**Title**
Cobamide and cobamide precursor cross-feeding: a genomic perspective



**Author**
Shelton, Amanda Nicole

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

# Cobamide and cobamide precursor cross-feeding: a genomic perspective

By

Amanda N. Shelton

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley.

Committee in charge:

Professor Michiko E. Taga, Chair
Professor Steven E. Lindow
Professor Matthew Traxler
Professor Ellen L. Simms

Summer 2019

Abstract

Cobamide and cobamide precursor cross-feeding: a genomic perspective

by

Amanda N. Shelton

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Michiko E. Taga, Chair

Many bacteria cannot synthesize all the nutrients they require. One type of nutrient that many bacteria must acquire from their environment is cobamides, a class of cofactors including vitamin $B_{12}$ and related analogs. Less than half of the bacteria that use cobamides produce them, suggesting that cobamides are widely shared in microbial communities. Cobamides are used in a variety of bacterial enzymes for functions including methionine synthesis, deoxyribonucleotide synthesis, carbon utilization, reductive dehalogenation, and others. Cobamides consist of a corrin ring macrocycle with a cobalt center. Attached to the cobalt is the upper ligand, which is part of the reactive center of the cofactor. The lower ligand, which may also be attached to the cobalt, is structurally variable and is generally from one of three classes: benzimidazoles, purines, or phenolics. Most bacteria that use cobamides can only use a subset of all cobamides, referred to as cobamide selectivity.

In Chapter 1, I provide background on cobamide structure, function, and biosynthesis. I then review known processes that contribute to an organism's observed cobamide selectivity. These mechanisms include strategies for making a cobamide: *de novo* biosynthesis, cobamide precursor salvaging, and cobamide remodeling. Uptake, regulation, and cobamide-dependent enzyme binding and activity also contribute to cobamide selectivity. Examples of cobamide sharing in a number of systems are also described.

To explore cobamide biosynthesis and dependence patterns in bacteria, I have taken a genomics-centered approach in Chapter 2. I assessed the potential of publically available genomes for use of cobamides and for cobamide biosynthesis based on the presence and absence of genes that have been experimentally validated, creating predictions for the phenotype of an organism. I analyzed 11,000 representative species from a dataset of all bacteria, and observed trends of cobamide biosynthesis and cobamide dependence potential at the phylum level.

To further validate the genomic prediction of cobamide phenotypes, the cobamide requirements of the human pathogen *Clostridioides (Clostridium) difficile* were interrogated in Chapter 3. This bacterium has seven cobamide dependent pathways encoded in its genome, and was

validated to be an ALA salvaging bacterium that produces pseudocobalamin. How cobamides affect the ability of *C. difficile* to grow and cause disease *in vivo* remains to be discovered, but the *in vitro* experiments revealed different cobamide preferences of methionine synthase and its cobamide-dependent ribonucleotide reductase. Its methionine synthase could surprisingly support growth with all nine cobamides and two cobamide precursors tested. *C. difficile* also exemplifies the strategy of metabolic flexibility, whereby it can use many cobamides and precursors to use cobamide-dependent metabolism, and also has options to bypass these pathways and not use cobamides.

The observation of the patchy distribution of the cobamide biosynthesis pathway across bacterial genomes led to the hypothesis that cobamide biosynthesis capability has been gained and lost multiple times in the domain Bacteria. In Chapter 4, I analyzed the cobamide biosynthesis genes in the order Clostridiales in the phylum Firmicutes by annotating the presence and absence of these genes onto phylogenetic trees. The Clostridiales in particular have many instances of putative gain or loss of some but not all genes in the biosynthesis pathway, possibly leading to the observed ALA-salvaging bacteria observed in the clade.

My work used comparative genomics to generate testable hypotheses about cobamide biosynthesis, cobamide precursor salvaging, and cobamide dependence in bacteria. Characterization of cobamide biosynthesis and dependence in *C. difficile* validated some of the genomic predictions and filled a knowledge gap in understanding the metabolism of this opportunistic pathogen. My final chapter summarizes some of the major findings of my PhD thesis, and describes a few aspects of cobamide sharing that remain to be investigated.

# Acknowledgements

My PhD advisor, Michiko Taga, has been so encouraging of my projects that are on the edge of our lab's expertise, especially for the bioinformatics. She has helped me develop so much as a scientist, in experimental design, in writing, and in presenting. The whole Taga lab has been welcoming from the beginning, and a source of support, making our lab a great place to learn. Postdocs Amrita Hazra, Florian Widner, Nicole Abreu, Sebastian Gude, and Zach Hallberg have been great sounding boards for both science and professional development. Kenny Mok was a great rotation mentor and has continued to teach me new techniques throughout my time in the lab. I've learned a lot from my fellow grad students Olga Sokolovskaya, Kris Kennedy, Gordon Pherribo, and Alexa Nicolas, especially since we each focus in a different area. They have also been a great source of support. Thanks especially to Amrita, Olga, and Kenny for all the help with the HPLC. Jong Duk Park, Anna Grimaldo, Joseph Maa, Clare Lou, Soohan Woo, Kim Kang, Victoria Innocent and other undergrads have kept the lab a lively place. I have to particularly thank undergrad Xun (Allen) Lyu for his work on the *C. difficile* genetics. Thank you to Luis Valentin-Alvarado for our conversations about *C. difficile* metabolism and molecular techniques during his rotation, we learned a lot together.

A number of people have taught me particular techniques or given me significant advice throughout my PhD. For cobamide biology: Martin Warren, Alison Smith, and Johan Kudahl. For *C. difficile*: Aimee Shen, Craig Ellermeier, and Drew Hryckowian. For phylogenetics: John Taylor, Gary Andersen, Christian Sieber, and Eric Dubinsky. I also have to thank my collaborators at the JCVI on the *bza* HMMs, who took on this project in response to a cold email about TIGRfams: David Haft and Samantha Jackson.

Thank you to my rotation labs: to Jill Banfield and Karthik Anantharaman for inspiring a love of genomics and looking for microbial potential, and to Lisa Alvarez-Cohen and Sara Gushgari for extensive training in anaerobic culturing. I also want to thank the professors who have served on my qualifying exam and thesis committees: Steve Lindow, Matt Traxler, Ellen Simms, Michael Shapira, and Michael Freeling.

Thank you to my undergrad research mentors, Richard Murray, Adrian Ponce, and Jared Leadbetter, as well as my microbiology professors Victoria Orphan, Dianne Newman, and Rob Philips, for cultivating my interest in biology and microbiology and preparing me well for starting graduate school at UC Berkeley.

PMB has been a great home these past few years and I am glad to have participated in PMBG, MSG, and BASIS. Thank you to the 2014 cohort for their support through the courses, quals, and now graduating, and especially to Brittney Nguyen and Cindy Amstutz for all the good times at Carleton St.

Thank you to Mom and Dad for supporting my curiosity and drive that led to me becoming a scientist. They have always been there to listen and encourage me when things get tough. Thank you to Emily Shelton for being a great friend and listener from childhood to today. I also have to thank my grandparents, aunts, and uncles for their support throughout the years. Thank you to my partner, Ben Yuan, for answering my questions about coding and computers at all hours. He has been so supportive and a great listener, despite the current distance, and has journeyed with me through Caltech and through our PhDs.

