ORIGINAL ARTICLE

The methanogenic redox cofactor $F_{420}$ is widely synthesized by aerobic soil bacteria

Blair Ney$^{1,2,5}$, F Hafna Ahmed$^{1,5}$, Carlo R Carere$^{3}$, Ambarish Biswas$^{4}$, Andrew C Warden$^{2}$, Sergio E Morales$^{4}$, Gunjan Pandey$^{2}$, Stephen J Watt$^{1}$, John G Oakeshott$^{2}$, Matthew C Taylor$^{5}$, Matthew B Stott$^{3}$, Colin J Jackson$^{1}$ and Chris Greening$^{2,6}$

$^{1}$Research School of Chemistry, Australian National University, Acton, Australian Capital Territory, Australia; $^{2}$The Commonwealth Scientific and Industrial Research Organisation, Land and Water Flagship, Clunies Ross Street, Acton, Australian Capital Territory, Australia; $^{3}$GNS Science, Wairakei Research Centre, Taupō, New Zealand and $^{4}$Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

$F_{420}$ is a low-potential redox cofactor that mediates the transformations of a wide range of complex organic compounds. Considered one of the rarest cofactors in biology, $F_{420}$ is best known for its role in methanogenesis and has only been chemically identified in two phyla to date, the Euryarchaeota and Actinobacteria. In this work, we show that this cofactor is more widely distributed than previously reported. We detected the genes encoding all five known $F_{420}$ biosynthesis enzymes ($cofC$, $cofD$, $cofE$, $cofG$ and $cofH$) in at least 653 bacterial and 173 archaeal species, including members of the dominant soil phyla Proteobacteria, Chloroflexi and Firmicutes. Metagenome datamining validated that these genes were disproportionately abundant in aerated soils compared with other ecosystems. We confirmed through high-performance liquid chromatography analysis that aerobically grown stationary-phase cultures of three bacterial species, Paracoccus denitrificans, Oligotropha carboxidivorans and Thermomicrobium roseum, synthesized $F_{420}$, with oligoglutamate sidechains of different lengths. To understand the evolution of $F_{420}$ biosynthesis, we also analyzed the distribution, phylogeny and genetic organization of the $cof$ genes. Our data suggest that although the $F_{o}$ precursor to $F_{420}$ originated in methanogens, $F_{420}$ itself was first synthesized in an ancestral actinobacterium. $F_{420}$ biosynthesis genes were then disseminated horizontally to archaea and other bacteria. Together, our findings suggest that the cofactor is more significant in aerobic bacterial metabolism and soil ecosystem composition than previously thought. The cofactor may confer several competitive advantages for aerobic soil bacteria by mediating their central metabolic processes and broadening the range of organic compounds they can synthesize, detoxify and mineralize.

The ISME Journal (2017) 11, 125–137; doi:10.1038/ismej.2016.100; published online 9 August 2016

Introduction

$F_{420}$ serves as the cofactor in the catalysis of some of the most chemically demanding redox reactions in biology. Among them are the one-carbon reactions of methanogenesis (Thauer, 1998; Shima et al., 2000; Hagemeier et al., 2003), the biosynthesis pathways of tetracycline antibiotics (Wang et al., 2013) and the biodegradation of pircate and aflatoxins (Ebert et al., 2001; Taylor et al., 2010; Lapalikar et al., 2012). The cofactor appears to have been selected for these roles because of its unique electrochemical properties compared with the ubiquitous flavin and nicotinamide cofactors FMN (flavin mononucleotide), FAD (flavin adenine dinucleotide) and NAD(P) (nicotinamide adenine dinucleotide (phosphate)) (Walsh, 1986; Greening et al., 2016a). As a 5-deazaflavin, $F_{420}$ is structurally and biosynthetically related to FMN and FAD, but exhibits distinct electrochemical properties because of several key substitutions. It has a relatively low redox potential of $\sim$340 mV under standard conditions and $\sim$380 mV under certain physiological conditions (de Poorter et al., 2005). This enables reduced $F_{420}$ ($F_{420}H_2$) to reduce a wide range of organic compounds otherwise recalcitrant to activation (Jacobson and Walsh, 1984; Greening et al., 2016a). As an obligate two-electron carrier, the cofactor can transform alkene, alkyne, alcohol and imine groups through hydride transfer reactions (Shima et al., 2000; Hagemeier et al., 2003;
Aufhammer et al., 2004; Shen et al., 2009; Wang et al., 2013.

Half a century since its discovery (Cheeseman et al., 1972), F<sub>420</sub> is still perceived to be a rare cofactor synonymous with methanogenesis (Greening et al., 2016a). It has been confirmed to be synthesized in just two phyla to date: the Euryarchaeota (Eirich et al., 1979) and Actinobacteria (Eker et al., 1980; Daniels et al., 1985). It is thought to serve as the primary catabolic electron carrier in multiple lineages of Euryarchaeota, including representatives of the methanogenic (Eirich et al., 1979), methanotrophic (Michaelis et al., 2002; Knittel et al., 2005) and sulfate-reducing orders (Lin and White, 1986). Genomic and spectroscopic evidence suggests that the cofactor is also synthesized in the aerobic ammonia-oxidizing phylum Thaumarchaeota (Spang et al., 2012). Among bacteria, the cofactor has been chemically identified only within the Actinobacteria, where its physiological roles remain under investigation. In these organisms, F<sub>420</sub> reduction is coupled to either glucose 6-phosphate or NADPH oxidation and hence is dependent on the pentose phosphate pathway (Greening et al., 2016a). The reduced cofactor (F<sub>420H</sub>) is reported to enhance the metabolic flexibility of mycobacteria by facilitating the catalysis of a wide range of reductions of endogenous and exogenous organic compounds (Ahmed et al., 2015; Greening et al., 2016a). F<sub>420</sub> also has roles in antibiotic synthesis and xenobiotic degradation in species of Streptomyces (Wang et al., 2013), Rhodococcus (Heiss et al., 2002) and Nocardioides (Ebert et al., 1999).

