Aug 5, 2021 rmk

Fox Cluster Determinants for Iron Biooxidation in the Extremely Thermoacidophilic Sulfolobaceae

James A. Counts, Nicholas P. Vitko, and Robert M. Kelly*

[1]*Department of Chemical and Biomolecular Engineering*

North Carolina State University, Raleigh, NC 27695-7905

Submitted to: Environmental Microbiology (March 2021)

Revision to EMI-2021-0402 submitted (August, 2021)

Running title: *Iron Biooxidation Genes in Extremophilic Sulfolobaceae*

Keywords: *Extreme thermoacidophiles, archaea, Sulfolobales, biooxidation, respiration*

Address correspondence to:* ***Robert M. Kelly*

Department of Chemical and Biomolecular Engineering

North Carolina State University

EB-1, 911 Partners Way

Raleigh, NC 27695-7905

Phone: (919) 515-6396

Fax: (919) 515-3465

Email: rmkelly@ncsu.edu

Summary

Within the extremely thermoacidophilic *Sulfolobaceae*, the capacity to oxidize iron varies considerably. While some species are prolific iron oxidizers (e.g., *Metallosphaera sedula*), other species do not oxidize iron at all (e.g., *Sulfolobus acidocaldarius*). Iron oxidation capacity maps to a genomic locus, referred to previously as the '*Fox* Cluster', that encodes putative proteins that are mostly unique to the *Sulfolobaceae*. The role of putative proteins in the *Fox* Cluster has not been confirmed, but proteomic analysis here of iron-oxidizing membranes from *M. sedula* indicates that FoxA2 and FoxB (both cytochrome c oxidase-like subunits) and FoxC (CbsA/cytochrome b domain-containing) are essential. Further, comparative genomics (locus organization and gene disruptions) and transcriptomics (polarity effects and differential expression) connects these genomic determinants with disparate iron biooxidation and respiration measurements among *Sulfolobaceae* species. While numerous homologous proteins can be identified for FoxA in genome databases (COX-like domains are prevalent across all domains of life), few homologs exist for FoxC or for most other *Fox* Cluster proteins. Phylogenetic reconstructions suggest this locus may have existed in early *Sulfolobaceae,* while the only other close homologs to the locus appear in the recently discovered candidate phylum *Marsarchaota*.

Introduction

Iron is the most ubiquitous metal element associated with life, present in almost every known organism, serving as a crucial intermediary for electron transfer that drives many enzymatic reactions, with the exception of some *Lactobacillus* spp., *e.g. Lactobacillus plantarum* which appears to thrive in an iron-rich environment where iron-mediated metabolisms produce copious Fenton-related free radicals (Archibald, 1983). Prior to the first rapid oxygenation of the Earth's atmosphere, iron was readily available in an aqueous state, likely leading to a large selective pressure for incorporating the multivalent metal within enzymatic centers across almost all domains of life (Ilbert and Bonnefoy, 2013). After oxygenation, iron availability became a limiting factor in many biotopes, leading to a plethora of scavenging mechanisms, including siderophores, ferritins, and reductase complexes (Sandy and Butler, 2009). In fact, its scarcity limits microbial activity in the carbon-fixing oceans (Tagliabue et al., 2017) and pathogenesis in the human body (Cassat and Skaar, 2013), sometimes encouraging risky evolutionary behavior (Bonnain et al., 2016). Where iron is not a limiting factor, there are examples of microbes subsisting on a spartan metabolism oxidizing ferrous iron.

Although there are examples of circumneutral pH iron metabolizers (Emerson et al., 2010; Hedrich et al., 2011), highly acidic conditions favor iron biooxidation-based metabolisms because of the thermodynamic stability of soluble Fe^{3+} (strong oxidant) species at lower pH, which is roughly nine orders of magnitude higher than at circumneutral conditions (Johnson et al., 2012). In particular, some polyanionic complexes are mediators of this increased solubility, such as sulfate, which readily forms complex ions with ferric iron (Welham et al., 2000). As a result, iron cycling plays an important role in microbial habitats, ranging from terrestrial solfataras to acidic waterways to iron mats (Gonzalez-Toril et al., 2003; Kozubal et al., 2012), especially in the presence of oxygen (Johnson et al., 2012; Ilbert and Bonnefoy, 2013). Thus, many extreme acidophiles (pH_{opt} < 3.0) are capable of oxidizing ferrous iron, funneling electrons into reducing or oxidizing reactions to meet bioenergetic needs (Appia-Ayme et al., 1999; Johnson et al., 2012).

However, the low electronic potential difference between Fe^{3+}/Fe^{2+} (+0.77 V) and O_2/H_2O (+1.12 V) (Ilbert and Bonnefoy, 2013), even at low pH, requires the reaction to occur largely under acidic extracellular conditions, where soluble iron is plentiful and easily turned over, as opposed to the circumneutral cytoplasmic space. Furthermore, this kind of harsh environmental niche limits life to a select group of microorganisms with an inverted membrane potential (in comparison to neutrophiles) and numerous other acid tolerance traits, as well as cellular features that protect against environments laden with heavy metals and sulfidic ores (Baker-Austin and Dopson, 2007; Wheaton et al., 2015). Many such organisms are chemolithoautotrophs, with limited capacity to utilize sugars or other complex carbohydrates that are inherently scarce in hot, acidic environments where these organisms are typically the primary producers.

The most thoroughly studied mechanism for iron biooxidation belongs to the mesoacidophilic bacterium, *Acidithiobacillus ferrooxidans* (f. *Thiobacillus ferrooxidans*), whose cytochrome c proteins (that oxidize iron) are localized to its outer membrane (Appia-Ayme et al., 1999), where pH is at its minimum. The liberated iron electron is passed to a periplasm-spanning copper oxidase (Bengrine et al., 1998), shuttling it either to a proton-importing NADH dehydrogenase (uphill) or to a proton-exporting terminal oxidase (downhill), both spanning the inner membrane (Bonnefoy and Holmes, 2012). The interplay between these two outcomes is crucial to maintaining the circumneutral cytosol, balancing proton importing activity from ATP synthase and NADH dehydrogenase (proton motive force generation is probably irrelevant) with proton export of the terminal oxidase (Ingledew, 1982; Baker-Austin and Dopson, 2007).

Mesoacidophilic, iron-oxidizing bacteria have been examined in some detail, but relatively little is known about iron oxidation in acidophilic archaeal species, despite the fact that iron oxidizers have been identified in the phyla *Euryarchaeota* and *Crenarchaeota*. One study in the mesophilic euryarchaeon *Ferroplasma acidiphilum* identified an iron-oxidizing complex exhibiting spectroscopic signatures and proteomic evidence for a type-aa_3 oxidase as well as a sulfocyanin (Castelle et al., 2015), containing copper with high oxidation potential, hypothesized as a similar

mechanism to *A. ferrooxidans* (Appia-Ayme et al., 1999). However, some key differences exist, such as a lack of cytochrome c, as well as the absence of periplasmic space present in bacteria, where bifurcation hypothetically takes place (Wheaton et al., 2015).

In the case of extremely thermoacidophilic archaea from the phylum *Crenarchaeota* (growth T_{opt} > 65 ˚C), a number of metal-mobilizing species have been isolated (11 genome-sequenced strains from the family *Sulfolobaceae* are considered here, 7 of which were examined to reinforce phenotypical analysis) (Lewis et al., 2021). Within this group, iron oxidation occurs to varying extents for reasons that were initially unclear. Thus far, transcriptomic studies with three of these archaea, *Sulfuracidifex metallicus* (f. *Sulfolobus metallicus*) (Bathe and Norris, 2007), *Metallosphaera sedula* (Auernik and Kelly, 2008; Auernik and Kelly, 2010b), and *Metallosphaera yellowstonensis*, as well as a metatranscriptomic analysis of iron mats in Yellowstone National Park (Norris Geyser Basin) (Kozubal et al., 2011), point to a group of iron-responsive, co-located genes, dubbed the 'Fox Cluster', although biochemical verification of the function of the encoded proteins has not been reported and some inconsistencies have existed to-date. Of interest here, were the genomic determinants of iron oxidation capacity among the *Sulfolobaceae*, whether and how iron oxidation proceeds, and how environmental adaptations arise in extreme thermoacidophiles to support their lifestyles in hot acid.

Materials and Methods

Cultivation procedures

All organisms were cultured aerobically with agitation in a species-modified DSM-88 Medium. All cultures were grown at 70˚C, except *S. metallicus* (65˚C), *Sulfolobus acidocaldarius* and *Sulfurisphaera tokodaii* (75 ˚C), and *Sulfurisphaera ohwakuensis* (85˚C). All cultures were grown at pH 2.0, except *S. tokodaii, Metallosphaera hakonensis, S. acidocaldarius* (all at pH 3.0), and *Acidianus sulfidivorans* (pH 1.0). Species-specific media supplementation included: 0.2 g/L yeast extract (YE) and 10 g/L elemental sulfur (S^0) (*Acidianus brierleyi*), 0.1 g/L YE and 10 g/L

pyrite (*A. sulfidivorans*), 1 g/L YE (*M. hakonensis*, *Metallosphaera prunae, Metallosphaera sedula, S. ohwakuensis*), 1 g/L NZ-amine, and 2 g/L sucrose (*S. acidocaldarius*), 0.2 g/L YE and 5 g/L S^0 (*S. metallicus*), 1 g/L each YE, D-glucose, casamino acids (*S. tokodaii*). For all cultures except those used for membrane proteomics, organisms were grown in 30 mL of culture volume within glass serum bottles vented with foam stoppers. Cells were enumerated with acridine orange stain under oil immersion microscopy.(Hobbie et al., 1977)

Proteomic isolation of iron biooxidizing complex

M. sedula was grown in a 2-L bioreactor (media above), sparged with air and agitated at 400 rpm, to a density of approximately 10^9 cells/mL. Cultures were centrifuged for 20 min at 18,000 x g and 4˚C. The decanted pellet was re-suspended in lysis buffer (85.3 mM citric acid, 14.7 mM sodium citrate dehydrate), then lysed with a French pressure cell at 18,000 psi. The lysate was centrifuged at 25,000 x g for 25 min at 4˚C. Membrane fraction isolation was based on a modified protocol (Castelle et al., 2008; Castelle et al., 2015), with ultracentrifugation (L8-70M Beckman) at 145,000 x g for 1 hour at 4˚C. The pellet was re-suspended in lysis buffer supplemented with 750 mM 6-aminocaproic acid, 10% v/v glycerol, 2% w/v n-Dodecyl-B-D-maltoside, then centrifuged at 145,000 x g for 1 h at 4˚C.