# Table of Contents

# Chapter 1

## Cobamides are shared metabolites, but there is added complexity

### Auxotrophy and nutrient exchange in microbial communities

Microbial requirements for exogenous nutrients are widespread. In the laboratory, we provide carbon and nitrogen sources, trace elements, and vitamins in the medium. In their natural habitat, microbes acquire these nutrients from both abiotic and biotic sources. Microbes share metabolites in both one-sided and mutually beneficial interactions with other organisms in their communities[1,2]. Estimates of bacterial auxotrophy, in which an organism requires a metabolite that it cannot produce, by genomic analysis have focused on amino acids. One study that examined auxotrophy for amino acids, some vitamins, and nucleobases found that up to 76% of genomes in their data set did not have complete biosynthetic pathways for all these metabolites, based on a threshold defining auxotrophy as lacking more than 50% of the genes in a pathway[3]. Automated genomic analysis found that up to 90% of about 6000 genomes did not have all 20 complete amino acid synthesis pathways[4]. While genomic analysis needs additional development to improve the accuracy of prediction of auxotrophy in diverse lineages[5], and is dependent on what genomes are present in the data set, genomics and culturing both show that many microbes are auxotrophic for one or more nutrients. Amino acid auxotrophs readily arise in laboratory-evolved *Escherichia coli* populations, and these auxotrophs acquire the needed amino acid from producer populations in the culture, showing that cross-feeding interactions, in which one organism acquires a nutrient from another, can easily occur[6]. Here I will focus on the shared class of metabolites called cobamides, enzyme cofactors that catalyze diverse reactions, and use genomics and experimental studies to explore cobamide and cobamide precursor auxotrophy. Only some microbes encode the complete *de novo* biosynthetic pathway, suggesting there is widespread cross-feeding of these molecules[7–11].

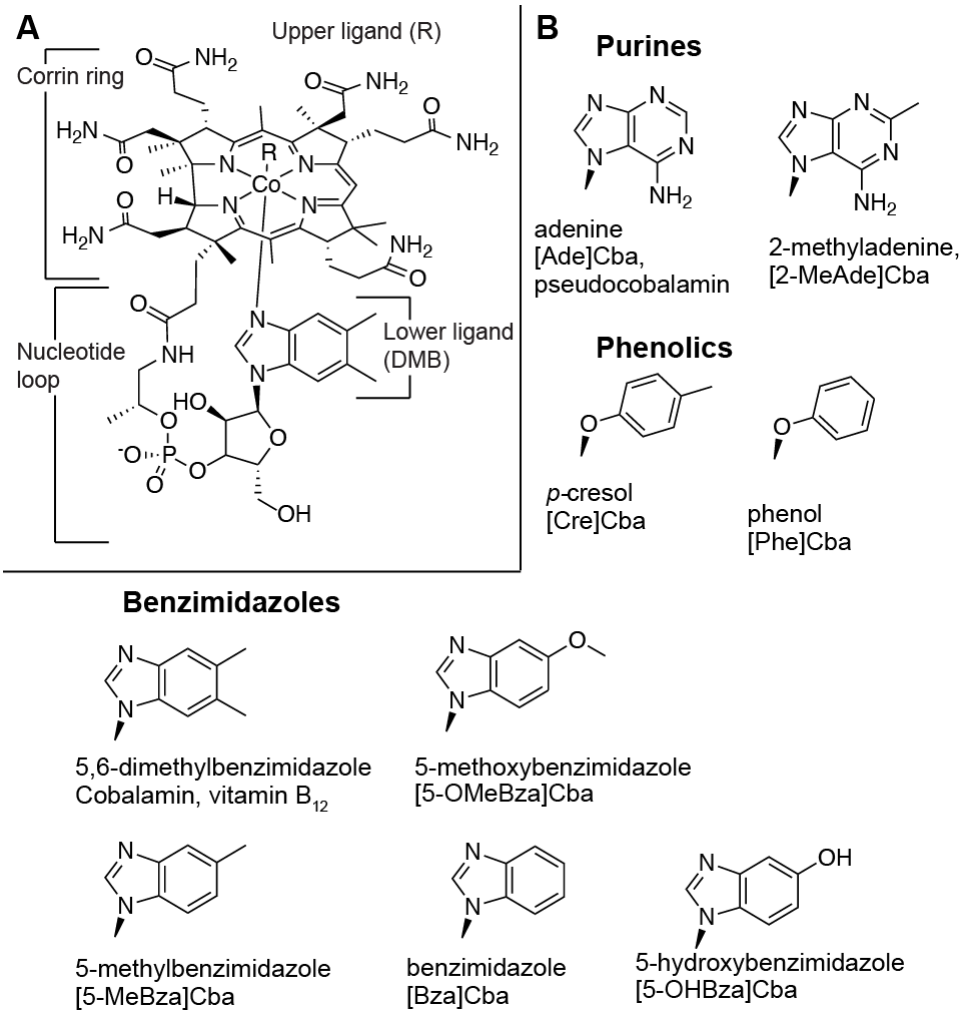



**Figure 1. Structure of cobamides. A.** Structure of cobalamin ($B_{12}$) with the corrin ring, nucleotide loop, and lower ligand 5,6-dimethylbenzimidazole (DMB) labeled. R, the upper ligand, can be CN, 5'-deoxyadeonsyl, methyl, or aquo. **B.** Alternative lower ligand structures. The name of the lower ligand is on the top row, and the abbreviation for the corresponding cobamide is on the bottom row.

## Cobamide structure and function

Cobamides are a family of enzyme cofactors that includes vitamin $B_{12}$ (Fig. 1). They consist of a modified tetrapyrrole that coordinates a cobalt center, known as a corrin ring. Any compound that has a corrin ring, including cobamides, is called a corrinoid. An upper ligand that coordinates to the cobalt ion is where the enzyme chemistry occurs. The upper ligand can be a cyano group in the inactive vitamin form; an aquo ($OH_2$) group; a methyl group used in methyltransferases reactions[12]; or a 5'-deoxyadenosyl group used for radical chemistry[13]. In some enzymes, like reductive dehalogenase and epoxyqueuosine reductase, no upper ligand is found in the active form, and instead the reactions are directly catalyzed by reactions with the cobalt center[14,15]. The lower ligand is tethered to the corrin ring by the nucleotide loop, and can affect the catalytic properties of the cofactor[16]. The lower ligand structure can vary, and there are three defined classes produced by bacteria and archaea: benzimidazoles, purines, and

phenolics[17] (Fig. 1B). Unlike benzimidazoles and purines, phenolic lower ligands cannot form a coordination bond to the central cobalt. The structure of the lower ligand affects enzyme binding and catalysis[18–22]. A complete corrinoid, also called a cobamide, has both an upper and lower ligand. Cobamides have been found to be produced only by bacteria and archaea[23].