F<sub>420</sub> is synthesized in three major steps in bacteria and archaea (Figure 1). In the first, a riboflavin precursor (5-amino-6-(D-ribitylamino)uracil) is condensed with tyrosine to form 8-hydroxy-5-deazaflavin, also known as F<sub>L</sub>; this step is catalyzed by the radical S-adenosylmethionine enzymes CofG and CofH that are fused into a single protein in some bacteria (known as CofGH or FbiC) (Choi et al., 2002; Philmus et al., 2015). Subsequently, LPPG (L-lactyl-2-diphospho-5'-guanosine) is proposed to be synthesized from 2-phospho-L-lactate by CofC (Grochowski et al., 2008) and transferred to F<sub>L</sub> by CofD (also known as FbiA) (Choi et al., 2001; Graupner and White, 2001; Graupner et al., 2002). The resulting LPPG sidechain is finally elongated with glutamate residues by the F<sub>420</sub>-γ-L-glutamyl ligase CofE (also known as FbiB) (Choi et al., 2001; Li et al., 2003; Nocek et al., 2007) that is fused with an FMN-dependent oxidoreductase in Actinobacteria (Bashiri et al., 2016). For reasons still not understood, the number of glutamate residues added varies between organisms, ranging from two to three in most methanogens (Gorris and van der Drift, 1994), four to five in Methanosarcina (Gorris and van der Drift, 1994) and five to seven in Mycobacterium (Bair et al., 2001). In addition to being an obligate intermediate in the F<sub>420</sub> biosynthesis pathway, F<sub>L</sub> is also synthesized independently by bacteria, archaea and eukaryotes for use as an antennal chromophore in DNA repair photolyases (Eker et al., 1988; Yasui et al., 1988; Kiener et al., 1989; Eppele and Carell, 1998).

We recently proposed that F<sub>420</sub> could be synthesized in a wider range of bacteria than currently described in the literature (Greening et al., 2016a). Our analysis of a protein superfamily, the flavin/deazaflavin oxidoreductases, identified putative F<sub>420</sub>-utilizing oxidoreductases in microorganisms other than the Euryarchaeota and Actinobacteria, including Chloroflexi, Proteobacteria and Firmicutes (Ahmed et al., 2015). In this work, we explored the distribution of the genes encoding the F<sub>420</sub> biosynthesis enzymes CofC, CofD, CofE, CofG and CofH in public genomes and metagenomes. This revealed that the genes required to synthesize F<sub>420</sub> are encoded in a broad range of aerobic bacteria and are widespread in soil and aquatic ecosystems. Using this information, we validated through pure culture studies that three representatives of the dominant soil phyla Proteobacteria and Chloroflexi synthesize F<sub>420</sub>. We propose that F<sub>420</sub> is much more widely distributed in microorganisms than previously reported and present a model of the evolution of the F<sub>L</sub> and F<sub>420</sub> biosynthesis pathways to explain the origin and dispersal of this cofactor.

Materials and methods

Gene sequence retrieval
The amino acid sequences of all known F<sub>420</sub> biosynthesis enzymes (CofC, CofD, CofE, CofG and CofH) represented in the NCBI (National Center for Biotechnology Information) Reference Sequence (RefSeq) database (Pruitt et al., 2007) were retrieved by Protein BLAST and PSI-BLAST (Altschul et al., 1990). The homologous proteins in Aigarchaeota, Bathyarchaeota, Geoarchaeota, Lokiarchaeota and Tectomicrobia were retrieved from the Joint Genome Institute’s Integrated Microbial Genomes database (Markowitz et al., 2012). Taxonomic annotations for the sequences were obtained from the NCBI Taxonomy database and sequences duplicated at the species level were deleted. Clustering on sequence similarity networks (Atkinson et al., 2009) generated using the Enzyme Function-Initiative Enzyme Similarity Tool (Gerlt et al., 2015) were used to identify homologs of characterized proteins from nonspecific hits. In this analysis, nodes represent individual proteins and edges represent the all-versus-all BLAST E-values (Altschul et al., 1990) between them. Closely related proteins form visual clusters, allowing the identification of sequences belonging to a protein family from those belonging to related families (Atkinson et al., 2009). Final sequence sets were obtained by decreasing logE-value cutoffs until no major changes in clustering were observed with large increases in cutoff value. Final logE-value...
cutoffs used to identify sequences were \(-20\) for CofC, \(-52\) for CofE and \(-60\) for CofD, CofG and CofH.

