunits, A and B. DMGDH is capable of utilizing both dimethylglycine and sarcosine as substrates.

this amino acid accounts for 11.5% of total amino acids and 20% of amino acid nitrogen [44]. Glycine is a constituent of glutathione, a tripeptide redox buffer that is the most abundant intracellular, non-protein molecule. It is worth noting that mitochondrial diseases, such as Leigh syndrome, are very often conditions of oxidative stress [14]. Perhaps FOXRED1 has a role in glutathione metabolism and the protection of complex I from oxidative stress. I will address the details of this hypothesis elsewhere.

Acknowledgments

I would like to thank Jeremy Wideman and Frank Nargang for critically reading the manuscript. I would particularly like to thank Jeremy Wideman and Bart Hazes for their support and advice. This work was supported by a National Science and Engineering Research Council Discovery Grant (NSERC 386351-2012).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbabio.2015.01.014.

References

[1] T. Gabaldón, Evolution of proteins and proteomes: a phylogenetics approach, Evol. Bioinformatics Online 1 (2005) 51–61.
[2] E.V. Koonin, Orthologs, paralogs, and evolutionary genomics, Annu. Rev. Genet. 39 (2005) 309–338.
[3] M. Pellegrini, E.M. Marcotte, M.J. Thompson, D. Eisenberg, T.O. Yeates, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 4285–4288.
[4] R.G. Efremov, L.A. Sazanov, Respiratory complex I: 'steam engine' of the cell? Curr. Opin. Struct. Biol. 21 (2011) 532–540.
[5] B. Andrews, J. Carroll, S. Ding, I.M. Fearnley, J.E. Walker, Assembly factors for the membrane arm of human complex I, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 18934–18939.
[6] K.R. Vinothkumar, J. Zhu, J. Hirst, Architecture of mammalian respiratory complex I, Nature 515 (2014) 80–84.
[7] V.K. Moparthi, C. Hägerhall, The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits, J. Mol. Evol. 72 (2011) 484–497.
[8] R.G. Efremov, R. Baradaran, L.A. Sazanov, The architecture of respiratory complex I, Nature 465 (2010) 441–445.
[9] R.G. Efremov, L.A. Sazanov, Structure of the membrane domain of respiratory complex I, Nature 476 (2011) 414–420.
[10] C. Hunte, V. Zickermann, U. Brandt, Functional modules and structural basis of conformational coupling in mitochondrial complex I, Science 329 (2010) 448–451.
[11] T. Gabaldón, D. Rainey, M.A. Huynen, Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (complex I), J. Mol. Biol. 348 (2005) 857–870.
[12] P. Cardol, Mitochondrial NADH:ubiquinone oxidoreductase (complex I) in eukaryotes: a highly conserved subunit composition highlighted by mining of protein databases, Biochim. Biophys. Acta 1807 (2011) 1390–1397.
[13] R.J. Janssen, L.G. Nijtmans, L.P. Heuvel, J.A. Smeitink, Mitochondrial complex I: structure, function and pathology, J. Inherit. Metab. Dis. 29 (2006) 499–515.
[14] F. Valsecchi, W.J. Koopman, G.R. Manjeri, R.J. Rodenburg, J.A. Smeitink, P.H. Willems, Complex I disorders: causes, mechanisms, and development of treatment strategies at the cellular level, Dev. Disabil. Res. Rev. 16 (2010) 175–182.
[15] V.F. Rhein, J. Carroll, S. Ding, I.M. Fearnley, J.E. Walker, NDUFAF7 methylates arginine 85 in the NDUFS2 subunit of human complex I, J. Biol. Chem. 288 (2013) 33016–33026.
[16] E. Fassone, A.J. Duncan, J.W. Taanman, A.T. Pagnamenta, M.I. Sadowski, T. Holand, W. Qasim, P. Rutland, S.E. Calvo, V.K. Mootha, M. Bitner-Glindzicz, S. Rahman, FOXRED1, encoding an FAD-dependent oxidoreductase complex-I-specific molecular chaperone, is mutated in infantile-onset mitochondrial encephalopathy, Hum. Mol. Genet. 19 (2010) 4837–4847.
[17] S.E. Calvo, E.J. Tucker, A.G. Compton, D.M. Kirby, G. Crawford, N.P. Burtt, M. Rivas, C. Guiducci, D.L. Bruno, O.A. Goldberger, M.C. Redman, E. Wiltshire, C.J. Wilson, D. Altshuler, S.B. Gabriel, M.J. Daly, D.R. Thorburn, V.K. Mootha, High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency, Nat. Genet. 42 (2010) 851–858.
[18] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[19] L.Y. Geer, A. Marchler-Bauer, R.C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, S.H. Bryant, The NCBI BioSystems database, Nucleic Acids Res. 38 (2010) D492–D496.