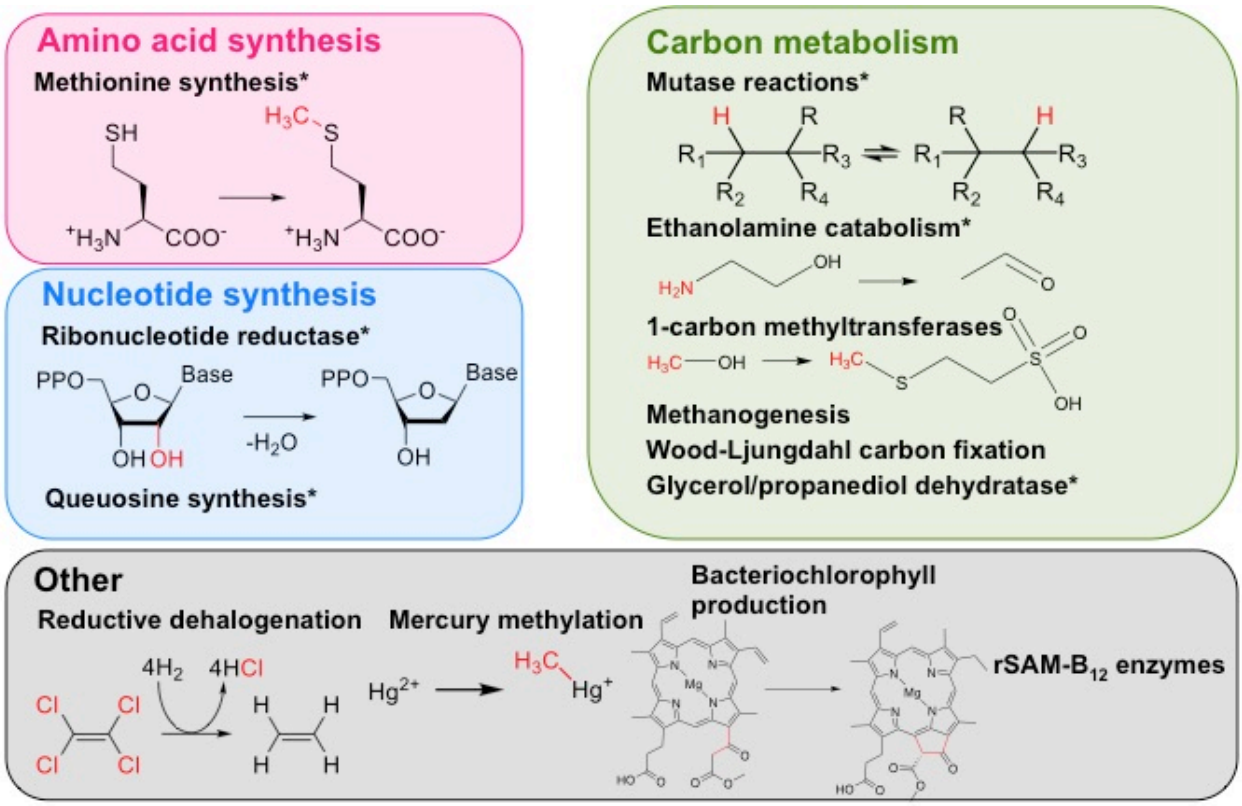



**Figure 2. Processes catalyzed by cobamides.** A selection of cobamide-dependent reactions carried out by cobamide-dependent processes. In the reactions shown, red atoms depict the change in the molecule caused by the cobamide-dependent reaction. Reactions are grouped by their biological function. Reactions with an asterisk (*) have known cobamide-independent alternatives. Abbreviations: rSAM, radical S-adenosylmethionine.

## Cobamides catalyze diverse reactions

Members of all three domains of life use cobamide-dependent enzymes. In the domain Eukaryota, clades that require cobamides use them for methionine synthase (MetH), for methylmalonyl-CoA mutase (MCM), which is involved in fatty acid metabolism by interconversion of succinate and propionate, and for deoxyribonucleotide synthesis in a few cases[13,24,25] (Fig. 2). Bacteria and archaea have many additional cobamide-dependent enzymes that are used in carbon metabolism, nucleotide synthesis, and other reactions (Fig. 2). A number of cobamide-dependent mutase enzymes are used by anaerobic bacteria to use glycerol, propanediol, ethanolamine, beta-lysine, D-ornithine, and other compounds as carbon and/or nitrogen sources[13]. Cobamides are involved in both acetogenesis and methanogenesis, which affect carbon mineralization in ecosystems[13]. Cobamide-dependent metabolisms like reductive dehalogenation can bioremediate industrial pollutants[26], or create a more toxic compound as in

mercury methylation[27]. Radical-SAM (rSAM) enzymes containing both cobamide and S-adenosylmethionine (SAM) cofactors function in pathways producing antibiotics and other secondary metabolites[28,29]. Cobamides have also been found in light-sensing proteins, in which the light-induced cleavage of adenosylcobalamin is used to change the conformation of the protein[30–32]. Chapter 2 examines the relative abundance of these diverse enzyme families in over 11,000 bacterial genomes.

Seven cobamide-dependent pathways have known cobamide-independent alternatives. This is how some lineages can perform essential functions such as deoxyribonucleotide synthesis without requiring cobamides. There are three classes of ribonucleotide reductases (RNRs), which catalyze the conversion of ribonucleotides to deoxyribonucleotides, of which only class II enzymes require cobamides[33]. As an alternative to MCM, some organisms have the methylcitrate pathway, a cobamide-independent pathway that interconverts propionate, a product of fatty acid or branched chain amino acid degradation, and succinate[34]. For methionine synthesis, organisms may encode the cobamide-independent alternative methionine synthase MetE or take up methionine from the environment[35]. However, in all of these cases, the cobamide-independent alternatives do not necessarily perform the function in exactly the same way. For example, MetH is thought to be much more efficient than MetE, because under certain stress conditions, MetE comprises up to 5% of the total protein in the cell[36–38]. Further, the cobamide-dependent ribonucleotide reductase (RNR), NrdJ, is essential for *Sinorhizobium meliloti* symbiosis with alfalfa. Replacing NrdJ with a heterologous cobamide-independent RNR cannot restore symbiosis[39]. The ability of cobamide-independent alternatives to glycerol and propanediol dehydratases, ethanolamine ammonia lyase, and epoxyqueuosine reductase to substitute for their cobamide-dependent counterparts has not been well-studied[40–45]. Cobamide-independent pathways must be taken in to account when studying the frequency of cobamide-dependence and the physiology of cobamide use in organisms that contain these pathways.