**Evolutionary analysis**
For robust phylogenetic tree construction, representative smaller sequence sets were generated by removing any sequences with >90% sequence identity using CD-HIT (Fu et al., 2012). This was first done for CofH and the same taxa as in the resulting sequence set were used for the phylogenetic analysis of the other proteins for consistency. RogueNaRok (Aberer et al., 2013) was used to remove fast-evolving sequences (all belonging to Firmicutes, Tectomicrobia, Thermoleophilia and Rubrobacteria) that otherwise caused long-branch attractions and hence unreliable evolutionary inferences (Anderson and Swofford, 2004). Sequences were aligned using MAFFT (Katoh and Standley, 2013) or MUSCLE (Edgar, 2004), with gaps and poorly aligned variable regions in the alignment removed using Gblocks (Castresana, 2000). Approximate maximum-likelihood trees were generated using Fasttree 2 (Price et al., 2010), with the JTT +CAT evolutionary model for 100 bootstrap replicates that were generated using SeqBoot in the Phylip package (Felsenstein, 2005). To analyze co-evolution of the F\(_{420}\) biosynthesis proteins, phylogenetic tree topologies were compared and correlated using MirrorTree (Kann et al., 2009). Genetic organization was compared using the Microbial Genomic Context Viewer (Overmars et al., 2013) and the Integrated Microbial Genomes database (Markowitz et al., 2012).

**Motif analysis and homology modeling**
BLAST and PSI-BLAST (Altschul et al., 1990) were used to retrieve sequences encoding probable F\(_{420}\)-dependent oxidoreductases from the NCBI reference genomes of *Oligotropha carboxidovorans*, *Paracoccus denitrificans* and *Thermomicrobium roseum*. Characterized representatives of the 20 previously described F\(_{420}\)-dependent oxidoreductase enzyme families were used as seed sequences (Greening et al., 2016a). Phyre2 (Kelley et al., 2015) was used to model protein structures based on these sequences based on solved protein structures with highest percentage amino acid sequence identity. The quality of the models at the global (full protein) and local (F\(_{420}\)-binding site) scales were assessed using ProQ2 (Ray et al., 2012). F\(_{420}\)-binding motifs were identified based on previous studies (Eguchi et al., 1984; Purwanti and Daniels, 1998; Aufhammer et al., 2004; Ahmed et al., 2015).
Evolution and distribution of the cofactor F$_{420}$
B Ney et al

128

Metagenome analysis
Metagenomes were screened for the presence of F$_{420}$ biosynthesis genes via a translated BLAST screen against the reference database as previously described (Greening _et al_., 2016b). In all, 19 publicly available metagenomes (11 ecosystem types) were randomly subsampled to an equal depth of 4 million reads with minimum read length >140 nucleotides. To remove false positives, hits within the initial screen were further sieved by removing any result with a minimum percentage identity <60 or minimum query coverage <40 amino acids.

Bacterial culturing and harvesting
_O. carboxidovorans_ OMS (Meyer and Schlegel, 1978) was grown on carbon monoxide oxidizer media supplemented with 36 mM sodium acetate (Meyer and Schlegel, 1978) and maintained on solid carbon monoxide oxidizer media supplemented with 1.2% (w/v) agar. _P. denitrificans_ strain PD1222 (de Vries _et al_., 1989) was grown and maintained in lysogeny broth liquid media and agar. _Mycobacterium smegmatis_ mc²155 (Snapper _et al_., 1990) was grown and maintained in lysogeny broth liquid media and agar supplemented with 0.05% (v/v) Tween 80. _T. roseum_ (Jackson _et al_., 1973) cultures were grown in Castenholz salts solution supplemented with 5 g l$^{-1}$ peptone and 2.5 g l$^{-1}$ sucrose (Houghton _et al_., 2015). Liquid cultures (500 ml) of _O. carboxidovorans_PD1243, _P. denitrificans_ and _M. smegmatis_ were grown to early stationary-phase in 2-l Erlenmeyer flasks in a rotary incubator (200 r.p.m., 37 °C). _T. roseum_ was grown in an equivalent manner in 1-l cultures at 130 r.p.m. and 68 °C. Cells from stationary-phase liquid cultures were harvested by centrifugation (4 °C, 10 000 g, 20 min) and the supernatant was discarded. Pellets were resuspended in 1 to 3 ml 50 mM sodium phosphate buffer (pH 7.0) and lysed by boiling. Lysates were centrifuged at 11 000 g for 10 min, and the supernatant was removed for analysis.

HPLC and liquid chromatography/mass spectrometry analysis
F$_{420}$ was detected by ion-pair reversed-phase high-performance liquid chromatography (HPLC) using an Agilent (Santa Clara, CA, USA) 1200 series HPLC system equipped with an autosampler, fluorescence detector and a Poroshell 120 EC-C18 2.1 × 100 mm, 2.7 µm column. Gradients of two HPLC buffers were used, A (20 mM ammonium phosphate, 10 mM tetra-butylammonium phosphate, pH 7.0) and B (100% acetonitrile), to separate F$_{420}$ species at high resolution by the length of their oligoglutamate tails. The applied gradient was 0–1 min 25% B; 1–10 min from 25% to 35% B; 10–13 min 35% B; 13–16 min from 35% to 40% B; 16–19 min from 40% to 25% B. Columns were extensively washed between sample runs, and test runs validated there was no carryover of F$_{420}$. For fluorescence detection, the samples were excited at 420 nm and emission spectra were recorded between 470 and 600 nm.