Blue Native (BN)-PAGE samples were buffer exchanged with BN-PAGE Wash Buffer (18 mM citric acid, 32 mM sodium citrate dehydrate, 50 mM 6-aminocaproic acid) three times using a 3kDa slide-a-lyzer and then concentrated using a 3 kDa MWCO polyethersulfone concentrator (ThermoFisher), centrifuged at 12,000 x g and 4˚C. Samples were resolved on a 4-16% Native-PAGE Bis-Tris Gel (ThermoFisher), with M_r approximations via NativeMark ladder (ThermoFisher). Duplicate gels were Coomassie-stained or stained with an iron biooxidation buffer (50 mM glycine, 10 mM ferrous sulfate heptahydrate) (Castelle et al., 2015) and incubated at 65˚C for 16 h. Sodium formate (at 50 mM) was used as a cytochrome c oxidase inhibitor for the in-gel iron oxidation stain. Both excised Coomassie bands and raw membrane preps were

LDS-PAGE resolved on 4-12% Bis-Tris NuPAGE polyacrylamide gels in MOPS buffer. The samples were reduced for 20 min at 99˚C with NuPAGE LDS Sample Buffer and NuPAGE Reducing Agent and cooled back to 4˚C prior to loading.

Proteomics

In order to map protein constituents back to the putative proteome of *M. sedula*, both excised complex bands from the BN-PAGE gel and the original solubilized membrane fraction were analyzed by liquid chromatography/mass spectrometry (LC-MS). The solubilized membrane fraction samples were buffer exchanged into a 0.1 M Tris, pH 8.0 buffer, using a 10 kDa MWCO vivaspin concentrator unit, then heated to 95˚C for 5 min and treated with dithiothreitol (DTT) for 5 additional min at 95 ˚C. Samples were cooled to 56 ˚C and treated with DTT for an additional 30 min. Samples were buffer exchanged to 8 M urea with 0.1 M Tris and alkylated with 50 mM iodoacetamide for 1 h. Samples were then washed several times with buffer exchange for 0.1 M $CaCl_2$ in 2M urea, then buffer exchanged to 0.1 M Tris, pH 7.5. Digestion was performed overnight at 37 ˚C with 0.2 μg trypsin and 0.25 μg chymotrypsin, subsequently quenched with formic acid and 0.001% zwittergent 3-16. Samples were lyophilized and stored at -20 ˚C, then reconstituted for LC-MS with 2% acetonitrile and 0.1% formic acid in water.

The gel band samples were prepared by destaining the complexes in 100 mM ammonium bicarbonate (ABC) in water and 50% (v/v) acetonitrile, then acetonitrile. Samples were denatured in 10 mM DTT/100 mM ABC for 10 min at 80 ˚C, then with freshly prepared DTT/ABC at 56 ˚C for 30 min. Samples were washed with acetonitrile, then treated for 30 min with 55 mM iodoacetamide to alkylate. Samples were again washed with acetonitrile, dried using vacuum evaporation, then digested with trypsin and chymotrypsin at 20 μg/μL and 25 μg/μL in 10 mM ABC and 10% (v/v) acetonitrile overnight (Vermachova et al., 2011; Min et al., 2015; Yin et al., 2015). Quenching was achieved with a 1:2 mix of 5% formic acid and acetonitrile incubated at 37 ˚C for 15 min. Samples were recovered from the gel with iterative acetonitrile and 100 mM ABC treatment three times

each in 50 μL for 15 min. The combined eluates were evaporated and stored at -20 ˚C, then reconstituted for LC-MS.

LC was performed with an Easy-nLC 1200 and a 3 μm, 0.075 mm x 20 mm Acclaim PepMap 100 C18 trap column, as well as a 2 μm, 0.075 mm x 250 mm EASY-Spray analytical column (ThermoFisher). Mobile phases transitioned from 2% to 80% acetonitrile, with 0.1% formic acid and the balance water. For MS, an Orbitrap Exploris 480 Mass Spectrometer (ThermoFisher) was utilized with a top-speed method. Data were mapped back to the *M. sedula* proteome, as well as potential sources of contaminants (*e.g.* keratin and trypsin/chymotrypsin) using Proteome Discoverer 2.4.0.305. The quality of PSM identification was set such that false discovery rates are reported as high (combined q-value less than 0.01), medium (0.01-0.05), or low; while all values are reported in supplementary tables 1, 2, and 3, only high/medium confidence identifications were considered.

Iron biooxidation assays

Cell suspensions were harvested in mid-exponential growth and centrifuged at 10,000 x g for 5 min, followed by washing with fresh DSM-88 base salts (no ferric chloride). In microtiter plates, whole cells were incubated with 10 mM ferrous sulfate heptahydrate dissolved in DSM-88 media (no carbon supplement), within a humidified vessel at pH 2 and 70˚C, for 16 h. After incubation, a sample (10μL) was mixed 1:8 (v/v) with 20% ethanol. Ferrous iron quantification was performed by mixing 8% (v/v) ammonium acetate buffer (16.5 M acetate, 3.25 M ammonium in water), 8% (v/v) phenanthroline (11.1 mM 1,10-phenanthroline, 14.8 M methanol in water), 5% (v/v) sample, and reserve water, then incubating at room temperature for 15 min in dark conditions. Samples were read on a spectrophotometer (Biotek) at 510 nm. Sample concentrations were compared to a standard curve of ferrous sulfate heptahydrate; the mean and standard error of triplicate samples are reported.

Indirect iron biooxidation was measured in an approximately 110 mL sealed vessel with a titanium membrane dissolved oxygen sensor (AppliSens). The vessel was equilibrated with pre-warmed (70˚C), dry-air saturated DSM-88 medium, without ferric iron (pH 2.0). Following single-point calibration, the vessel was charged with approximately 5 x 10^9 cells, as determined by epifluorescence cell enumeration (acridine orange stain). Following 5 min of basal oxygen consumption, the vessel was charged to 10 mM with ferrous sulfate suspended in DSM-88 medium without ferric iron, and measurements were recorded for an additional 25 min at 30 s intervals. Species reactions were done with triplicate biological repeats; the oxygen consumption rate was reported as the mean of slopes from linear regression, within the linear window of data collection. Oxygen concentrations were approximated assuming Henry's Law for dry air at 70 ˚C.

Identification of conserved 'fox' genes

Full genome sequences for the available type-strain *Sulfolobaceae* species were obtained from NCBI. Previously designated '*fox*' genes and any unnamed genes in between Msed_RS02400 to Msed_RS02490 (*M. sedula* genome was used as a reference) were used in a BLASTP (Altschul et al., 1990) search throughout the other *Sulfolobaceae*. Potential homologs for comparisons within the order were best-hits with an E-value < 10^{-15}. Functional identifications reported in **Figure 1** were made using annotations from BLASTP-identified homologs (from the non-redundant database).

G+C content of coding sequences was determined with the R package Seqinr (version 3.4-5) (Charif and Lobry, 2007). The data were analyzed and visualized using ggplot2 (version 3.1.1). Broad taxonomical phylogenetic reconstructions were created from protein sequence alignments with the MUSCLE (Edgar, 2004) algorithm (version 3.8.425), and tree construction via FastTree (Price et al., 2010) (version 2.1.5) with default parameters. For FoxA and FoxC protein databases, BLASTP searches were screened with a 50% query and 30% identity cutoff (in comparison to the respective genes).

Reverse transcript screen

Mid-exponential cultures were snap-cooled on dry ice/ethanol and centrifuged at 10,000 x g for 20 min. The cell pellets were washed once with tris-EDTA buffer. The resultant pellets were lysed in TRIzol reagent (Invitrogen) and nucleic acids extracted with chloroform. Cleanup was performed on a RNeasy column (Qiagen) with DNase-I treatment, per manufacturer recommendations. Resultant RNA was quantified using a Nanodrop spectrophotometer (ThermoFisher). One microgram of total RNA from each species was reverse transcribed using iScript RT Supermix (Bio-Rad) and target *fox* genes (*fox*ABCDW) were amplified using Phusion PCR polymerase (New England Biolabs) with an annealing temperature of 60˚C for 25 cycles. Primers utilized in this experiment are provided in **Supplementary Information**.

Results

Relationship between the Fox Cluster and iron biooxidation in Metallosphaera sedula

M. sedula was initially isolated from an Italian solfatara and found to be a prolific metal mobilizer (Huber et al., 1989). *M. sedula's* growth physiology has been examined in some detail, including its CO_2-fixation through the 3-hydroxypropionate/4-hydroxybutyrate cycle (Berg et al., 2007; Fuchs, 2011; Han et al., 2012; Hawkins et al., 2013; Fuchs and Berg, 2014), and oxidation of iron (Auernik and Kelly, 2008) and other metals (Auernik and Kelly, 2010b; Mukherjee et al., 2012; Wheaton et al., 2019). The *M. sedula* genome (Auernik et al., 2008) encodes homologs to Fox Cluster proteins, first identified in the iron-oxidizing extreme thermoacidophilic archaeon *S. metallicus* (Bathe and Norris, 2007) and later in *M. yellowstonensis* (Kozubal et al., 2011). **Figure 1** shows the transcript percentiles of Fox Cluster components (based on normalized logarithmic distribution of oligonucleotide microarray data) for *M. sedula* across numerous conditions and data sets, including autotrophic, heterotrophic, and mixotrophic growth conditions (Auernik and Kelly, 2008; Auernik and Kelly, 2010a, b; Mukherjee et al., 2012; Hawkins et al., 2013; Wheaton et al., 2019). This analysis shows that many Fox Cluster genes are constitutively transcribed at

very high levels, regularly falling in the upper 90th percentile of quantifiable transcripts. Further, this suggests the *fox* genes from *M. sedula* are not stimulated to a measurable degree (Auernik and Kelly, 2008), contrasting transcriptomic data from *S. metallicus*, *M. yellowstonensis*, and hot-acid spring metatranscriptomes, where the presence of soluble or mineralized iron induced *fox* gene expression (Bathe and Norris, 2007; Kozubal et al., 2011). This meta-analysis explains an earlier suggestion that CbsAB-SoxLN (a separate oxidase cluster) might play an important role in metal biooxidation for *M. sedula* (Auernik and Kelly, 2008), in contrast to the other studied metal biooxidizers, as a potential misinterpretation of constitutive expression.

The precise biochemical roles of the Fox Cluster components have not been confirmed. Thus, an attempt was made to isolate proteins involved in iron biooxidization, presumably localized to the cellular membrane given the putative annotation of transmembrane domains in many Fox Cluster components and the biochemistry of iron oxidation (**Figure 1**). Ultracentrifuged and detergent-solubilized *M. sedula* membrane fractions were BN-PAGE resolved **(Figure 2B),** yielding two bands observed at approximately 300 and 600 kDa. When stained with a ferrous sulfate-glycine buffer (**Figure 2A**), these two bands became even more pronounced and yellow/black in coloration, likely due to the localized production of ferric iron-sulfate complexes, indicating that protein complexes in these bands were responsible for oxidizing ferrous iron under acidic conditions (Castelle et al., 2015). Interestingly, the activity observed in the lower molecular mass band appears comparable to that observed in the high mass band, despite a much less intense staining with Coomassie. This iron oxidation activity was inhibited in the presence of formic acid, which behaves as an inhibitor of cytochrome c oxidases, via binding to the a_3 heme group in its oxidized state (Nicholls, 1975). These two bands were excised and separated by denaturing LDS-PAGE. Silver staining of both lanes (**Figure 2C**) shows many similarly migrating proteins in both complexes. Additionally, many lower mass constituents (approximately 20-60 kDa) feature prominently in both complexes as well as the original membrane fraction (**Figure 2D**), while several high mass components are absent in the resolved complexes.