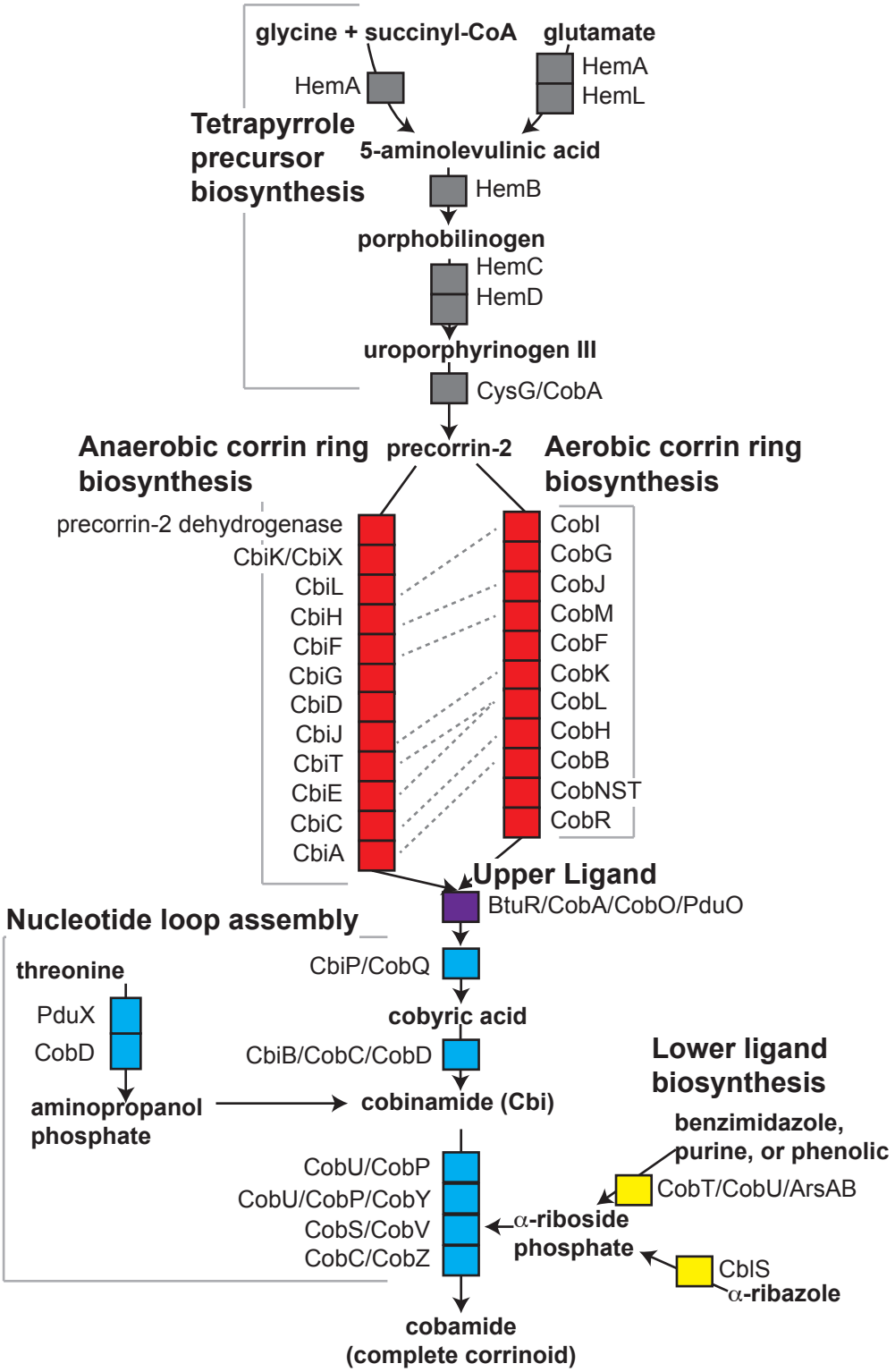



**Figure 3. Cobamide biosynthesis genes**. The main enzymes of *de novo* cobamide biosynthesis are depicted in this schematic as colored squares, and key intermediates are in bold text. Labels for pathway sections correspond to the labels of the cobalamin structure in Fig 1A and have the following colors: tetrapyrrole precursor biosynthesis, gray; corrin ring biosynthesis, red; upper ligand attachment, purple;

nucleotide loop assembly, cyan; lower ligand biosynthesis, yellow. Dashed lines indicate homologous enzymes in the corrin ring section of the pathway. Enzyme names are given for each step in the pathway.

## Cobamide biosynthesis

Previous estimates of cobamide dependence suggested that up to 75% of microbial species use cobamides for at least one process, but far fewer contain the entire *de novo* cobamide biosynthesis pathway[7–9]. The complete pathway is approximately 30 genes (Fig. 3), with the first 5-7 genes being shared with other tetrapyrrole synthesis pathways including heme, chlorophyll, and siroheme (Fig. 4)[23]. The cobamide biosynthesis pathway was genetically and biochemically characterized in *Pseudomonas denitrificans,* which has the aerobic branch of corrin ring synthesis, and in *Propionibacterium shermanii, Bacillus megaterium,* and *Salmonella enterica* serovar Typhimurium, which have the anaerobic branch (Fig. 3, red shapes)[46,47]. Overall, the biosynthesis pathway is similar in all lineages with the exception of the timing of the cobalt insertion and ring contraction to form the corrin ring, as discussed below, but there are nonhomologous replacements of genes in some lineages in the canonical pathways of these model organisms, especially in the domain Archaea[48–51].

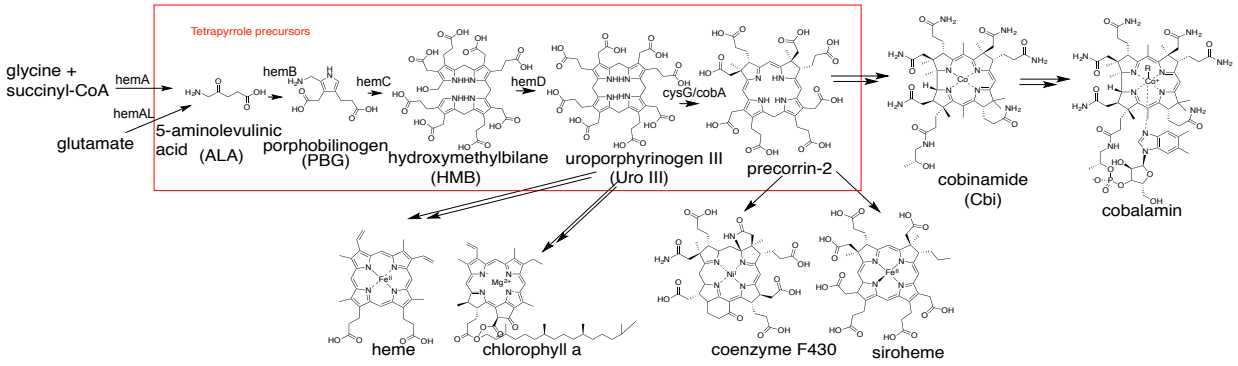



**Figure 4**. Tetrapyrrole synthesis. Tetrapyrrole precursors and branch points to heme, chlorophyll, coenzyme F430, siroheme, and cobalamin are shown.

The early steps of cobamide biosynthesis are shared with all tetrapyrroles (Fig. 4). In Alphaproteobacteria, glycine and succinyl-CoA are combined by HemA to form the first committed tetrapyrrole intermediate, 5-aminolevulinic acid (ALA)[52]. In all other bacteria, glutamate is converted by HemA and HemL to ALA[53,54]. Two molecules of ALA are combined by HemB to form the pyrrole porphobilinogen (PBG), and then four molecules of PBG are combined by HemC and HemD to form uroporphyrinogen III (UroIII), a tetrapyrrole macrocycle, where traditional heme and chlorophyll synthesis through the intermediate protoporphyrin IX split from the corrinoid pathway [55]. UroIII is then converted to precorrin-2 via CysG/CobA (Fig. 3, gray, Fig. 4).