The genetic determinants of F$_{420}$ biosynthesis are widely encoded by aerobic soil bacteria
We initially retrieved the sequences of all F$_{420}$ biosynthesis genes in publicly available genomes in order to understand the distribution and evolution of F$_{420}$ in microorganisms (Supplementary Table S1). A total of 653 bacterial and 173 archaeal species named in the NCBI Reference Sequence database encoded all five proteins specifically required to synthesize F$_{420}$ (CoF, CoFΔ, CoFE, CoFG and CoFH) (Supplementary Table S2).

Among archaea, the genes encoding F$_{420}$ biosynthesis enzymes were unsurprisingly widespread in the Euryarchaeota, specifically in all six validated methanogenic orders, Halobacteria and Archaeoglobi. Biosynthesis genes were also detected in full suites in the Thaumarchaeota and Geoarchaeota, and in partial repertoires in the metagenome-derived incomplete genomes of Aigarchaeota, Bathyarchaeota and Lokiarchaeota representatives (Figure 2a and Supplementary Table S3). Among bacteria, F$_{420}$ biosynthesis genes were present in the majority of sequenced Actinobacteria, as well as multiple species within the phyla Proteobacteria (classes: Alpha, Beta and Gamma) and Chloroflexi (classes: Thermomicrobia, Ktedonobacteria and Ardenticatenia) (Figure 2a and Supplementary Table S3). Within these phyla often predominating in soil environments, species harboring _cof_ genes were obligate aerobes and facultative anaerobes rather than obligate anaerobes. For example, these genes were widespread in the aerobic lineages of Chloroflexi, but not anaerobic lineages such as the Dehalococci. In addition, _cof_ genes were disproportionally abundant in known soil- and marine-dwelling orders of Proteobacteria (for example, Rhizobiales and Alteromonadales) compared with host-associated orders (for example, Rickettsiales and Enterobacteriales) (Supplementary Table S1). Finally, F$_{420}$ biosynthesis genes were also detected in two Firmicutes species and within the Tectomicrobia (Supplementary Table S3), a recently discovered relative abundance of F$_{420}$ species, concentrated F$_{420}$ standard was purified from a recombinant F$_{420}$ overexpression strain of _M. smegmatis_ mc²4517 (Bashiri _et al_., 2010) as previously described (Isabelle _et al_., 2002). The standard was lyophilized for storage, resuspended in 50 µl ultrapure water and serially diluted. The sample was analyzed on a Waters (Milford, MA, USA) LCT Premier OA-TOF (orthogonal acceleration time-of-flight) mass spectrometer. Samples were ionized by electrospray in negative ion mode at a rate of 5 µl per min with a capillary voltage of 2500 V, desolvation temperature of 150 °C and source temperature of 100 °C. Deconvolution of the mass data was facilitated by MaxEnt software (Princeton, NJ, USA).

Results

The genetic determinants of F$_{420}$ biosynthesis are widely encoded by aerobic soil bacteria

We initially retrieved the sequences of all F$_{420}$ biosynthesis genes in publicly available genomes in order to understand the distribution and evolution of F$_{420}$ in microorganisms (Supplementary Table S1). A total of 653 bacterial and 173 archaeal species named in the NCBI Reference Sequence database encoded all five proteins specifically required to synthesize F$_{420}$ (CoF, CoFΔ, CoFE, CoFG and CoFH) (Supplementary Table S2).

Among archaea, the genes encoding F$_{420}$ biosynthesis enzymes were unsurprisingly widespread in the Euryarchaeota, specifically in all six validated methanogenic orders, Halobacteria and Archaeoglobi. Biosynthesis genes were also detected in full suites in the Thaumarchaeota and Geoarchaeota, and in partial repertoires in the metagenome-derived incomplete genomes of Aigarchaeota, Bathyarchaeota and Lokiarchaeota representatives (Figure 2a and Supplementary Table S3). Among bacteria, F$_{420}$ biosynthesis genes were present in the majority of sequenced Actinobacteria, as well as multiple species within the phyla Proteobacteria (classes: Alpha, Beta and Gamma) and Chloroflexi (classes: Thermomicrobia, Ktedonobacteria and Ardenticatenia) (Figure 2a and Supplementary Table S3). Within these phyla often predominating in soil environments, species harboring _cof_ genes were obligate aerobes and facultative anaerobes rather than obligate anaerobes. For example, these genes were widespread in the aerobic lineages of Chloroflexi, but not anaerobic lineages such as the Dehalococci. In addition, _cof_ genes were disproportionally abundant in known soil- and marine-dwelling orders of Proteobacteria (for example, Rhizobiales and Alteromonadales) compared with host-associated orders (for example, Rickettsiales and Enterobacteriales) (Supplementary Table S1). Finally, F$_{420}$ biosynthesis genes were also detected in two Firmicutes species and within the Tectomicrobia (Supplementary Table S3), a recently discovered relative abundance of F$_{420}$ species, concentrated F$_{420}$ standard was purified from a recombinant F$_{420}$ overexpression strain of _M. smegmatis_ mc²4517 (Bashiri _et al_., 2010) as previously described (Isabelle _et al_., 2002). The standard was lyophilized for storage, resuspended in 50 µl ultrapure water and serially diluted. The sample was analyzed on a Waters (Milford, MA, USA) LCT Premier OA-TOF (orthogonal acceleration time-of-flight) mass spectrometer. Samples were ionized by electrospray in negative ion mode at a rate of 5 µl per min with a capillary voltage of 2500 V, desolvation temperature of 150 °C and source temperature of 100 °C. Deconvolution of the mass data was facilitated by MaxEnt software (Princeton, NJ, USA).
candidate phylum with wide biosynthetic capacity (Wilson et al., 2014). Putative F$_{420}$-dependent oxidoreductases from multiple families were identified in the genomes of the phyla predicted to synthesize F$_{420}$ (Supplementary Table S4).