Given the significant contribution of the complexes to the overall membrane fraction, semi-quantitative liquid chromatography/mass spectrometry was performed to identify components of the full membrane preparation, as well as the active membrane complexes. Peptide spectral matches (PSMs) were mapped back to the putative proteome of *M. sedula*. After controlling for trypsin, chymotrypsin and keratin contamination from sample preparation, a large number of PSMs were identified for all of the samples (a full listing is given in **Supplementary Tables 1, 2,** and **3** for the full membrane fraction, high M_r band, and low M_r band, respectively). Within the full membrane sample, a number of Fox proteins were identified among those with high scores, accounting for their posterior error probabilities, including FoxA2 (Msed_RS02490), FoxB (Msed_RS02465), and FoxC (Msed_RS02455), while no FoxA1 (Msed_RS02485) was detected (**Table 1**). These are particularly meaningful identifications given that *foxA2* encodes a putative cytochrome c oxidase subunit I (transcribed in transcriptional meta-analysis at much higher levels than *foxA1,* a putative paralog), while *foxB* encodes a putative quinol oxidase (cytochrome c oxidase subunit II) and *foxC* encodes a putative cytochrome b-containing protein. Further, PSMs from FoxA2 and FoxC were relatively abundant in the low M_r complex, as well as some PSMs from FoxB. For FoxA2 and FoxB, obvious corresponding bands appeared in the 2D and 1D denaturing gels around 65 and 20 kDa, while no significant bands corresponded to the expected M_r of FoxC (56.7 kDa without a putative signal peptide).

A few other Fox proteins were identified in the analysis, including small hypotheticals (Msed_RS02475 and Msed_RS02460) and FoxE (Msed_RS02440) with few identifiable peptides, and only one peptide in the case of Msed_RS02460 and Msed_RS02440. Additionally, a putative AAA family ATPase (Msed_RS02415), conserved in the region of the *fox* operon in several *Sulfolobaceae*, was detected in the full membrane sample and the high M_r complex, but not the low M_r complex. In terms of other putative respiratory complexes, traces from the SoxABCDD'L, SoxEFGHIM, DoxBCE, and SoxNL-CbsAB loci (**Supplementary Table 4**) were detected in the full membrane fraction. However, only traces of SoxC were found in the low M_r

complex, while none of these components were detected in the high M_r complex. Interestingly, there were no detectable rusti-/sulfo-cyanin-like proteins detected in the samples, except for small amounts of SoxE, which along with SoxM and SoxG was detected in the full membrane fraction.

Given the observation that some extracellular proteins of thermoacidophilic archaea may exhibit glycosylation (Wang et al., 2019), post-electrophoresis staining of the membrane fraction was performed (**Figure 2E**), showing a number of stain-sensitive proteins. The most intensely stained bands appear to correspond with constituents that are absent in the iron-staining complexes, while bands between 48-66 kDa (3 total) and two more near 18 kDa could correspond with components of the these complexes. While *in silico* prediction of archaeal glycosylation is limited by few validated examples, prediction software based primarily upon eukaryotic glycoproteins identified a number of potential N-linked glycosylation sites in the Fox proteins that were examined (**Table 1**). In the case of FoxB and FoxC, these sites occur in between PSMs detected during LC/MS, further suggesting these moieties could interfere with the accurate detection of all motifs. In previously identified N-linked glycosylation of a membrane protein from *S. acidocaldarius,* a unique hexamer was identified, suggesting that glycosylation groups could add considerable shifts to the proteins of interest here (Hettmann, 1998).

Sulfolobaceae comparative genomic identification of the iron biooxidation determinants

Genome sequence information from the archaeal *Sulfolobaceae* had previously been limited to a few species from the genera *Sulfuracidifex*, *Metallosphaera*, and *Acidianus*, but recently was expanded to all seven named genera and many of the family's type-strains (Counts et al., 2018; Counts et al., 2020a; Counts et al., 2020b). This was due in part to early difficulties using Sanger sequencing for genomes that are highly repetitive, difficult to assemble, and lacked reference genomes. This database was subsequently expanded to include several new type strains, including *Acidianus copahuensis* and *Sulfodiicoccus acidiphilus* (Sakai and Kurosawa, 2017; Urbieta et al., 2017). While signatures of the Fox Cluster are evident among certain

Sulfolobales, there is some variability in locus content and structure. The *Fox* Cluster across eleven species, shown in **Figure 3**, is both well-conserved in terms of genetic content (high bi-directional homology) and highly syntenous. At least three sets of genes are universally conserved in terms of identity, orientation, and proximity. Further, the organization of these genes seems to follow the evolutionary partitioning of the order *Sulfolobales*.

Among these gene groupings in **Figure 3**, the blue region is of particular interest because it contains two genes (*foxB* and *foxC*) with high homology to metal cofactor-binding domains. In particular, FoxB has a conserved domain for a quinol oxidase and FoxC contains a putative heme b binding domain. Interestingly, this region is almost always adjacent, or fewer than 10 genes removed, from the primary subunit (*foxA*) of a putative cytochrome c oxidase-like (COXA-like) protein with putative heme and copper binding domains. Another important region in the locus with high conservation is the *foxVEUF* region, containing three genes, as well as a small unannotated putative open reading frame, conserved in all reported *Sulfolobaceae* metal mobilizers, except for *Metallosphaera cuprina,* whose iron oxidation capacity could not be determined here because the species is not available. In addition, there are several small proteins (encoded in *foxW and foxYZ*) that appear to be highly syntenous: *foxW* with all species' *foxBCD* locus (except for *M. hakonensis*, due to transposition), and *foxYZ* proximal to either *foxB* or *foxA*. Otherwise, a notable inconsistency from these loci is *foxGH*, which can be far removed (or completely absent in the case of *Sulfodiicoccus acidiphilus*). This result is particularly intriguing because past results have suggested this locus is induced in the presence of sulfur or sulfidic ores for *S. metallicus* (Bathe and Norris, 2007*)*, while *S. acidiphilus* was recently reported to be the first *Sulfolobaceae* member whose growth is actually inhibited in the presence of elemental sulfur (Sakai and Kurosawa, 2017).

Iron biooxidation capacity among reported extremely thermoacidophilic iron biooxidizers

Several members of the archaeal family *Sulfolobaceae*, whose genomes encode components of the *Fox* Cluster, oxidize either aqueous iron or iron-sulfidic ores (Wheaton et al., 2015). In order to examine this trait and characterize metal biooxidation in these species, measurements were taken, including direct determination of iron oxidation and respiration in the presence of iron. Colorimetric identification of soluble ferrous or total iron in actively growing cultures (**Figure 4A** and **4B)** suggest that, contrary to initial reports, *M. hakonensis* from DSMZ culture collections does not oxidize iron to any detectable levels (as compared to *S. acidocaldarius* and a negative control without any cells). Interestingly, the two most proficient iron oxidizers were *M. prunae* and *M. sedula,* which oxidized nearly 90% of the supplied iron and generated the largest Fe^{3+}/Fe^{2+} ratios, suggesting robust oxidation and a high redox equilibrium. For the other species: *Acidianus brierleyi, Acidianus sulfidivorans, S. metallicus, Sulfurisphaera ohwakuensis,* and *Sulfurisphaera tokodaii*, lower redox ratios were observed, but 25-55% of the ferrous iron originally supplied to the cultures was oxidized over the course of sixteen hours. In the case of *Sulfurisphaera ohwakuensis,* this capability was not previously reported, making it one of the most thermoacidophilic iron oxidizers to date (optimal growth temperature 85˚C; optimal pH 2.0).

Additionally, a shorter timeframe experiment measured O_2 consumption (cellular respiration) in response to ferrous iron addition for all of these species over the course of 30 minutes and the same physiological conditions. Mid- to late-exponential phase cultures were washed to remove residual growth media and used to inoculate a dry-air saturated basal salts solution in closed vessels. After allowing for a brief equilibrium, soluble ferrous iron was added to the reactor vessel and dissolved oxygen levels monitored. Oxygen consumption was determined by assuming Henry's Law of infinite dilution of oxygen at the reactor conditions (70˚C, pH 2.0). Representative traces (aggregates of three replicates) are shown in **Figure 4C and 4D.** The abiotic control for iron biooxidation was approximately $1/80^{th}$ that observed for the most proficient

iron oxidizer, *M. sedula*, suggesting a low rate of auto-oxidation during this time frame. Unsurprisingly, *M. hakonensis* (**Figure 4D,** O_2 consumption: 18.6 ± 3.68 μmol/min) also exhibited this markedly low respiration rate. However, *S. tokodaii* (13.8 ± 9.22 μmol/min) also demonstrated a low respiration rate under these conditions. Other species exhibited rates ranging between 40.8-153 μmol/min, including *M. prunae,* which despite oxidizing comparable amounts of iron over an extended period, showed an O_2-consumption rate 1/3rd that of *M. sedula*. *M. sedula* was the most proficient iron oxidizer and the organism to exhibit the highest rates of respiration over the course of the assay nearing complete depletion O_2 over the course of 30 minutes.

Reverse transcription of key Fox Cluster genes

Given the conservation of the *fox* operon but disparate iron biooxidation for a number of species, reverse transcription of key *fox* genes was examined using RNA collected from heterotrophic growth conditions for *M. sedula, M. prunae, M. hakonensis,* and *S. tokodaii* (**Figure 5**). Optimizing the experiment for RNA input and post reverse transcription PCR cycles (25 here), qualitative differences were easily identified among these separate species. Unsurprisingly, the largest qualitative transcription levels for the gene groups around *foxA* and *foxC* in any of the organisms was *M. sedula*, agreeing with the analysis of existing transcriptomic data sets from *M. sedula* (**Figure 1**), where *M. sedula* shows relatively high constituitive transcription of these genes. This lends credence to the importance of these genes in iron biooxidation for the most proficient iron oxidizer examined in this study and explains the predominance of FoxA2 and FoxC PSMs identified by LC/MS. Additionally, *foxA1* was not transcribed at high levels. Intriguingly for *M. prunae*, no apparent *foxA* transcription occurs, although additional qPCR (**Figure 5** inset) demonstrated that the gene is transcribed at roughly three orders of magnitude (1,000-fold) less than *foxA2* (from *M. sedula*), but at similar levels to *foxA1* (also *M. sedula*). This likely accounts

for a difference in presence of FoxA2 in the *M. sedula* membrane and the lowered oxidation rate for *M. prunae* shown in **Figure 4**.