At this point, cobamide biosynthesis splits from all other tetrapyrroles (Fig. 4) to form the corrin ring, which is modified with a ring contraction, gains additional methyl and amide substituents, and a cobalt ion center. This section of the pathway requires 10-12 genes and has two distinct branches, the oxygen-requiring late cobalt insertion pathway (aerobic), and the

oxygen-sensitive early cobalt insertion pathway (anaerobic) (shown in red in Fig. 3)[23,46]. Genomes generally encode only one of these alternative pathways for synthesizing the corrin ring. Some genes in each pathway are orthologous to each other (Fig 3, dashed lines). The shared end product of the corrin ring pathway, cobyrinic acid, is where nucleotide loop assembly begins (Fig. 3, cyan). Here is where the corrinoid is now adenosylated, adding a 5'-deoxyadenosyl group as the upper ligand (Fig. 3, purple). The cobyrinic acid intermediate is further modified and an aminopropanol linker that connects the corrin ring to the lower ligand is attached, forming the intermediate cobinamide (Cbi). Production of this linker has been shown to be catalyzed by PduX and CobD in *S. enterica*[56–59], and the kinase PduX is replaced by BluE in Rhodobacterales[60]. CobD is replaced by a specific L-serine phosphate decarboxylase for the formation of an ethanolamine linker in organisms that produce norcobamides[61]. The kinase is just one example where certain clades use different enzymes to carry out the same step in cobamide biosynthesis. Lineages missing known genes for specific steps in the cobamide biosynthesis pathway are candidates for discovery of new cobamide biosynthesis genes[48–51]. Separate from the aminopropanol linker, the lower ligand is activated to an α-riboside phosphate by CobT[62] (Fig. 3, yellow). Finally, the lower ligand is connected to the aminopropanol linker via CobS and the phosphate removed by CobC[63].

**Genetic determinants of cobamide lower ligand structure**

Variation in the lower ligand structure is known to be encoded by certain synthesis and attachment genes for each class of lower ligands (Fig. 1B)[64–69]. There are two independent pathways to make 5,6-dimethylbenzimidazole (DMB), the lower ligand for vitamin $B_{12}$. The oxygen-dependent pathway uses BluB to convert flavin mononucleotide to DMB[66,67,70,71]. The anaerobic pathway uses BzaABCDE or BzaFCDE to convert 5-aminoimidazole ribotide to DMB[68,72]. Organisms with only some of these *bza* genes make other benzimidazoles[68].

For the class of phenolyl cobamides, the CobT homologs ArsA and ArsB heterodimer activate phenol and *p*-cresol to form α-riboside phosphates that can be attached to the nucleotide loop[64,65]. So far, the production of phenolyl cobamides appears to be restricted to the class Negativicutes in the phylum Firmicutes[64] (Chapter 2). Based on isotope labeling in *Sporomusa ovata,* the source of phenolic compounds for the lower ligand is likely the degradation of aromatic amino acids[73].

The genomic determinants of the class of purinyl cobamides are not well known, but the absence of known benzimidazole and phenolic lower ligand determinants is likely a strong signal. Recent work has suggested that *cobT* sequences segregate into subfamilies based on the lower ligand attached, although many homologs tested *in vitro* or in heterologous hosts are able to activate diverse lower ligands[49,69,74,75]. A gene tree of some *cobT* homologs shows the *cobT* genes of organisms that make purinyl cobamides are a separate group from those that make benzimidazolyl cobamides[69]. However, recent analysis of distant *cobT* homologs from Cyanobacteria and Archaea shows additional groupings[49]. Additional study of CobT and other potential lower ligand determinants is needed to find a genetic basis for purinyl cobamide biosynthesis preference.

**Processes that contribute to cobamide selectivity**

Although the reactive center of all cobamide cofactors is the cobalt and upper ligand, all studies reported thus far indicate that a given organism cannot use all cobamides with differing lower ligands for its cobamide-dependent metabolisms. This is referred to as cobamide

preference or cobamide selectivity. For example, eukaryotic algae showed no growth with pseudocobalamin under MetH-dependent conditions, but could use cobalamin[11]. In contrast, a *metE* loss-of-function mutant of *Salmonella enterica,* using MetH, was able to use a wider variety of cobamides to grow: cobalamin, pseudocobalamin, [Phe]Cba, and [Cre]Cba, but growth with other cobamides was not reported[19,64]. Additionally, *Bacteroides thetaiotaomicron* could use benzimidazolyl and purinyl cobamides, but not phenolyl cobamides for MetH-dependent growth[9]. These examples show that organisms using cobamides for the same function have different cobamide selectivity. In Chapter 3, I explore the cobamide preference of *Clostridioides (Clostridium) difficile*. These preferences are often observed by the ability or inability of cobamides to support growth under cobamide-requiring conditions, and a number of processes contribute to the observed specificity *in vivo* (Fig. 5).