We also analyzed the distribution of the two genes (cofG and cofH) required for synthesis of F$_o$, both an independent chromophore and an F$_{420}$ precursor (Figure 2a and Supplementary Table S3). It has been postulated that F$_o$ is universally distributed given it is used in DNA photolyases across the three domains of life (Kiener et al., 1989; Mees et al., 2004; Petersen and Ronan, 2010). Our analysis revealed, however, that cofG and cofH genes were almost exclusively encoded by predicted F$_{420}$ producers, that is, those organisms that harbor all five F$_{420}$ biosynthesis genes. The only exceptions were Cyanobacteria, Chlorophyta and Streptophyta (Supplementary Table S1) that encoded the genetic determinants of F$_o$ biosynthesis (cofG and cofH) but not F$_{420}$ synthesis (cofC, cofD and cofE). These three lineages comprise oxygenic phototrophs known to encode F$_o$-utilizing photolyases (Mees et al., 2004; Petersen and Ronan, 2010). As previously observed (Graham et al., 2003), the cofG and cofH genes are fused into a single open reading frame (referred hereafter as cofGH) in most Actinobacteria, some Proteobacteria and phototrophic eukaryotes.

To explore the ecological role of F$_{420}$, we surveyed publicly available metagenomes for the genetic determinants of F$_{420}$ synthesis (Supplementary Table S5). In total, a full complement of F$_{420}$ biosynthesis genes were identified in all 17 soil, water and sediment metagenomes analyzed (Supplementary Table S5). These genes were especially abundant in the aerated forest and agricultural soil samples, accounting for ~0.02% of total sequence reads (Figure 2b). Consistent with the microbial community composition of soil ecosystems (Janssen, 2006), the majority of the identified reads matched those from Actinobacteria, Alphaproteobacteria and Chloroflexi (Figure 2c). Based on metagenomes, F$_{420}$ biosynthesis genes were also abundant in eight aquatic environments. In these ecosystems, genes encoding F$_o$ synthases (cofG and cofH) were 1.4 to 1.8 times more abundant relative to those encoding F$_{420}$-specific biosynthesis enzymes (cofC, cofD and cofE) (Figure 2b), suggesting that these environments harbor large communities of both F$_o$-utilizing phototrophs and F$_{420}$-synthesizing chemotrophs (Figure 2c). F$_{420}$ biosynthesis genes closely related to those encoded in Euryarchaeota were unsurprisingly common in marine sediments known to harbor high concentrations of methanocycling archaea (Figure 2c). Consistent with their community composition, F$_{420}$ biosynthesis genes were detected in very low abundance within gut ecosystems (Supplementary Table S5).

Representatives of the Proteobacteria and Chloroflexi synthesize F$_{420}$

To validate these genome-based predictions, we investigated whether isolates of the phyla Proteobacteria and Chloroflexi synthesized F$_{420}$. We cultured three bacteria to test for F$_{420}$ production
under conditions that would promote aerobic, heterotrophic growth: *O. carboxidovorans*, an obligate aerobe and facultative lithothroph of the alphaproteobacterial order Rhizobiales (Meyer and Schlegel, 1978); *P. denitrificans*, a facultative anaerobe and facultative lithothroph of the alphaproteobacterial order Rhodobacterales (de Vries et al., 1989); and *T. roseum*, a thermophilic obligate aerobe and obligate heterotroph of the phylum Chloroflexi (Jackson et al., 1973). *F420* was detected in whole-cell lysates of all three bacteria as well as within the positive control *M. smegmatis* (Figure 3). In HPLC analysis, *F420* species from all four organisms eluted with retention times (Figure 3a) identical to those from a purified *F420* standard validated by mass spectrometry (Figure 3b). When excited at 420 nm, all samples emitted fluorescence ($\lambda_{\text{max}} = 480$ nm) (Figure 3c) with a spectrum characteristic of *F420* (Eirich et al., 1978).

The quantity of *F420* produced and distribution of oligoglutamate chain lengths differed between the strains analyzed (Table 1). Whereas *T. roseum* and *M. smegmatis* produced large quantities of *F420*, normalized production was $\sim 100$-fold lower for the proteobacterial strains under comparable conditions (Figure 3a); *F420* was nevertheless unambiguously and reproducibly detected in such organisms. Ion-pair HPLC was used to resolve the *F420* species based on the length of their oligoglutamate side chains (Figure 3a). Mass spectrometry of the purified standard confirmed the lengths of the dominant HPLC peaks (Figure 3b). The number of glutamate residues attached differed between species, with 4 to 6 glutamates detected in *P. denitrificans*, 5 to 7 detected in *O. carboxidovorans* and 6 to 8 dominating in the *T. roseum* and *M. smegmatis* samples (Table 1).