In the case of the other metal-mobilizers, *S. tokodaii* and *M. hakonensis*, issues with transcription of genes from the *foxBWCD* locus were apparent. For the former, a frameshift mutation appears to disrupt the sequence of *foxC*, potentially requiring the production of a fusion protein from the existing N-terminal domain and the out of frame but orthologous C-terminal domain is translated separately. Thus, the lower C-terminal domain transcription levels may impede production or assembly. *For M. hakonensis*, there is an apparent transposon disrupting the *foxW* gene (confirmed by its absence in RT-PCR). Furthermore, this disruption may have a polar effect on *foxC*, as no transcript is visible on the gel. In this case again, the *foxA* duplicates appear to show little basal transcription in *M. hakonensis*. Finally, *foxD* had the least or no visible transcription in all of the strains studied, supporting its absence in the proteomic samples, and calling into question its proposed criticality in the *Fox* membrane complex.

Evolutionary track of the "fox" genes

There are few, if any, homologs for the *Fox* Cluster genes in other *Sulfolobaceae*, other than proteins containing broadly conserved domains, such as the cytochrome oxidases that have both lower homology and alternative putative or confirmed biochemical purposes. This raises the question: Were the *Fox* Cluster genes obtained through horizontal gene transfer (HGT) or passed down through evolutionary processes? The possibility of horizontal gene transfer was explored for the species listed in **Figure 3**. This analysis was done from two perspectives: parametrically based on G+C content of genes (including in the codon positions; **Figure 6**) and phylogenetically comparing FoxA and FoxC specifically to available homologous proteins (**Figure 7**).

In the case of G+C analysis, **Figure 6** shows that the "*fox"* genes are not outliers, as would be expected if the genes were the result of a recent HGT event and had not ameliorated to the genetic background of the individual genomes. Further, the genes are well-distributed within the

whiskers of each box plot with few exceptions, suggesting either deeper evolutionary origins or amelioration of the genes to suit the codon preferences of the individual species. The phylogenetic analysis (**Figure 7**) shows that the *foxA* and its putative cytochrome c oxidase-like domain are widely conserved across the prokaryotic kingdoms, with almost all of the archaeal constituents forming a tighter and divergent clade from other bacterial constituents. Meanwhile *foxC* is relatively unique from an evolutionary perspective, with no close relatives beyond the archaea. Inserts of the trees in **Figure 7** show that the proteins have largely evolved along taxonomical lines, as one would expect. Interestingly, but not surprisingly, there are numerous proteins with high sequence similarity to both FoxA and FoxC that also form other terminal oxidase complexes present throughout the *Sulfolobales (Lewis et al., 2021)*. Interestingly, these homologous proteins are actually separated from the best-hit FoxA and FoxC proteins by branches containing homologs identified in the new candidate phylum *Marsarchaeota*. This suggests earlier evolutionary events differentiating these proteins role in ferrous iron oxidation, which is consistent with this candidate phylum's presence in iron mats in Yellowstone National Park (Jay et al., 2018). Intriguingly, this new phylum was found in higher density just below a more oxic layer containing signatures of the species *M. yellowstonensis*.

Discussion

Microbiological, biochemical and transcriptomic data support the essentiality of FoxA, FoxB, and FoxC for iron biooxidation by extremely thermoacidophilic archaea. Specifically, their presence in the proteomic membrane compartment in conjunction with clear aerobic iron oxidation by these same membrane fractions (**Figure 2**), as well as iron and oxygen-dependent biochemistries of *in vivo* cells (**Figure 4**), queue into their essentiality for iron biooxidation. Previous reports suggested that FoxD ($M_r \approx 35$ kDa) and FoxG ($M_r \approx 69$ kDa) play an essential role in stabilizing FoxC and shuttling electrons downhill, although no evidence of either one of the transmembrane-heavy proteins was seen in the proteomics or the LDS-PAGE analysis from the

active membrane fractions. Although heavy glycosylation has been observed in some extracellular crenarchaeal proteins (Hettmann, 1998; Albers et al., 2017), resulting in poor resolution in the standard LDS buffer and an inability of proteases to access cleavage motifs for proteomic analysis (Wang et al., 2019), while large hydrophobic transmembrane domains can prove difficult to resolve by standard denaturing gel techniques.

Consideration of comparative genomics data, in conjunction with transcriptomic and phenotypical information, revealed more about the key genes in the biooxidation process. In the archaea considered here, there were at least three interesting traits that further support the essentiality of *foxA* and *foxC* to iron oxidation. In the case of *M. prunae* and *M. sedula*, all genes share 100% amino acid sequence identity in the Fox Cluster, with the exception of FoxA. In this case, the version of FoxA in *M. prunae* appears as a fusion of the complementary N- and C-terminal domains of FoxA1 and FoxA2, respectively in *M. sedula*, with the promoter region of *foxA1*. Given the much lower levels of *foxA* transcription (**Figure 5**), lower respiratory rates, but comparable iron biooxidation compared to *M. sedula* (**Figure 4**), one might assume that the fusion protein performs a similar task more efficiently or is regulated by some environmental stimulus, instead of being constitutively transcribed. It is possible that this ortholog has much higher oxygen affinity and is regulated in response to changing oxygen tension, a regulatory trait observed in other *Sulfolobales* as a potential way of managing the beneficial and deleterious effects of oxygen bioenergetics (Simon et al., 2009).

This was a curious result, given *M. prunae* was originally isolated from a uranium mining slag heap and determined to have increased resistance to uranium toxicity versus *M. sedula* (despite failing to meet the genome-based definition of speciation), possibly through more finely-tuned iron biooxidation (Mukherjee et al., 2012). In addition, we have observed delayed metal-dependent respiration for *M. prunae* upon the removal and reintroduction of iron in its growth media that is in contrast to *M. sedula,* which can now be rationalized by the disparate regulation of the *foxA* gene(s) (Wheaton et al., 2019). Additionally, this suggests an alternative uranium

toxicity mitigation process (Mukherjee et al., 2012) in which decreased iron biooxidation activity may help prevent excessive production of toxic hexavalent uranium species from insoluble metal ores via indirect biooxidation by ferric iron. This couples with an RNase-mediated mechanism of suppression (Mukherjee et al., 2017) that balances modest metal biooxidation (to maintain pH homeostasis) with preventing the buildup of electrophilic ferric iron species in the presence of highly toxic heavy metals.

In the case of FoxC, both *M. hakonensis* and *S. tokodaii* provide potential validation of its essentiality. In both cases, little or no iron biooxidation was observed despite reports of growth on iron substrates; this is perhaps an example of unintended laboratory-directed evolution, although fortuitously consequential in this circumstance. In the case of *M. hakonensis,* this may be the result of a transpositional disruption of the *foxW* gene (**Figure 3**), leading to its loss and the potential disruption of transcription (polar effect) for the *foxC* gene, since both have no apparent transcription in **Figure 5.** In the case of *S. tokodaii*, the genome sequence suggests a disruption of the full protein sequence by a SNP (**Figure 3**) that potentially displaces the planar and axial coordinating residues of the protein in two separate domains, based on previous modeling of the FoxC protein (Kozubal et al., 2011).

The recent addition of several metal-biooxidizers to this group provides further evidence for the importance of these proteins and several others, via synteny, in metal biooxidation. In particular, the loci with *foxBWCD, foxYZ,* and *foxVEUFJ* have demonstrated synteny among all of the species examined here and a few other species confirmed to oxidize metals in laboratory cultures. While the functions of *foxABC* (type-aa3 cytochrome/quinol oxidase I, quinol oxidase subunit II, and cytochrome b-containing protein, respectively) are bioinformatically apparent, all of the other proteins show no remarkable similarity to any known protein in publicly accessible databases. Furthermore, while FoxABC have remarkably high similarity (> 60% across the board), almost all of the other proteins show larger evolutionary drift (30-60% similarity), with the exception of FoxV and FoxE, which are both above 50% for all species compared to *M. sedula.*

Given the induction of the *foxVEUFJ* locus and their large numbers of transmembrane domains, one putative role is in the assembly and delivery of cofactors to the larger cytochrome oxidase complex, a role without candidates in current explanations.

Finally, the *foxGH* locus, believed to be involved in signal transduction or in the cytochrome b complex (Kozubal et al., 2011), is dislocated from the other genes in some species and is absent in *Sulfodiicocuus acidiphilus*. While there is no report as to the ability of this microbe to utilize iron (only inorganic substrates containing sulfur were tested), there is a report that the organism cannot grow in the presence of elemental sulfur (Sakai and Kurosawa, 2017). Given the original analysis describing the gene pair as pyrite- (and not Fe^{2+}) induced (Bathe and Norris, 2007), it now seems possible that the genes are tied to some form of sulfur/sulfidic ore metabolism absent in the new species and not directly involved in iron metabolism. In further support, BLAST searches suggest the absence of the *foxGH* genes in a newly discovered lineage, the *Marsarchaeota*, with distant links between the *Crenarchaea* and other phyla of the kingdom *Archaea* (Jay et al., 2018). In fact, these organisms were proposed to have a major role in iron cycling in the microaerobic region of iron mats, although several of the binned genomes have clear homologs to *foxABCDEF.*

Another gene proposed to perform an essential role are copper oxidases, such as those in the metal-mobilizing mesophilic bacterium *A. ferroxidans (Quatrini et al., 2007)*, as well as the mesophilic euryarchaeon *F. acidiphilum* (Castelle et al., 2015), that serves in a putative role as a facilitator of redox balance, spanning the periplasmic (bacteria) or pseudo-periplasmic (archaea) space. However, the lack of a true periplasmic space in the archaea calls into question the necessity of a “nanowire-like” protein, although in *A. ferroxidans,* it is hypothesized to serve as a branching point between uphill (NAD+ reduction) and downhill (oxygen reduction) electron flow (Elbehti et al., 2000). Previous analysis has identified at least four copper oxidases, two of the sulfocyanin type and two of the rusticyanin (one with plastocyanin-like) type, in *M. sedula* (Auernik et al., 2008) and several copper oxidases in *M. yellowstonensis* (Kozubal et al., 2011). Here,

comparative genomics suggests that there is a rusticyanin (Msed_RS04905) homolog in all of the studied genomes, with the exception of *S. acidiphilus* and *S. tokodaii* (Identity above 40% and coverage over 65%). However, the protein from *M. sedula* also shares homology with proteins of non-metal-mobilizing species, such as *Saccharrolobus solfataricus* and *Sulfolobus islandicus*. The other protein of interest, encoded in Msed_RS06125206, has only a few homologs in the *Metallosphaera spp.* as well as in the *Marsarchaeota*, *Sulfolobus islandicus* strains and *Sulfodiicoccus* acidiphilus. Given a lack of strong conservation and synteny (the genes are always displaced from the *fox* locus, in contrast to *A. ferrooxidans*), the protein may serve a more universal electron transfer role or act as a critical carrier in an ancillary process.