[20] I.V. Grigoriev, H. Nordberg, I. Shabalov, A. Aerts, M. Cantor, D. Goodstein, A. Kuo, S. Minovitsky, R. Nikitin, R.A. Ohm, R. Otillar, A. Poliakov, I. Ratnere, R. Riley, T. Smirnova, D. Rokhsar, I. Dubchak, The genome portal of the Department of Energy Joint Genome Institute, Nucleic Acids Res. 40 (2012) D26–D32.
[21] P. Flicek, M.R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C.G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A.K. Kähäri, S. Keenan, E. Kulesha, F.J. Martin, T. Maurel, W.M. McLaren, D.N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H.S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S.J. Trevanion, A. Vullo, S.P. Wilder, M. Wilson, A. Zadissa, B.L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T.J. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D.R. Zerbino, S.M. Searle, Ensembl 2014, Nucleic Acids Res. 42 (2014) D749–D755.
[22] C.A. Cuomo, B.W. Birren, The fungal genome initiative and lessons learned from genome sequencing, Methods Enzymol. 470 (2010) 833–855.
[23] B. Hazes, CDSbank: taxonomy-aware extraction, selection, renaming and formatting of protein-coding DNA or amino acid sequences, BMC Bioinforma. 15 (2014) 61.
[24] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.
[25] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[26] S. Guindon, F. Lethiec, P. Duroux, O. Gascuel, PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference, Nucleic Acids Res. 33 (2005) W557–W559.
[27] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[28] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview Version 2–a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191.
[29] O. Penn, E. Privman, H. Ashkenazy, G. Landan, D. Graur, T. Pupko, GUIDANCE: a web server for assessing alignment confidence scores, Nucleic Acids Res. 38 (2010) W23–W28.
[30] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.M. Claverie, O. Gascuel, Phylogeny.fr: robust phylogenetic analysis for the non-specialist, Nucleic Acids Res. 36 (2008) W465–W469.
[31] A. Boc, A.B. Diallo, V. Makarenkov, T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks, Nucleic Acids Res. 40 (2012) W573–W579.
[32] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, Proc. Gateway Computing Environments Workshop, 2010, pp. 1–8.
[33] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucleic Acids Res. 42 (2014) D222–D230.
[34] J.E. Lawson, S.H. Park, A.R. Mattison, J. Yan, L.J. Reed, Cloning, expression, and properties of the regulatory subunit of bovine pyruvate dehydrogenase phosphatase, J. Biol. Chem. 272 (1997) 31625–31629.
[35] R. Douce, J. Bourguignon, M. Neuburger, F. Rebeille, The glycine decarboxylase system: a fascinating complex, Trends Plant Sci. 6 (2001) 167–176.
[36] D. Leys, J. Basran, N.S. Scrutton, Channelling and formation of 'active' formaldehyde in dimethylglycine oxidase, EMBO J. 22 (2003) 4038–4048.
[37] T. Makiuchi, T. Nozaki, Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa, Biochimie 100 (2014) 3–17.
[38] R.M. de Graaf, G. Ricard, T.A. van Alen, I. Duarte, B.E. Dutilh, C. Burgtorf, J.W. Kuiper, G.W. van der Staay, A.G. Tielens, M.A. Huynen, J.H. Hackstein, The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*, Mol. Biol. Evol. 28 (2011) 2379–2391.
[39] I. Wawrzyniak, M. Roussel, M. Diogon, A. Couloux, C. Texier, K.S. Tan, C.P. Vivarès, F. Delbac, P. Wincker, H. El Alaoui, Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*, Int. J. Parasitol. 38 (2008) 1377–1382.
[40] F. Denoeud, M. Roussel, B. Noel, I. Wawrzyniak, C. Da Silva, M. Diogon, E. Viscogliosi, C. Brochier-Armanet, A. Couloux, J. Poulain, B. Segurens, V. Anthouard, C. Texier, N. Blot, P. Poirier, G.C. Ng, K.S. Tan, F. Artiguenave, O. Jaillon, J.M. Aury, F. Delbac, P. Wincker, C.P. Vivarès, H. El Alaoui, Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite, Genome Biol. 12 (2011) R29.
[41] F. Mao, P. Dam, J. Chou, V. Olman, Y. Xu, DOOR: a database for prokaryotic operons, Nucleic Acids Res. 37 (2009) D459–D463.
[42] M. Wyss, R. Kaddurah-Daouk, Creatine and creatinine metabolism, Physiol. Rev. 80 (2000) 1107–1213.
[43] B.E. Reuber, C. Karl, S.A. Reimann, S.J. Mihalik, G. Dodt, Cloning and functional expression of a mammalian gene for a peroxisomal sarcosine oxidase, J. Biol. Chem. 272 (1997) 6766–6776.
[44] W. Wang, Z. Wu, Z. Dai, Y. Yang, J. Wang, G. Wu, Glycine metabolism in animals and humans: implications for nutrition and health, Amino Acids 45 (2013) 463–477.