Having established that species from the phyla Proteobacteria and Chloroflexi synthesize *F420*,

---

**Figure 3** Chemical detection of *F420* in dominant soil phyla. (a) HPLC traces showing *F420* from cell lysates of different species relative to a purified *F420* standard. The traces show the intensity of the fluorescence emitted ($\lambda_{\text{excitation}} = 420$ nm, $\lambda_{\text{emission}} = 480$ nm). The dotted lines show the times at which each species started to elute. (b) Mass spectra confirming the molecular weight and oligoglutamate side chain length of the purified *F420* standard. (c) Fluorescence emission spectra of *F420* from cell lysates of five different species against a purified *F420* standard.
we investigated whether any F_{420}-dependent enzymes that have been characterized in Archaea and Actinobacteria are conserved in the genomes of *P. denitrificans*, *O. carboxidovorans* and *T. roseum*. We identified close homologs of Fgd (F_{420}-reducing glucose 6-phosphate dehydrogenase) in the Chloroflexi and Fno (F_{420}-reducing NADPH dehydrogenase) in the two Proteobacteria. Putative F_{420}H_2-dependent reductases from two superfamily cycles, the luciferase-like hydride transferases (Greening *et al.*, 2016a) and flavin/deazaflavin oxidoreductases (Ahmed *et al.*, 2015), were also encoded in the three organisms (Supplementary Table S6). Multiple sequence alignments and protein homology models confirmed the residues involved in F_{420} binding are almost entirely conserved in these sequences, including residues that interact with the isoalloxazine ring and oligoglutamate chain, that are not found in the flavins FAD and FMN (Supplementary Figure S1). The presence of these sequences and the high conservation of the F_{420}-binding sequence motifs provide support that these organisms encode F_{420}-dependent oxidoreductases.

**F_{420} biosynthesis pathways evolved through bacterial-to-archaeal horizontal gene transfers**

Finally, we analyzed the evolution of the F_{420} biosynthesis enzymes (CofC, CofD, CofE, CofG and CofH) to understand the origin and distribution of the cofactor. With respect to all five proteins, phylogenetic tree topologies (Figure 4a) and sequence similarity networks (Supplementary Figure S2) were similar. In all cases, two large clades corresponding to the euryarchaeotal and actinobacterial enzymes flanked the trees and networks, and there were smaller middle clades corresponding to homologous enzymes in Proteobacteria, Chloroflexi, Tectomicrobia, Firmicutes, Thaumarchaeota and deep-branching actinobacterial lineages such as Acidimicrobia. A MirrorTree analysis comparing the topologies of the phylogenetic trees indicated that the five Cof enzymes co-evolved (with Pearson’s correlation coefficients between 0.71 and 0.93) (Supplementary Figure S3). This is consistent with selection for all five subunits being required to produce functional F_{420}. The F, biosynthesis proteins CofG and CofH in Cyanobacteria have diverged further from their homologs in F_{420}-synthesizing organisms (Supplementary Figures S2 and S4), consistent with their lack of co-evolution with the rest of the pathway.

To better understand the evolutionary origin of these proteins, phylogenetic trees were rooted with related families containing the same protein fold (Figure 4a), namely biotin synthase (a radical S-adenosylmethionine enzyme) for CofG and CofH, adenosylcobinamide guanosine transferase (MobA) for CofC and YvcK (hypothetical protein family) for CofD. CofE adopts a novel protein fold and does not share any significant sequence homology with any other known proteins (Nocek *et al.*, 2007) and thus is presented as an unrooted tree with no evolutionary origin inferred. Our analysis indicates that CofG is likely to have originated in an ancestral archaeon. Subsequently, the functionally and structurally related CofH evolved from an ancient duplication of the cofG gene within an early methanogenic euryarchaeon. In contrast, the roots of the CofC and CofD trees indicate these enzymes originated in an early actinobacterium. These findings are consistent with the recent work suggesting cofC and cofD were among the metabolic genes acquired from bacteria during the diversification of archaea (Nelson-Sathi *et al.*, 2015). Our observations suggest that, in contrast to the F, syntheses, enzymes specific for F_{420} biosynthesis evolved in bacteria, most probably an early actinobacterium, and were subsequently horizontally transferred into archaea.

The genetic organization of the cof genes is highly variable (Figure 4b). The five opening reading frames are genomically separated in two methanogenic orders (Methanococcales and Methanomicrobiales) and are partially clustered in Actinobacteria, Chloroflexi, Tectomicrobia and all other archaea. In contrast, cof genes are organized as complete operons within Proteobacteria; consistent with the presence of two distinct clusters of proteobacterial sequences in all phylogenetic trees, these operons were organized in two distinct configurations (designated Proteobacteria I and Proteobacteria II). Our analysis suggests that the cof genes initially evolved separately, but became increasingly syntenic because of selective pressures to inherit all five enzymes for F_{420} production. This analysis shows that the deepest-rooted branches of the CofGH fusion protein are between the classes Actinobacteria and Acidimicrobia. This suggests the CofGH fusion occurred early in actinobacterial evolution, presumably to enhance the catalytic efficiency of the
coordinated radical S-adenosylmethionine reactions required for F₄₂₀ synthesis (Philmus et al., 2015). In some organisms, F₄₂₀-reducing dehydrogenases (Fgd, Fno and Mer) and putative F₄₂₀H₂-dependent reductases also appear to be operonic with the cof genes (Figure 4b).