Conclusion

In this report, we demonstrate the essentiality of FoxA, FoxB, and FoxC in the process of iron biooxidation in the extremely thermoacidophilic archaea and provide additional evidence for the role of unconstrained *foxABWC* transcription in the proficient biooxidization capacity of *M. sedula*. This suggests a potential application in metal biooxidation biotechnological applications, which have been dominated by mesophilic species and the few more characterized archaeal thermoacidophiles (Wheaton et al., 2015; Gumulya et al., 2018; Johnson, 2018). Interestingly, this process is deeply rooted in the extremely thermoacidophilic aerobic archaea and even appears in a somewhat distant archaeal lineage in roughly the same genetic constitution. This process may be a far distant transfer event or a more recent convergent evolution emerging in the wake of an oxygenated world, exclusively in environments dominated by heavy inorganic mineral dissolution and copious energy dissipation from globally distributed geothermal activity.

Acknowledgments

JAC and NPV acknowledge assistance from Mashkurel Haque, Alexandra Comer, Samuel Hamacher, and Shadia Taylor. This work was supported by the US Air Force Office of Scientific

Research (FA9550-17-1-0268, FA9550-20-1-0216) and the US National Science Foundation
565 (CBET-1802939). JAC acknowledges support from an NIH Biotechnology Traineeship (2T32GM008776). The proteomics analysis was performed in part by the Molecular Education, Technology and Research Innovation Center (METRIC) at NC State University, which is supported by the State of North Carolina.

References

Albers, S., Elichler, J., and Aebi, M. (2017) Archaea. In Essentials of Glycobiology. Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M. et al. (eds). Cold Springs Harbor, NY: Cold Springs Harbor Laboratory Press.

*Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. J Mol Biol **215**: 403-410.*

*Appia-Ayme, C., Guiliani, N., Ratouchniak, J., and Bonnefoy, V. (1999) Characterization of an operon encoding two c-type cytochromes, an aa(3)-type cytochrome oxidase, and rusticyanin in Thiobacillus ferrooxidans ATCC 33020. Appl Environ Microbiol **65**: 4781-4787.*

*Archibald, F. (1983) Lactobacillus plantarum, an organism not requiring iron. FEMS Microbiol Lett **19**: 29-32.*

*Auernik, K.S., and Kelly, R.M. (2008) Identification of components of electron transport chains in the extremely thermoacidophilic Metallosphaera sedula through iron and sulfur compound oxidation transcriptomes. Appl Environ Microbiol **74**: 7723-7732.*

*Auernik, K.S., and Kelly, R.M. (2010a) Physiological versatility of the extremely thermoacidophilic archaeon Metallosphaera sedula supported by transcriptomic analysis of heterotrophic, autotrophic, and mixotrophic growth. Appl Environ Microbiol **76**: 931-935.*

*Auernik, K.S., and Kelly, R.M. (2010b) Impact of molecular hydrogen on chalcopyrite bioleaching by the extremely thermoacidophilic archaeon Metallosphaera sedala. Appl Environ Microbiol **76**: 2668-2672.*

*Auernik, K.S., Maezato, Y., Blum, P.H., and Kelly, R.M. (2008) The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon Metallosphaera sedula provides insights into bioleaching-associated metabolism. Appl Environ Microbiol **74**: 682-692.*

*Baker-Austin, C., and Dopson, M. (2007) Life in acid: pH homeostasis in acidophiles. Trends Microbiol **15**: 165-171.*

*Bathe, S., and Norris, P.R. (2007) Ferrous iron- and sulfur-induced genes in Sulfolobus metallicus. Appl Environ Microbiol **73**: 2491-2497.*

*Bengrine, a., Guiliani, N., Appia-Ayme, C., Jedlicki, E., Holmes, D.S., Chippaux, M., and Bonnefoy, V. (1998) Sequence and expression of the rusticyanin structural gene from Thiobacillus ferrooxidans ATCC33020 strain. Biochim Biophys Acta **1443**: 99-112.*

*Berg, I.A., Kockelkorn, D., Wolfgang, B., Fuchs, G., Buckel, W., and Fuchs, G. (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. Science **318**: 1782-1786.*

*Bonnain, C., Breitbart, M., and Buck, K.N. (2016) The Ferrojan Horse Hypothesis: Iron-Virus Interactions in the Ocean. Front Mar Sci **3**.*

*Bonnefoy, V., and Holmes, D.S. (2012) Genomic insights into microbial iron oxidation and iron uptake strategies in extremely acidic environments. Environ Microbiol **14**: 1597-1611.*

*Cassat, J.E., and Skaar, E.P. (2013) Iron in infection and immunity. Cell Host Microbe **13**: 509-519.*

*Castelle, C., Guiral, M., Malarte, G., Ledgham, F., Leroy, G., Brugna, M., and Giudici-Orticoni, M.-T. (2008) A new iron-oxidizing/O2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile Acidithiobacillus ferrooxidans. J Biol Chem **283**: 25803-25811.*

*Castelle, C.J., Roger, M., Bauzan, M., Brugna, M., Lignon, S., Nimtz, M. et al. (2015) The aerobic respiratory chain of the acidophilic archaeon Ferroplasma acidiphilum: A membrane-bound complex oxidizing ferrous iron. Biochim Biophys Acta, Bioenerg **1847**: 717-728.*

Charif, D., and Lobry, J.R. (2007) SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In Structural Approaches to Sequence Evolution: Molecules, Networks, Populations. Bastolla, U., Porto, M., Roman, H.E., and Vendruscolo, M. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 207-232.

Counts, J.A., Vitko, N.P., and Kelly, R.M. (2018) Complete genome sequences of extremely thermoacidophilic metal-mobilizing type strain members of the Archaeal Family Sulfolobaceae, Acidianus brierleyi DSM-1651, Acidianus sulfidivorans DSM-18786, and Metallosphaera hakonensis DSM-7519. Microbiol Res Announc ***7****.*

Counts, J.A., Vitko, N.P., and Kelly, R.M. (2020a) Genome sequences of five type strain members of the archaeal Family Sulfolobaceae, Acidianus ambivalens, Acidianus infernus, Stygiolobus azoricus, Sulfuracidifex metallicus, and Sulfurisphaera ohwakuensis. Microbiol Resour Announc ***9****.*

Counts, J.A., Willard, D.J., and Kelly, R.M. (2020b) Life in hot acid: a genome-based reassessment of the archaeal order Sulfolobales. Environ Microbiol.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res ***32****: 1792-1797.*

Elbehti, A., Brasseur, G., and Lemesle-Meunier, D. (2000) First evidence for existence of an uphill electron transfer through the bc(1) and NADH-Q oxidoreductase complexes of the acidophilic obligate chemolithotrophic ferrous ion-oxidizing bacterium Thiobacillus ferrooxidans. J Bacteriol ***182****: 3602-3606.*

Emerson, D., Fleming, E.J., and McBeth, J.M. (2010) Iron-oxidizing bacteria: an environmental and genomic perspective. Annu Rev Microbiol ***64****: 561-583.*

Fuchs, G. (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu Rev Microbiol ***65****: 631-658.*

Fuchs, G., and Berg, I.A. (2014) Unfamiliar metabolic links in the central carbon metabolism. J Biotechnol ***192****: 314-322.*

Gonzalez-Toril, E., Llobet-Brossa, E., Casamayor, E.O., Amann, R., and Amils, R. (2003) Microbial ecology of an extreme acidic environment, the Tinto River. Appl Environ Microbiol ***69****: 4853-4865.*

*Gumulya, Y., Boxall, N.J., Khaleque, H.N., Santala, V., Carlson, R.P., and Kaksonen, A.H. (2018) In a quest for engineering acidophiles for biomining applications: challenges and opportunities. Genes **9**.*

*Han, Y., Hawkins, A.S., Adams, M.W.W., and Kelly, R.M. (2012) Epimerase (Msed_0639) and mutase (Msed_0638 and Msed_2055) convert (S)-methylmalonyl-coenzyme a (CoA) to succinyl-coA in the Metallosphaera sedula 3-hydroxypropionate/4-hydroxybutyrate cycle. Appl Environ Microbiol **78**: 6194-6202.*

*Hawkins, A.S., Han, Y., Bennett, R.K., Adams, M.W.W., and Kelly, R.M. (2013) Role of 4-hydroxybutyrate-CoA synthetase in the CO_2 fixation cycle in thermoacidophilic archaea. J Biol Chem **288**: 4012-4022.*

*Hedrich, S., Schlomann, M., and Johnson, D.B. (2011) The iron-oxidizing proteobacteria. Microbiology **157**: 1551-1564.*

*Hettmann, T. (1998) Cytochrome b558/566 from the Archaeon Sulfolobus acidocaldarius. A novel highly glycosylated, membrane-bound B-type hemoprotein. J Biol Chem **273**: 12032-12040.*

*Hobbie, J.E., Daley, R.J., and Jasper, S. (1977) Use of nuclepore filters for counting bacteria by fluorescence microscopy. Appl Environ Microbiol **33**: 1225-1228.*

*Huber, G., Spinnler, C., Gambacorta, A., and Stetter, K.O. (1989) Metallosphaera sedula gen. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaebacteria. Syst Appl Microbiol **12**: 38-47.*

*Ilbert, M., and Bonnefoy, V. (2013) Insight into the evolution of the iron oxidation pathways. Biochim Biophys Acta **1827**: 161-175.*

*Ingledew, W.J. (1982) Thiobacillus ferrooxidans the bioenergetics of an acidophilic chemolithotroph. Biochim Biophys Acta, Rev Bioenerg **683**: 89-117.*

*Jay, Z.J., Beam, J.P., Dlakić, M., Rusch, D.B., Kozubal, M.A., and Inskeep, W.P. (2018) Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. Nat Microbiol **3**: 732-740.*

*Johnson, D. (2018) The evolution, current status, and future prospects of using biotechnologies in the mineral extraction and metal recovery sectors. Minerals **8**.*

*Johnson, D.B., Kanao, T., and Hedrich, S. (2012) Redox Transformations of iron at extremely low pH: Fundamental and applied aspects. Front Microbiol **3**.*

*Kozubal, M.a., Dlakic, M., Macur, R.E., Inskeep, W.P., Dlakić, M., Macur, R.E., and Inskeep, W.P. (2011) Terminal oxidase diversity and function in "Metallosphaera yellowstonensis": gene expression and protein modeling suggest mechanisms of Fe(II) oxidation in the Sulfolobales. Appl Environ Microbiol **77**: 1844-1853.*

*Kozubal, M.a., Macur, R.E., Jay, Z.J., Beam, J.P., Malfatti, S.a., Tringe, S.G. et al. (2012) Microbial iron cycling in acidic geothermal springs of Yellowstone National Park: Integrating molecular surveys, geochemical processes, and isolation of novel Fe-active microorganisms. Front Microbiol **3**: 1-16.*

Lewis, A.M., Recalde, A., Brasen, C., Counts, J.A., Nussbaum, P., Bost, J. et al. (2021) The biology of thermoacidophilic archaea from the order Sulfolobales. FEMS Microbiol Rev.