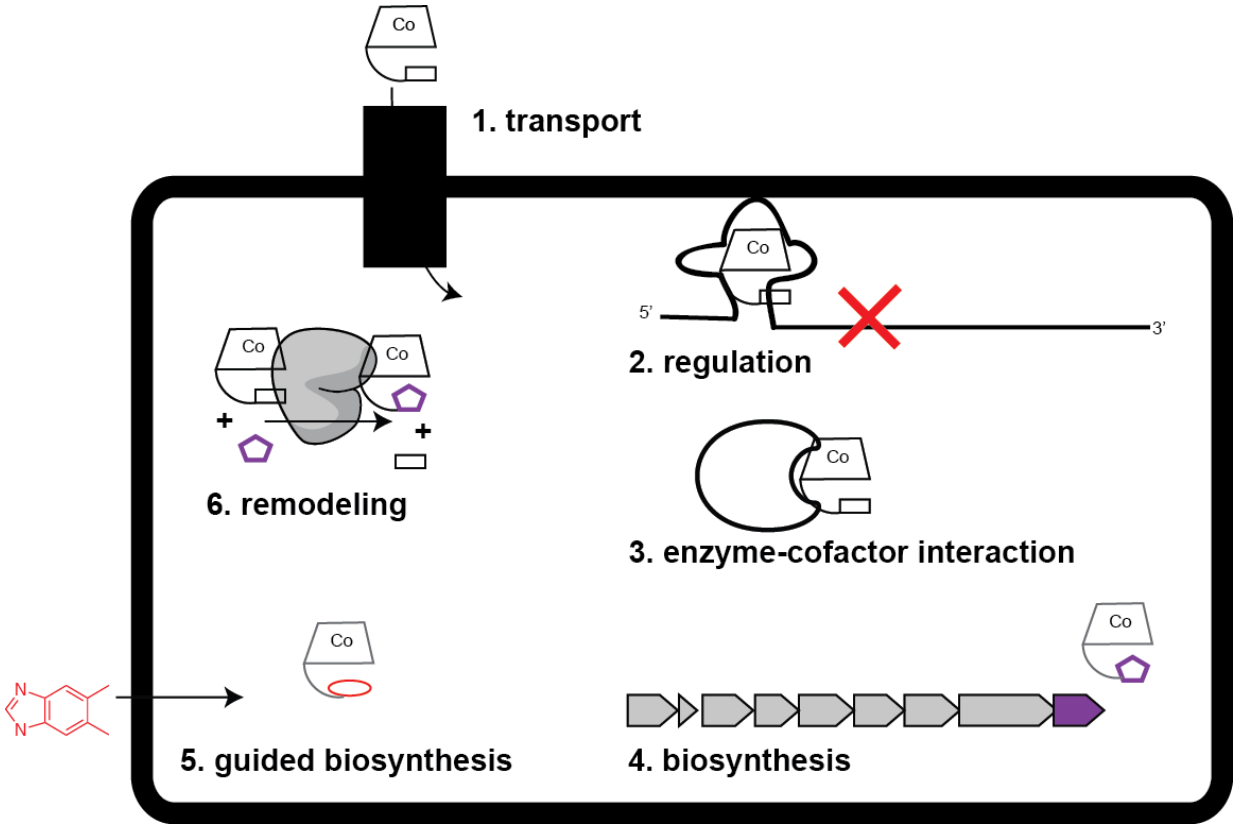



**Figure 5. Processes that affect cobamide selectivity in bacteria.** 1. Uptake of cobamides through Btu(B)FCD or other transporters. 2. Regulation by DNA-binding proteins or cobamide riboswitches that turn off gene expression upon RNA-cobamide binding. 3. Cobamide-dependent enzymes interacting with the cofactor including binding of the cofactor and activity of the holoenzyme. 4. Cobamide biosynthesis and lower ligand determinant genes. 5. Incorporation of exogenous lower ligands (guided biosynthesis). 6. Exchange of lower ligands via cobamide remodeling.

An organism's observed cobamide preference is the result of a combination of multiple cellular processes including transport, regulation, the cobamide-dependent enzymes themselves, biosynthesis of the cofactor, and replacement of the upper ligand by the adenosyltransferase. Whether adenosyltransferases selectively adenosylate cobamides based on

the lower ligand present is unknown, but the other processes that contribute to cobamide selectivity have been studied in some systems. The ABC transporter BtuFCD is the best-studied cobamide import system (Fig. 5, #1), which can import both cobamides and the late intermediate cobinamide (Cbi)[76–79]. In Gram-negative bacteria, transport also requires the outer membrane binding protein BtuB, which has been determined to be a major component of lower ligand-related specificity in those systems[9]. In *Bacteroides thetaiotaomicron*, mutants of three different *btuB* homologs have differing competitive advantages depending on the cobamide provided, suggesting that the different BtuB homologs have distinct cobamide affinities[9]. Additional cobalamin importers have been discovered or predicted more recently[8,80–82]. Detailed studies of cobamide selectivity of transporters other than Btu(B)FCD with a panel of cobamides with differing upper and lower ligands have not yet been conducted.

Some selectivity has been observed in cobamide-dependent regulation. The common regulatory element that responds to cobamides is the cobalamin riboswitch, which is a ligand-binding RNA in the 5' untranslated region of mRNAs[83]. The cobalamin riboswitches studied so far negatively regulate expression of co-transcribed genes when bound to their cognate ligand (Fig. 5, #2)[7,9,83–88]. The *E. coli btuB* riboswitch has different binding affinity for different cobamides and Cbi, suggesting that cobamide structure can affect regulation of downstream genes, but the change in affinity due to a difference in lower ligand was modest[85]. Cobalamin riboswitches from a single organism, *Desulfitobacterium hafniense*, have different sequences and have a 1000-fold difference in binding affinities for adenosylcobalamin *in vitro*[87]. However, the measured binding affinity did not correlate with the repression of the genes measured by qPCR[87]. Further work is needed to probe the effect of specific lower ligands on ligand-binding by cobalamin riboswitches and to determine how that correlates with changes in gene expression and impacts physiology of the organism.

Cobamide-dependent enzymes each have their own cobamide selectivity. Both binding of the enzyme to the cofactor and catalytic activity of the holoenzyme have been measured (Fig. 5, #3). There are two main modes of cobamide-enzyme binding: base-off, in which the lower ligand does not coordinate the cobalt, and base-on, in which the lower ligand is coordinated to the cobalt. There is a conserved cobamide-binding motif in many cobamide-dependent enzymes[89,90], so insights into cobamide binding selectivity in one model enzyme might be applied to many enzymes. In enzymes that bind the cofactor in the base-on configuration, phenolyl cobamides that cannot form a coordination bond between the lower ligand and the cobalt are expected to be nonfunctional, such as in the class II RNR (NrdJ). For example, NrdJ-dependent growth of *S. meliloti* was supported by benzimidazolyl cobamides but not by phenolyl cobamides[74]. In contrast, MetH homologs bind cobalamin in a base-off configuration[89], and the effect of the class of lower ligand on cofactor binding and activity is not as easily predicted in base-off enzymes. *Spirulina platensis* MetH heterologously expressed in *E. coli* had similar activity with both cobalamin and pseudocobalamin, but had higher apparent affinity for pseudocobalamin[22]. Homologs of another base-off enzyme, MCM, from different organisms show different binding affinities, with the native cobamides from each organism binding best[18]. With the exception of the phenolyl cobamides, cofactor binding *in vitro* correlated well with MCM-requiring growth *in vivo*[18]. Panels of cobamide selectivity have been limited to a few model enzymes, so it is currently difficult to generalize predictions of enzyme selectivity to untested systems.

In organisms that produce cobamides, the lower ligand determinants such as the benzimidazole biosynthesis genes and the phenolic lower ligand activation genes are good predictors of which cobamide it can make, as discussed above (Fig. 5, #4). But environmental

factors can also affect which cobamide is made. One process is called guided biosynthesis, in which some bacteria that can synthesize cobamides *de novo* can incorporate exogenously provided lower ligands in place of their native lower ligand (Fig. 5, #5), even to their own detriment[19,21,74,91–94]. For example, the acetogen *Sporomusa ovata* natively produces [Cre]Cba, but when provided with exogenous benzimidazoles, it instead produces benzimidazolyl cobamides, which cannot support its growth on methanol or 3,4-dimethoxybenzoate[93]. *Sulfurospiriullum multivorans*'s natively produced cofactor for reductive dehalogenation is norpseudocobalamin, but the enzyme becomes non-functional when exogenous DMB or other benzimidazoles are added to the culture[21,94]. Guided biosynthesis can also provide access to cobamides that have more activity than the natively produced cobamide. When *S. enterica* is grown anaerobically, it produces pseudocobalamin, which cannot support its growth on ethanolamine[19]. However, it can attach DMB provided exogenously to create cobalamin and use that cofactor for ethanolamine catabolism[19]. Thus, depending on the context, guided biosynthesis can either allow access to cobamide-dependent metabolism that cannot be used with the native cobamide, or prevent growth of the organism. Benzimidazole compounds are found in multiple environments and are therefore available for guided biosynthesis, but their sources are unknown[95].