**Discussion**

Until now, the vast majority of research into F₄₂₀ has focused on the roles of the cofactor in methane cycling and tuberculosis infection (Greening et al., 2016a). However, our findings demonstrate that the cofactor is widely distributed among diverse taxa and ecosystems, and is likely to play a role in a broader array of metabolic and ecological phenomena than is currently recognized. We have demonstrated, using a combination of comparative genomics and analytical chemistry approaches, that F₄₂₀ is widely distributed in three of the five most dominant soil phyla (Actinobacteria, Proteobacteria and Chloroflexi) and provided genomic evidence that the cofactor is encoded in other phyla.
Evolution and distribution of the cofactor F420
B Ney et al

(Tectomicrobia, Firmicutes, Thaumarchaeota and Geoarchaeota). The finding that F420 biosynthesis genes are widely distributed within aerobic bacterial and archaeal taxa indicates that F420 is far from a niche methanogenic cofactor. Furthermore, the abundance of these genes in soil ecosystems suggests that F420 influences the biological and chemical composition of soils.

There are three feasible explanations for the observed distribution of F420 in biological systems: (1) F420 biosynthesis evolved pre-LUCA (last universal common ancestor); (2) F420 biosynthesis evolved in an archaeon and was acquired by bacteria; or (3) F420 biosynthesis evolved in a bacterium and was acquired by archaea. Of these, the first hypothesis is unsupported given that cof gene evolution does not parallel 16S rRNA gene phylogenies and the cofactor appears to be completely absent from proposed deep-branching bacteria (for example, Aquificae and Thermotogae) and archaea (for example, Crenarchaeota and Thermococci). The widely assumed archaea-to-bacteria hypothesis remains plausible, but is challenged by phylogenetic trees suggesting a bacterial origin for CofC and CofD. The bacteria-to-archaea hypothesis for F420 acquisition is more robustly supported by our phylogenetic trees and is also consistent with the findings of the large-scale Martin analysis on interdomain lateral gene transfers (Nelson-Sathi et al., 2015).

In Figure 4c, we present an evolutionary model of the origin and distribution of F420. Our studies strongly support that the genes required for F420 synthesis (cofG and cofH) evolved independently and probably earlier than the genes required for F420 synthesis from F0 (cofC, cofD and cofE). We hypothesize that an early methanogen synthesized F0 but not F420 using the products of ancient cofG and cofH genes. This cofactor may have initially sustained many of the redox functions now supported by F420, including hydrogenotrophic methanogenesis. Our evolutionary analysis suggests that early actinobacterium subsequently acquired cofG and cofH (later fused into cofGH) and evolved the cofC, cofD and likely cofE genes required to produce F420; the lactyloligoglutamyl derivative of F0. Multiple horizontal gene transfer events thereafter led to the acquisition of these genes in the archaea and other bacteria. The euryarchaeotal branches of the CofC, CofD, and CofE phylogenetic trees mirror those of 16S rRNA gene trees (Hedderich and Whitman, 2013). This is consistent with their genes being acquired from Actinobacteria early in the evolution of Euryarchaeota (Figure 4), resulting in the capacity for F420 biosynthesis being vertically inherited by the six methanogenic orders, Archaeoglobi, Halobacteria and likely anaerobic methanotrophs (ANME). unicellular eukaryotes (Chlorophyta and Streptophyta) appear to have acquired the chromophore F0 by laterally acquiring the fused cofGH gene from Proteobacteria. The origin of the cyanobacterial cofG and cofH remains unclear because of their evolutionary distance from the rest of the sequences; scenarios involving acquisition from methanogens or Proteobacteria are both plausible.

We propose that the evolutionary driving force resulting in F0 synthesis was the need for redox cofactors with distinct electrochemical properties to flavins and nicotinamides to drive the unique reactions of hydrogenotrophic methanogenesis. By hijacking the riboflavins synthesis pathway with CofG and CofH, an early euryarchaeon would have produced a cofactor with the ideal electrochemistry to mediate the central reactions of methanogenesis: H2 oxidation, methenyl and methylene reduction and NADP reduction. F0 could have functioned as the primordial methanogenic redox cofactor as it has near-identical electrochemical properties to F420, including a standard redox potential of −340 mV and obligate two-electron reactivity (Walsh, 1986; Greening et al., 2016a). Furthermore, it has been demonstrated that F0 can functionally substitute for F420 in various enzymatic processes in vitro, including in the crucial methanogenesis enzymes Frh (F420-reducing hydrogenase) and Fno (F420-dependent NADP reductase) (Yamazaki and Tsai, 1980; Muth et al., 1987). The current distribution of cof genes emphasizes that deazaflavins have primarily been selected for their electrochemical and not photochemical properties. With the exception of cyanobacterial and eukaryotic phototrophs, all organisms appear to have retained cofG and cofH genes primarily to generate F420 as a redox cofactor rather than F0 as an independent chromophore. We propose that the need to produce a charged derivative of F0, F420, served as the main selection pressure for the evolution and dispersal of CofC, CofD and CofE. The feature that distinguishes F420 from its precursor F0 is the presence of a negatively charged sidechain containing phosphate and glutamate groups (Eirich et al., 1978). Thus, whereas F0 diffuses through membranes, F420 is retained in the cytosol (Glas et al., 2009). The electrostatic properties of F420 may also facilitate higher affinity enzyme–cofactor interactions according to recent structural studies (Greening et al., 2016a). We also show here that the length of the oligoglutamate sidechain of F420 varies between phyla and progressively increases between the Euryarchaeota (Gorris and van der Drift, 1994), Proteobacteria, Actinobacteria and Chloroflexi (Table 1). We are presently investigating the reasons and mechanisms behind these variations.