*Min, L., Choe, L.H., and Lee, K.H. (2015) Improved protease digestion conditions for membrane protein detection. Electrophoresis **36**: 1690-1698.*

*Mukherjee, A., Wheaton, G.H., Blum, P.H., and Kelly, R.M. (2012) Uranium extremophily is an adaptive, rather than intrinsic, feature for extremely thermoacidophilic Metallosphaera species. Proc Natl Acad Sci, USA **109**: 16702-16707.*

*Mukherjee, A., Wheaton, G.H., Counts, J.A., Ijeomah, B., Desai, J., and Kelly, R.M. (2017) VapC toxins drive cellular dormancy under uranium stress for the extreme thermoacidophile Metallosphaera prunae. Environ Microbiol **19**: 2831-2842.*

*Nicholls, P. (1975) Formate as an inhibitor of cytochrome c oxidase. Biochem Bioph Res Co **67**: 610-616.*

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments [WWW document].

*Quatrini, R., Lefimil, C., Veloso, F.A., Pedroso, I., Holmes, D.S., and Jedlicki, E. (2007) Bioinformatic prediction and experimental verification of Fur-regulated genes in the extreme acidophile Acidithiobacillus ferrooxidans. Nucleic Acids Res **35**: 2153-2166.*

*Sakai, H.D., and Kurosawa, N. (2017) Sulfodiicoccus acidiphilus gen. nov., sp. nov., a sulfur-inhibited thermoacidophilic archaeon belonging to the order Sulfolobales isolated from a terrestrial acidic hot spring. Int J Syst Evol Microbiol **67**: 1880-1886.*

*Sandy, M., and Butler, A. (2009) Microbial Iron Acquisition: Marine and Terrestrial Siderophores. Chem Rev **109**: 4580-4595.*

*Simon, G., Walther, J., Zabeti, N., Combet-Blanc, Y., Auria, R., van der Oost, J., and Casalot, L. (2009) Effect of O2 concentrations on Sulfolobus solfataricus P2. FEMS Microbiol Lett **299**: 255-260.*

*Tagliabue, A., Bowie, A.R., Boyd, P.W., Buck, K.N., Johnson, K.S., and Saito, M.A. (2017) The integral role of iron in ocean biogeochemistry. Nature **543**: 51-59.*

*Urbieta, M.S., Rascovan, N., Vazquez, M.P., and Donati, E. (2017) Genome analysis of the thermoacidophilic archaeon Acidianus copahuensis focusing on the metabolisms associated to biomining activities. BMC Genomics **18**: 445.*

*Vermachova, M., Purkrtova, Z., Santrucek, J., Jolivet, P., Chardot, T., and Kodicek, M. (2011) New protein isoforms identified within Arabidopsis thaliana seed oil bodies combining chymotrypsin/trypsin digestion and peptide fragmentation analysis. Proteomics **11**: 3430-3434.*

Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., Su, Z., de Oliveira, G.A.P., Osinski, T. et al. (2019) An extensively glycosylated archaeal pilus survives extreme conditions. Nat Microbiol.

*Welham, N.J., Malatt, K.A., and Vukcevic, S. (2000) The effect of solution speciation on iron-sulfur-arsenic-chloride systems at 298 K. Hydrometallurgy **57**: 209-223.*

*Wheaton, G., Counts, J., Mukherjee, A., Kruh, J., and Kelly, R. (2015) The confluence of heavy metal biooxidation and heavy metal resistance: implications for bioleaching by extreme thermoacidophiles. Minerals **5**: 397-451.*

*Wheaton, G.H., Vitko, N.P., Counts, J.A., Dulkis, J.A., Podolsky, I., Mukherjee, A., and Kelly, R.M. (2019) Extremely thermoacidophilic Metallosphaera species mediate mobilization and oxidation of vanadium and molybdenum oxides. Appl Environ Microbiol **85**: e02805-02818.*

*Yin, L., Vener, A.V., and Spetea, C. (2015) The membrane proteome of stroma thylakoids from Arabidopsis thaliana studied by successive in-solution and in-gel digestion. Physiol Plant **154**: 433-446.*

TABLE AND FIGURE LEGENDS

Table 1. Mass Spectrometry and Amino Acid Sequence Analysis of *M. sedula fox* Genes. List of *fox* genes and their detection by liquid chromatography mass spectrometry in the membrane fraction and both complex bands from BN-PAGE. Sequence analysis shows positional data for sequence length in amino acids in yellow (AA), LC/MS-detected peptide spectral matches (PSMs) in blue (MS), position of trypsin/chymotrypsin-digestible sites in red (CS), likelihood of N-linked glycosylation site based on the prokaryotic-trained algorithm GlycoPPv1.0 (GP), and SEC signal peptide prediction likelihood in purple based on the SignalPv5.0 “Archaeal” algorithm (SP). For each sample, the coverage of the full amino acid sequence in PSMs and the number of PSMs are reported.

Figure 1. Fox Cluster in *M. sedula*. Annotations are based on *S. metallicus* transcriptome, for genes responsive to iron and BLASTP searches. M_r and pI were calculated in the absence of putative cleavage sequences using EXPASY. Transmembrane (TM) domains were determined with TMHMM v. 2.0 and Signal peptides (SP) were determined with SignalP v. 5.0 with setting “Archaea.” Reported transcriptional data were compiled from previous *M. sedula* transcriptional studies. For signal peptides, †FoxC has a putative Sec/SPI site at the sequence AYG^AD with 0.9193 likelihood, ‡FoxB has a putative Sec/SPI site at the sequence LET^QY with 0.8257 likelihood, all other likelihoods were below 0.1. Values in parenthesis for TMs represent values minus the signal peptide region. Bracketed gene names are previously unnamed genes, conserved in all species from this study. Orange shading is 5’ to 3’, green shading is 3’ to 5’.

Figure 2. Gel Electrophoresis of *M. sedula* membrane fractions. (A) BN-PAGE of membrane fraction with ferrous iron/glycine buffer under physiologically-relevant conditions: 65 ˚C and pH ~ 3.0. (B) BN-PAGE of membrane fraction stained with Coomassie Brilliant Blue (image in black and white). (C) LDS-PAGE of bands from Blue-Native PAGE gel (B) stained with Coomassie Brilliant Blue. (D) LDS-PAGE of original membrane preparation (1-dimension). (E) Glycoprotein staining of membrane preparation under denaturing conditions with Pro-Q Emerald 300. NativeMark, BenchMark, and CandyCane (Invitrogen), serve as molecular weight markers (bands reported in kDa). †Non-glycosylated marker proteins.

Figure 3. Conservation of the '*Fox*' Stimulon with key loci and homology. Names of genes are derived from previous work and non-named genes are colored according to their best match in a BLASTP search considering bi-directional queries. Where paralogs are assumed, genes are numbered based on their appearance. The entire stimulon is oriented based upon the detection of *foxA*. Boxing demonstrates gene sets that are syntenous in terms of content, orientation, and order among all species. The gene between *foxE* and *foxF* is an unannotated putative reading frame of conserved sequence identity.

Figure 4. Iron Biooxidation for Reported Iron Biooxidizing Species. Oxidation of soluble iron (10 mM at pH = 2 and 70˚C) over 16 h by reported metal-mobilizers from the order *Sulfolobales*. Quantification measured as total and ferrous soluble iron via 1,10-phenanthroline assay. A) Fraction of total iron oxidized, calculated as *[Initial concentration = 10 mM] – [Ferrous iron]*. B) Ratio of ferric to ferrous iron, calculated as ([Total Iron] – [Ferrous Iron])/[Ferrous Iron]. Oxygen consumption in the presence of iron over the course of 30 minutes. Oxidation cell was equilibrated with dry air-saturated DSM-88 medium (without carbon sources) at pH = 2 and 70˚C. Cells were added at 0 minutes and iron (10 mM) at 5 min. C) Trace of oxygen consumption, as a percent of original oxygen present. D) Rate of oxygen consumption in the presence of iron, over stable window of traces (10-15 minutes), with oxygen concentrations calculated via Henry's Law for air at relevant conditions. Negative controls were abiotic samples containing media without cells. Species abbreviations: *Acidianus brierleyi (Abri), Acidianus sulfidivorans (Asul), Metallosphaera hakonensis (Mhak), Metallosphaera prunae (Mpru), Metallosphaera sedula (Msed), Sulfolobus acidocaldarius (Saci), Sulfuracidifex metallicus (Smet), Sulfurisphaera ohwakuensis (Sohw),* and *Sulfurisphaera tokodaii (Stok)*

Figure 5. RT-PCR of *Fox* Cluster genes using RNA collected from heterotrophic growth conditions (further detailed in Methods). In each case, the first lane shows a ladder with the 300, 200, and 100 bp standards from 100 bp (NEB) ladder. No primers were designed to produce a product longer than 210 bp and smaller products occurred only when required to fit the primers within the annotated open reading frame of the gene. *fox* genes are labeled based on their appearance in the genome, where *M. sedula* and *M. hakonensis* contain two and three *fox*A gene duplicates, respectively. For *S. tokodaii*, two primers pairs were used for the *fox*C gene that cover the two putative domains of the gene which is dislocated by a frameshift mutation. The upper right insert shows the relative-fold transcription levels of *fox*A genes in *M. sedula* and *M. prunae*, via qPCR.

Figure 6. GC Content of the *Fox* Locus in the Sulfolobaceae. Analysis of GC content in coding regions and each of the three nucleotide positions of codons. *Fox* genes are shown in red for each species. All data points are the collective mean of dinucleotide content for a particular ORF, as predicted through the Prokka pipeline. Whiskers extend 1.5*IQR (Interquartile range) from the 1st and 3rd quartile values. Abbreviations are *Acidianus brierleyi (Abri),* Candidatus *Acidianus copahuensis (Acop), Acidianus manzaensis (Aman), Acidianus sulfidivorans (Asul), Metallosphaera hakonensis (Mhak), Metallosphaera prunae (Mpru), Metallosphaera sedula (Msed), Metallosphaera yellowstonensis (Myel), Sulfodiicoccus acidiphilus (Sacd), Sulfuracidifex metallicus (Smet), Sulfurisphaera tokodaii (Stok).*

Figure 7. Phylogenetic Analysis of FoxA and FoxC. BLASTP searches were performed using Msed_RS02490 (FoxA2) and Msed_RS02455 (FoxC) as queries. In both cases, data sets were curated to minimize excessive and spurious comparisons. For FoxA2, subjects had to cover at least 50% of the query and produce an E-value below 10^{-14}. For FoxC, only a query cutoff of 40% was used, removing a few partial domain proteins. Both trees, generated by MUSCLE alignment and FastTree are unrooted. Each region has dots showing the constituents on the basis of major phyla (color), while the boundaries show kingdom (solid:bacteria; dotted:archaea). “Others” was used as a designation for taxa with only a single constituent or for unclassified/environmental samples. Stars are meant to guide the reader to the relevant insert and show where the branches originate in the larger topography, focusing on close homologues of FoxA2 and FoxC. Yellow triangle showing positioning of FoxA2 and purple triangle showing the positioning of FoxA1. Distance scale bars represent estimated substitutions per site.