Another process that requires metabolites from the environment is cobamide precursor salvaging. Bacteria that do not have a complete cobamide biosynthesis pathway but have the nucleotide loop assembly genes may be able to take up intermediates such as Cbi and complete cobamide biosynthesis. They may produce a lower ligand *de novo* or attach exogenous lower ligands. Cbi salvaging has been observed experimentally, and Chapter 2 and Chapter 3 describe examples of the newly validated ALA salvaging, in which the genome is only missing the ALA synthesis genes. An example of Cbi salvaging is in *E. coli,* which can take up Cbi and produce [2-MeAde]Cba, or with added DMB, make cobalamin[68,96]. Members of the order Thermotogales are also able to salvage Cbi to make cobalamin[97], and in this case, the ability to do so was laterally transferred into the clade[98]. Another similar mechanism of taking up cobamide precursors involves uptake of α-ribazole by CblT, phosphorylation by CblS, and attachment of the activated lower ligand (α-ribazole phosphate) by nucleotide loop assembly genes (Fig. 3, yellow)[99,100]. *Listeria innocua* is an example of an organism that has the complete *de novo* cobamide biosynthesis pathway with the exception of *cobT*, but can acquire lower ligands in this way[100]. However, α-riboside uptake has only been observed with benzimidazole lower ligands, and it is unknown if other α-ribosides are acquired in this way. α-ribazole has been observed in microbial exudates, but how and why it is released by microbes is unknown[101,102].

Some bacteria can remodel cobamides by removing the lower ligand of a cobamide and attaching a new one produced endogenously, or an exogenous lower ligand (Fig. 5, #6). The only published genetic determinant for this activity is CbiZ, which hydrolyzes the amide in the nucleotide loop to produce cobyric acid[103–106]. To be able to attach another lower ligand, the bacterium must additionally have the genes to synthesize the nucleotide loop. *Dehalococcoides mccartyi* is an example of an organism with this set of enzymes. When provided with exogenous cobamides and benzimidazole lower ligands, it is able to convert cobamides that it cannot use for reductive dehalogenation to benzimidazolyl cobamides that it can use[107,108]. Recent work by Kenny Mok in our lab has found a different corrinoid-remodeling enzyme in *Akkermansia muciniphila*, and additional mechanisms for remodeling may remain to be discovered (Mok and Taga, unpublished). Remodeling is a process that ensures that the organism can get the correct cobamide no matter what forms are available in the environment. To use remodeling

as a strategy to acquire a useable cobamide, organisms need to have access to the correct lower ligand either by encoding its biosynthesis or by acquiring it from another community member.

**Cobamides are shared in microbial communities**

Bacteria have been shown to have a number of ways to acquire the cobamides they need. Some have selective transport, some can synthesize the cobamides they need, and some can remodel the available cobamides or salvage intermediates to form the cobamide they need. Bioinformatic analysis of microbial genomes has shown that less than half to three fourths of microbes in each study that encode enzymes that use cobamides contain all the genes to make them *de novo*, so many species must acquire environmental cobamides or cobamide precursors to use their cobamide-dependent metabolism[7–11]. Some genomics studies have focused on particular habitats, and there is variation in the predicted percentage of cobamide auxotrophs depending on the environment. In the human gut, two different studies found that only 31-42% of isolate genomes had the complete cobamide biosynthesis pathway[9,10], but 80% had cobamide-dependent enzymes[9], suggesting these cobamide auxotrophs must acquire cobamides from other community members or the host diet. In a 13-member freshwater enrichment community, metagenomics analysis revealed that 12 members can use cobamides, but only five could produce them[109]. Since two of the cyanobacteria members presumably make purinyl cobamides, it is possible that the cobamide auxotrophs would need to modify these cobamides, but the cobamide preference of the auxotrophs in this study was not determined experimentally. Genomics studies suggest that cobamides must be shared between members of microbial communities and can predict which cobamides are made by producers, but cannot give complete information on which cobamides are required by cobamide auxotrophs. Genomics analysis can also generate hypotheses about the directionality of cobamide sharing based on predicted autotrophies in both cobamide producer and consumer genomes, but experimental testing is required to validate these predictions.

In natural and synthetic cocultures, mutualistic metabolite sharing has been observed. Since the amount of cobalamin in the open ocean is too low to support growth[110], close interactions between bacteria and eukaryotic algae are thought to be a common way for the algae to acquire vitamin $B_{12}$. Cobamide-producing bacteria are frequently co-isolated with algae[111,112]. In lab cultures, eukaryotic algae produce reduced carbon that co-associated bacteria consume, and the bacteria provide the algae with vitamin $B_{12}$ to use for MetH [111–114]. One particular example is the synthetic pairing of the alga *Lobomonas rostrata* and the bacterium *Mesorhizobium loti.* In this case, production of vitamin $B_{12}$ is induced in the bacterium by the alga, and the presence of the bacterium induces gene expression changes in the alga [111,114,115]. Another model mutualism involved exchange of B vitamins and precursors between the alga *Ostreococcus tauri* and the bacterium *Dinoroseobacter shibae* in a stable co-culture [116]. Mechanistic studies with algae-bacteria cobamide sharing have involved organisms isolated from the same environment that were not isolated together; although mutualistic nutrient cross-feeding occurred, the pairs of strains used in these studies did not co-evolve. The fact that these co-cultures are so readily created suggests that this mode of nutrient sharing is common.

In other consortia, cobamide production is not rewarded with a product from the cobamide auxotroph. For example, in a tri-culture containing *Dehalococcoides mccartyi*, *Pelosinus fermentans*, and *Desulfovibrio vulgatus*, *D. mccartyi* uses cobamides produced by *P. fermentans* for reductive dehalogenation and consumes $H_2$ produced by *D. vulgatus*. However, *D. mccartyi* does not directly provide a service to *P. fermentans*[108]. Similarly, *Geobacter* strains produce cobamides that *D. mccartyi* can use for reductive dehalogenation when DMB

is added to the coculture, but there is no apparent benefit for *Geobacter*[117]. The fact that cobamide sharing occurs even without benefit to the producer leads to questions about how cobamides are released and how cobamide release evolves, which is unknown and may vary between systems. When culturing a cobamide producing bacterium such as a cyanobacterium alone, one can find cobamides in the medium[118]. One hypothesis is that cobamides are released through cell lysis. Work in our lab is ongoing to further explore this potential mechanism of cobamide release.