F420 may confer a number of competitive advantages to these bacteria by mediating both endogenous and exogenous metabolic processes. F420 is known to serve as a redox cofactor in an increasing number of endogenous processes in mycobacteria, including central carbon metabolism (Bashiri et al., 2008), cell wall modification (Purwantini and Mukhopadhyay, 2013), antioxidant production (Ahmed et al., 2015) and possibly quinone reduction (Gurumurthy et al., 2013). Although the cofactor is synthesized in high
levels under oxic conditions (Figure 2), phenotypic evidence suggests it is particularly important for survival under hypoxia (Gurumurthy et al., 2013). In this condition, mycobacteria may increasingly depend on low-potential cofactors such as F420 to maintain redox homeostasis (Berney et al., 2014; Cook et al., 2014). Many of the F420H2-dependent oxidoreductases involved in mycobacterial metabolism are also conserved in other taxa, where they may have similar roles (Supplementary Table S5). F420 is likely to be particularly important in the redox metabolism of Chloroflexi and other Actinobacteria given representatives of such phyla produce large amounts of the compound even under optimal growth conditions (Figure 3a). A central role for F420 in the metabolism of Proteobacteria seems less likely, given the levels detected of the cofactor in O. carboxidovorans (Proteobacteria I lineage) and P. denitrificans (Proteobacteria II lineage) were low (Figure 3a). It is likely that transcription of the cof genes is constitutively repressed in such organisms, but is activated in response to environmental stresses or other signals. Consistently, TetR and MarR family transcriptional repressors were identified immediately downstream of the cof operons in most Proteobacteria, including P. denitrificans and O. carboxidivorans.

The richest literature on the physiological roles of F420 in bacteria concerns the metabolism of complex organic compounds. F420H2-dependent reductases mediate the biodegradation of nitroaromatic explosives (Ebert et al., 2001), triphenyl dyes (Guerra-Lopez et al., 2007) and furanocoumarins (Taylor et al., 2010), as well as the biosynthesis of tetracycline and pyrrolobenzodiazepine antibiotics (Li et al., 2009; Wang et al., 2013). A role for F420 in equivalent processes seems particularly likely for the Chloroflexi and Tectomicrobia; both phyla contain an abundance of F420H2-dependent reductases (Supplementary Tables S3 and S6) and are reputed equivalent processes seems particularly likely for the Chloroflexi and Tectomicrobia; both phyla contain an abundance of F420H2-dependent reductases (Supplementary Tables S3 and S6) and are reputed for their biosynthetic versatility and biodegradative capacities (Björnsson et al., 2002; Wilson et al., 2014). In turn, these F420-dependent processes are likely to affect the biological composition of soil ecosystems, for example, by shaping the antibiotics arms race. They may also influence the chemical composition of soil environments by controlling the levels of complex organic compounds produced and consumed. Better understanding of the endogenous and exogenous roles of F420 depends on defining the functions of the numerous putative F420-dependent oxidoreductases encoded in aerobic bacteria and archaea (Supplementary Table S3). In turn, further understanding the physiological and ecological significance of F420 may open opportunities in the fields of pharmaceuticals, biocatalysis and bioremediation.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr Ghader Bashiri (University of Auckland) for providing the fbiABC overexpression plasmid, Professor William Jacobs Jr (Albert Einstein College of Medicine) for supplying Mycobacterium smegmatis mc² 4517 and the three anonymous reviewers for their helpful suggestions. This work was supported by Australian Research Council research Grants (DE120102673, DP130102144) awarded to CJJ; Australian National University Summer and PhD scholarships awarded respectively to BN and FHA; and a CSIRO Office of the Chief Executive Postdoctoral Fellowship awarded to CG.

Author contributions

CG, CJJ, BN, FHA, CRC, MBS, MCT, ACW, JGO, SJW and GP designed experiments. BN, FHA, CG, CRC, AB, SJW, and CJJ performed experiments. CG, CJJ, BN, FHA, CRC, MBS, MCT, ACW, JGO, SJW and EA supervised students and researchers. CG, CJJ, FHA, BN and JGO wrote the paper.

References


Bashiri G, Squire CJ, Moreland NJ, Baker EN. (2008). Crystal structures of F420-dependent glucose-6-

The ISME Journal


Evolution and distribution of the cofactor F$_{420}$
B Ney et al.


Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)