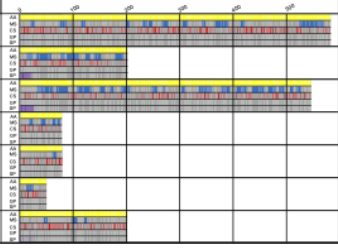
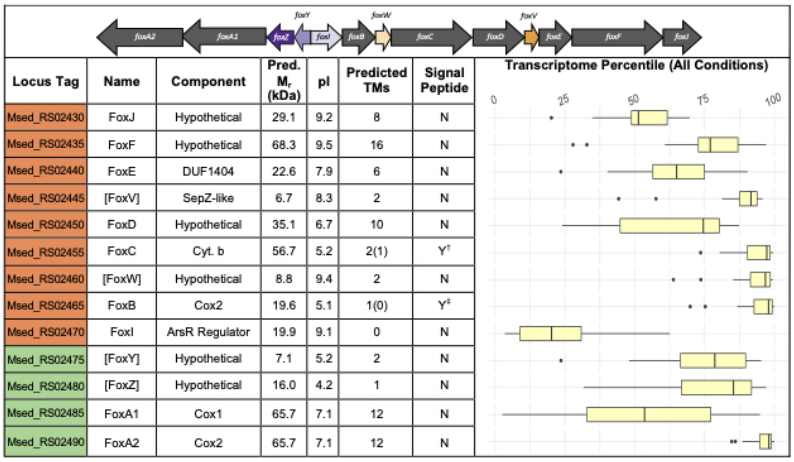
			Membrane Fraction			High MW Complex		Low MW Complex	
Protein Name	Locus ID	Mr (kDa)	Coverage (%)	#PSMs	Sequence Analysis	Coverage (%)	#PSMs	Coverage (%)	#PSMs
FoxA2	Msed_RS02490	65.7	29	297		7	16	24	134
FoxB	Msed_RS02465	19.6	58	223				11	6
FoxC	Msed_RS02455	56.7	49	554		3	9	28	86
FoxY	Msed_RS02475	7.1	46	60					
FoxW	Msed_RS02460	8.8	10	2					
FoxV	Msed_RS02445	7.1	21	6					
FoxE	Msed_RS02440	22.6	12	16					

Table 1. Mass Spectrometry and Amino Acid Sequence Analysis of M. sedula fox Genes. List of fox genes and their detection by liquid chromatography mass spectrometry in the membrane fraction and both complex bands from BN-PAGE. Sequence analysis shows positional data for sequence length in amino acids in yellow (AA), LC/MS-detected peptide spectral matches (PSMs) in blue (MS), position of trypsin/chymotrypsin-digestible sites in red (CS), likelihood of N-linked glycosylation site based on the prokaryotic-trained algorithm GlycoPPv1.0 (GP), and SEC signal peptide prediction likelihood in purple based on the SignalPv5.0 "Archaeal" algorithm (SP). For each sample, the coverage of the full amino acid sequence in PSMs and the number of PSMs are reported.



Locus Tag	Name	Component	Pred. M, (kDa)	pI	Predicted TMs	Signal Peptide
Msed_RS02430	FoxJ	Hypothetical	29.1	9.2	8	N
Msed_RS02435	FoxF	Hypothetical	68.3	9.5	16	N
Msed_RS02440	FoxE	DUF1404	22.6	7.9	6	N
Msed_RS02445	[FoxV]	SepZ-like	6.7	8.3	2	N
Msed_RS02450	FoxD	Hypothetical	35.1	6.7	10	N
Msed_RS02455	FoxC	Cyt. b	56.7	5.2	2(1)	Y†
Msed_RS02460	[FoxW]	Hypothetical	8.8	9.4	2	N
Msed_RS02465	FoxB	Cox2	19.6	5.1	1(0)	Y‡
Msed_RS02470	FoxI	ArsR Regulator	19.9	9.1	0	N
Msed_RS02475	[FoxY]	Hypothetical	7.1	5.2	2	N
Msed_RS02480	[FoxZ]	Hypothetical	16.0	4.2	1	N
Msed_RS02485	FoxA1	Cox1	65.7	7.1	12	N
Msed_RS02490	FoxA2	Cox2	65.7	7.1	12	N

Figure 1. Fox Cluster in M. sedula. Annotations are based on S. metallicus transcriptome, for genes responsive to iron and BLASTP searches. Mr and pI were calculated in the absence of putative cleavage sequences using EXPASY. Transmembrane (TM) domains were determined with TMHMM v. 2.0 and Signal peptides (SP) were determined with SignalP v. 5.0 with setting "Archaea." Reported transcriptional data were compiled from previous M. sedula transcriptional studies. For signal peptides, †FoxC has a putative Sec/SPI site at the sequence AYG^AD with 0.9193 likelihood, ‡FoxB has a putative Sec/SPI site at the sequence LET^QY with 0.8257 likelihood, all other likelihoods were below 0.1. Values in parenthesis for TMs represent values minus the signal peptide region. Bracketed gene names are previously unnamed genes, conserved in all species from this study. Orange shading is 5' to 3', green shading is 3' to 5'.

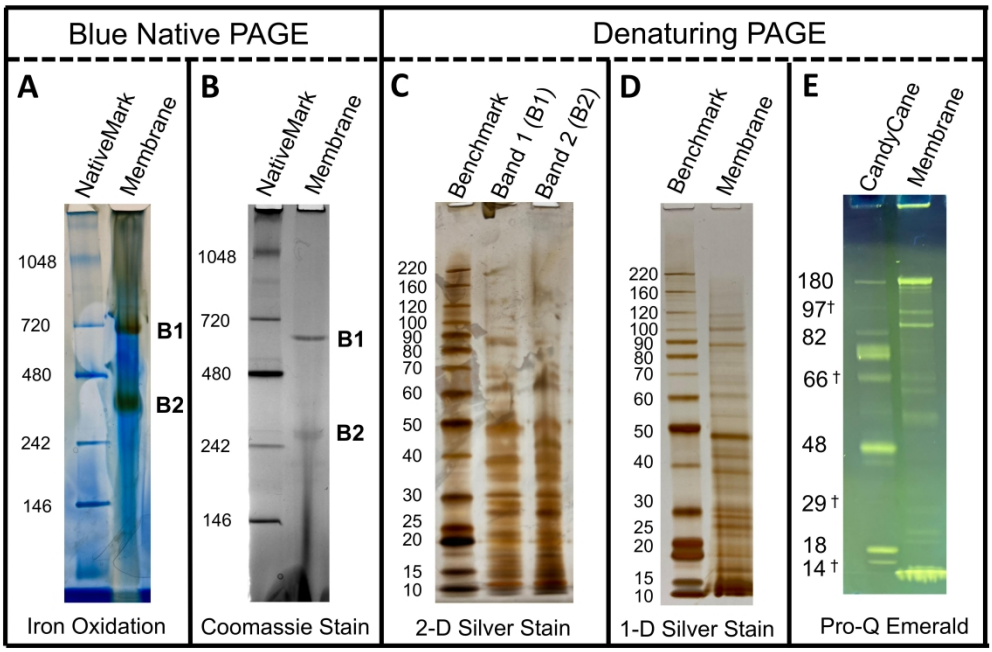


Figure 2. Gel Electrophoresis of M. sedula membrane fractions. (A) BN-PAGE of membrane fraction with ferrous iron/glycine buffer under physiologically-relevant conditions: 65 °C and pH ~ 3.0. (B) BN-PAGE of membrane fraction stained with Coomassie Brilliant Blue. (C) LDS-PAGE of bands from Blue-Native PAGE gel (B) stained with Coomassie Brilliant Blue. (D) LDS-PAGE of original membrane preparation (1-dimension). E) Glycoprotein staining of membrane preparation under denaturing conditions with Pro-Q Emerald 300.

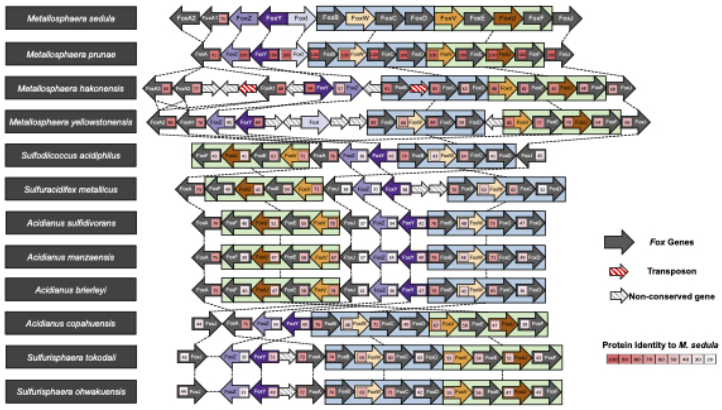


Figure 3. Conservation of the 'Fox' Stimulon with key loci and homology. Names of genes are derived from previous work and non-named genes are colored according to their best match in a BLASTP search considering bi-directional queries. Where paralogs are assumed, genes are numbered based on their appearance. The entire stimulon is oriented based upon the detection of foxA. Boxing demonstrates gene sets that are syntenous in terms of content, orientation, and order among all species. The gene between foxE and foxF is an unannotated putative reading frame of conserved sequence identity.

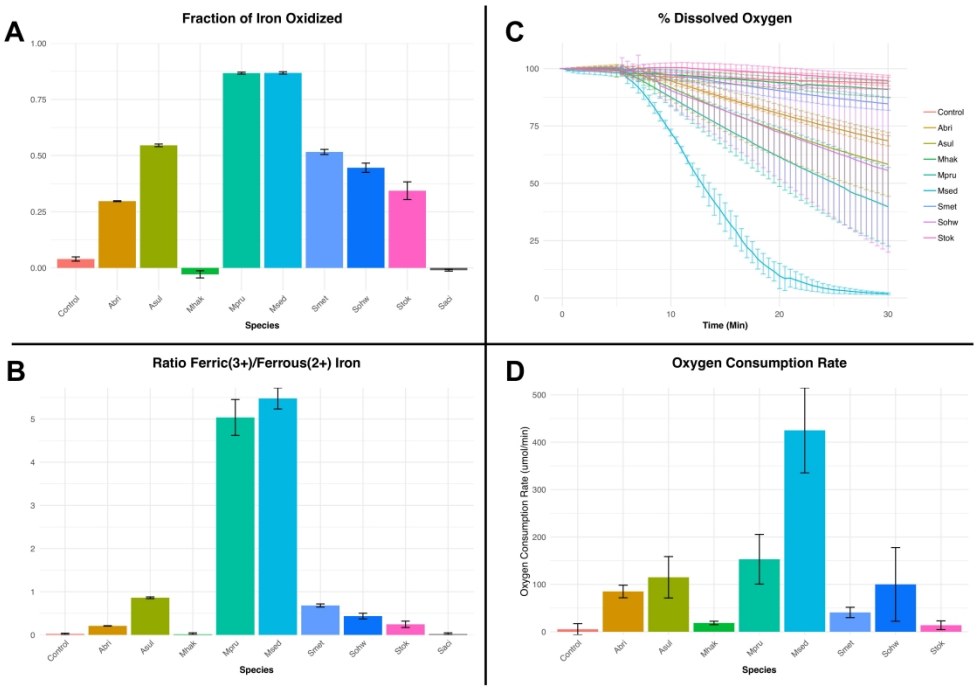


Figure 4. Iron Biooxidation for Reported Iron Biooxidizing Species. Left: Oxidation of soluble iron (10 mM) at pH = 2 and 70°C) over 16 h by reported metal-mobilizers from the order Sulfolobales. Quantification measured as total and ferrous soluble iron via 1,10-phenanthroline assay (error bars are based on standard deviation of n = 3 replicates). A) Fraction of total iron oxidized, calculated as [Initial concentration = 10 mM] – [Ferrous iron] (error bars are standard deviation of n=3 replicates). B) Ratio of ferric to ferrous iron, calculated as ([Total Iron] – [Ferrous Iron])/[Ferrous Iron]. Right: Oxygen consumption in the presence of iron over the course of 30 minutes. Oxidation cell was equilibrated with dry air-saturated DSM-88 medium (without carbon sources) at pH = 2 and 70°C. Cells were added at 0 minutes and iron (10 mM) at 5 min. C) Trace of oxygen consumption, as a percent of original oxygen present. D) Rate of oxygen consumption in the presence of iron, over stable window of traces (10-15 minutes), with oxygen concentrations calculated via Henry's Law for air at relevant conditions. Negative controls were abiotic samples containing media without cells. Species abbreviations: Acidianus brierleyi (Abri), Acidianus sulfidivorans (Asul), Metallosphaera hakonensis (Mhak), Metallosphaera prunae (Mpru), Metallosphaera sedula (Msed), Sulfolobus acidocaldarius (Saci), Sulfolobus metallicus (Smet), Sulfurisphaera ohwakuensis (Sohw), and Sulfurisphaera tokodaii (Stok)

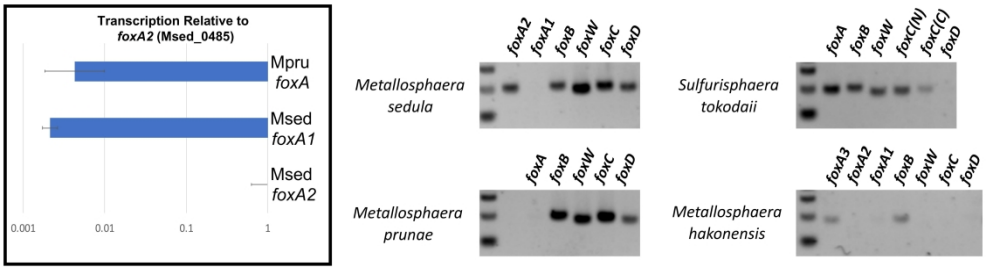


Figure 5. RT-PCR of Fox Cluster genes using RNA collected from 'normal' growth conditions (further detailed in Methods). In each case, the first lane shows a ladder with the 300, 200, and 100 bp standards from 100 bp (NEB) ladder. No primers were designed to produce a product longer than 210 bp and smaller products occurred only when required to fit the primers within the annotated open reading frame of the gene. fox genes are labeled based on their appearance in the genome, where M. sedula and M. hakonensis contain two and three foxA gene duplicates, respectively. For S. tokodaii, two primers pairs were used for the foxC gene that cover the two putative domains of the gene which is dislocated by a frameshift mutation. The upper right insert shows the relative-fold transcription levels of foxA genes in M. sedula and M. prunae, via qPCR.

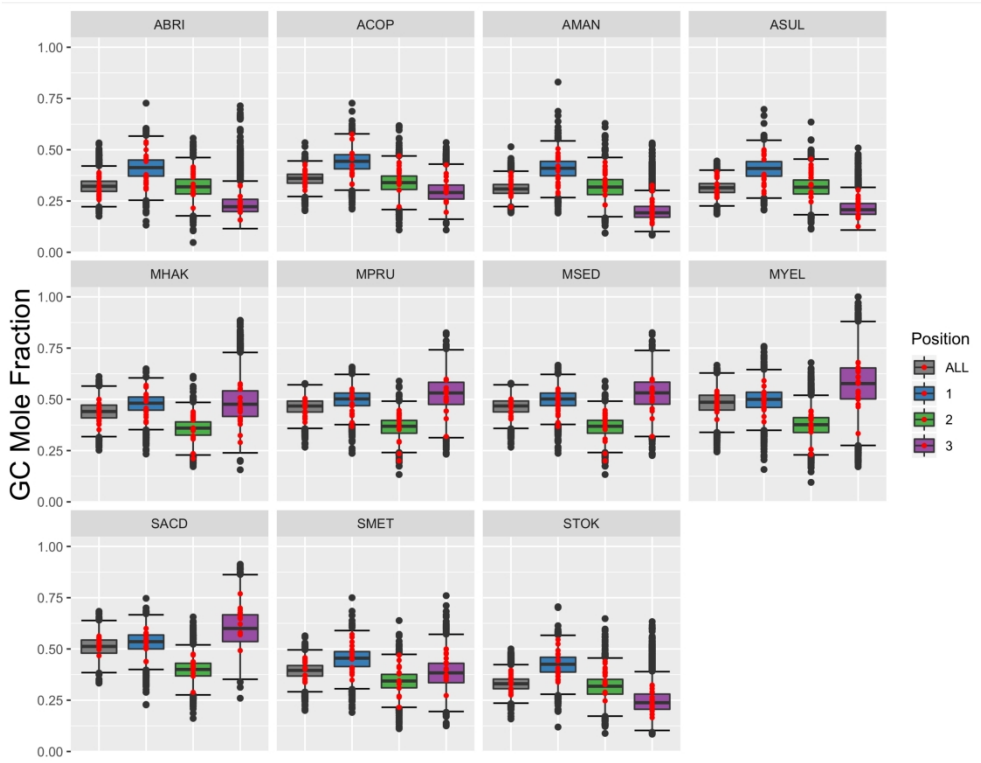


Figure 6. GC Content of the Fox Locus in the Sulfolobaceae. Analysis of GC content in coding regions and each of the three nucleotide positions of codons. Fox genes are shown in red for each species. All data points are the collective mean of dinucleotide content for a particular ORF, as predicted through the Prokka pipeline. Whiskers extend 1.5*IQR (Interquartile range) from the 1st and 3rd quartile values. Abbreviations are Acidianus brierleyi (Abri), Candidatus Acidianus copahuensis (Acop), Acidianus manzaensis (Aman), Acidianus sulfidivorans (Asul), Metallosphaera hakonensis (Mhak), Metallosphaera prunae (Mpru), Metallosphaera sedula (Msed), Metallosphaera yellowstonensis (Myel), Sulfodiicoccus acidiphilus (Sacd), Sulfolobus metallicus (Smet), Sulfurisphaera tokodaii (Stok).

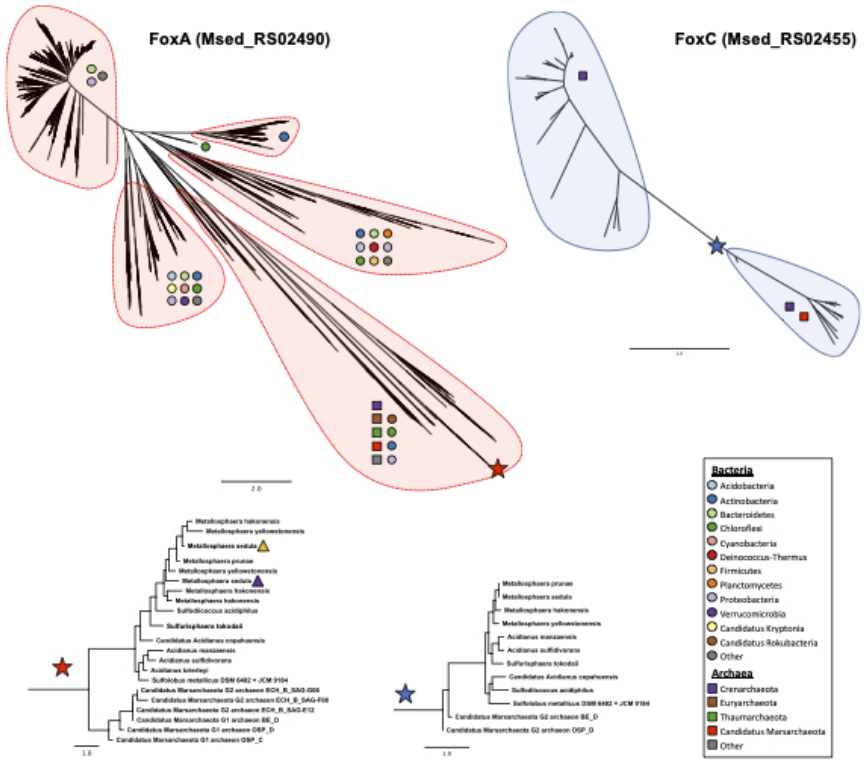


Figure 7. Phylogenetic Analysis of FoxA and FoxC. BLASTP searches were performed using Msed_RS02490 (FoxA2) and Msed_RS02455 (FoxC) as queries. In both cases, data sets were curated to minimize excessive and spurious comparisons. For FoxA2, subjects had to cover at least 50% of the query and produce an E-value below 10-14. For FoxC, only a query cutoff of 40% was used, removing a few partial domain proteins. Both trees, generated by MUSCLE alignment and FastTree are unrooted. Each region has dots showing the constituents on the basis of major phyla (color), while the boundaries show kingdom (solid:bacteria; dotted:archaea). “Others” was used as a designation for taxa with only a single constituent or for unclassified/environmental samples. Stars are meant to guide the reader to the relevant insert and show where the branches originate in the larger topography, focusing on close homologues of FoxA2 and FoxC. Yellow triangle showing positioning of FoxA2 and purple triangle showing the positioning of FoxA1. Distance scale bars represent estimated substitutions per site.