**Frameworks for loss of cobamide biosynthesis**

Why are so many organisms that use cobamides unable to produce them *de novo*? As the complete cobamide biosynthesis pathway is long, cobamide biosynthesis is thought to be costly. A number of frameworks provide possible explanations for gene loss and metabolic dependencies. The Black Queen Hypothesis (BQH) describes the adaptive advantage of losing a costly metabolic function when the function is sufficiently "leaky"[3,119,120]. An experimental example of the BQH was a serially-passaged culture of *E. coli* that was chromosomally deficient in catalase production carrying a catalase-containing plasmid[121]. Catalase activity is a "leaky" function because it reduces peroxide concentration in the medium for all cells. A catalase-positive and a catalase-negative population stably co-existed in a negative frequency-dependent manner without reciprocity as predicted by the BQH in the laboratory evolution[121]. The BQH suggests that loss of the cobamide biosynthesis pathway is under positive selection as long as there remain sufficient cobamides available in the environment. In mutually beneficial interactions, coevolution allows one partner to lose the biosynthesis pathway and maintain the association by providing another service to the partner producing cobamides[1]. Co-culturing of *E. coli* strains auxotrophic for individual amino acids found that 43 out of 54 pairings resulted in the auxotrophs having a greater growth rate than the prototrophic ancestor, showing the benefit to cross-feeding[122]. A synthetic example of this in cobamide cross feeding is the algae-bacterium symbiosis discussed above[35,111].

In general, there are many ways to acquire carbon, nitrogen, and energy sources without using cobamide-dependent metabolisms. Seven known processes carried out by cobamide-dependent enzymes have cobamide-independent alternatives as discussed above. For some cobamide-dependent metabolisms, bacteria will use alternative carbon and energy sources if cobamides are unavailable. This metabolic flexibility allows many bacteria to be facultatively cobamide-dependent. I speculate that in some lineages, constant availability of cobamides or ways to bypass cobamide-dependent processes by using metabolisms that do not require cobamides can place cobamide biosynthesis under relaxed selection.

**Conclusions**

Cobamide cofactors are used in diverse microbial enzymes, and microbes have evolved many strategies to acquire the specific cofactor they need. Understanding cobamide cross-feeding in microbial communities requires examination of cobamides, their precursors, and lower ligands. Further study on the mechanistic determinants of cobamide specificity at the levels of transport, regulation, and cobamide-dependent enzymes will improve the ability to predict which cobamides can and cannot be used by an organism based on its genome. In the following chapters, I focus on cobamide biosynthesis, highlighting bacterial genomes containing partial cobamide pathways and potential precursor salvaging activity as an underexplored strategy for acquiring a specific cobamide.

# Chapter 2

## Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics

**Published:** Amanda N. Shelton, Erica C. Seth, Kenny C. Mok, Andrew W. Han, Samantha N. Jackson, David R. Haft, Michiko E. Taga. "Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics." *The ISME Journal* (2019) **13,** 789-804.

**Abstract**

The vitamin $B_{12}$ family of cofactors known as cobamides are essential for a variety of microbial metabolisms. We used comparative genomics of 11,000 bacterial species to analyze the extent and distribution of cobamide production and use across bacteria. We find that 86% of bacteria in this data set have at least one of 15 cobamide-dependent enzyme families, yet only 37% are predicted to synthesize cobamides *de novo*. The distribution of cobamide biosynthesis and use vary at the phylum level. While 57% of Actinobacteria are predicted to biosynthesize cobamides, only 0.6% of Bacteroidetes have the complete pathway, yet 96% of species in this phylum have cobamide-dependent enzymes. The form of cobamide produced by the bacteria could be predicted for 58% of cobamide-producing species, based on the presence of signature lower ligand biosynthesis and attachment genes. Our predictions also revealed that 17% of bacteria have partial biosynthetic pathways, yet have the potential to salvage cobamide precursors. These include a newly defined, experimentally verified category of bacteria lacking the first step in the biosynthesis pathway. These predictions highlight the importance of cobamide and cobamide precursor salvaging as examples of nutritional dependencies in bacteria.

### Introduction

Microorganisms almost universally reside in complex communities where individual members interact with each other through physical and chemical networks. A major type of chemical interaction is nutrient salvaging, in which microbes that lack the ability to synthesize particular required nutrients (termed auxotrophs) obtain these nutrients from other organisms in their community[123]. By understanding which organisms require nutrients and which can produce them, we can predict specific metabolic interactions between members of a microbial community[124]. With the development of next-generation sequencing, the genome sequences of tens of thousands of bacteria from diverse environments are now available, leading to the possibility of predicting community interactions based on the genomes of individual members. However, the power to predict the metabolism of an organism by analyzing its genome remains limited.

The critical roles of cobamides (the vitamin $B_{12}$ family of enzyme cofactors) in the metabolism of humans and diverse microbes have long been appreciated. Only recently, however, has cobamide-dependent metabolism been recognized as a potential mediator of microbial interactions[123,125,126]. Cobamides are used in a variety of enzymes in prokaryotes, including those involved in central metabolic processes such as carbon metabolism and the biosynthesis of methionine and deoxynucleotides[23] (Fig. 1). Some of the functions carried out

by cobamide-dependent pathways, such as acetogenesis via the Wood-Ljungdahl pathway in anaerobic environments, can be vital in shaping microbial communities[127]. Cobamides are also used for environmentally and industrially important processes such as reductive dehalogenation and natural product synthesis[13,128].

*De novo* cobamide biosynthesis involves approximately 30 steps[46], and the pathway can be divided into several segments (Fig. 2). The first segment, tetrapyrrole precursor biosynthesis, contains the first five steps of the pathway, most of which are also common to the biosynthesis of heme, chlorophyll, and other tetrapyrroles. The next segment, corrin ring biosynthesis, is divided into oxygen-sensitive (anaerobic) and oxygen-dependent (aerobic) routes, depending on the organism. These two alternative pathways then converge at a late intermediate, which is further modified to form the cobamide (Fig. 2, nucleotide loop assembly). The latter portion of the pathway involves adenosylation of the central cobalt ion followed by the synthesis and attachment of the aminopropanol linker and lower axial ligand (Fig. 2). Investigation of cobamide salvaging must account for structural diversity in the lower ligand (Fig. 2B), as only a subset of cobamide cofactors can support growth of any individual organism[9,11,20,21,93,107,117]. Recent work has identified many of the genetic determinants for the biosynthesis of the benzimidazole class of lower ligands[66–68,70,72] and attachment of phenolic lower ligands[64,65] (Fig. 2).

Previous analyses of bacterial genomes have found that less than half to three fourths of prokaryotes that require cobamides are predicted to make them[7,8], suggesting that cobamide salvaging may be widespread in microbial communities. Analyses of cobamide biosynthesis in the human gut[9,10] and in the phylum Cyanobacteria[11] further reinforce that cobamide-producing and cobamide-dependent bacteria coexist in nature. These studies provide valuable insights into the extent of cobamide use and biosynthesis in bacteria, but are limited in the diversity and number of organisms studied and have limited prediction of cobamide structure.

Here, we have analyzed the genomes of over 11,000 bacterial species and generated predictions of cobamide biosynthesis, dependence, and structure. We predict that 86% of sequenced bacteria are capable of using cobamides, yet only 37% produce cobamides *de novo*. We were able to predict cobamide structure for 58% of cobamide producers. Additionally, our predictions revealed that 17% of bacteria can salvage cobamide precursors, of which we have defined a new category of bacteria that require early tetrapyrrole precursors to produce cobamides.

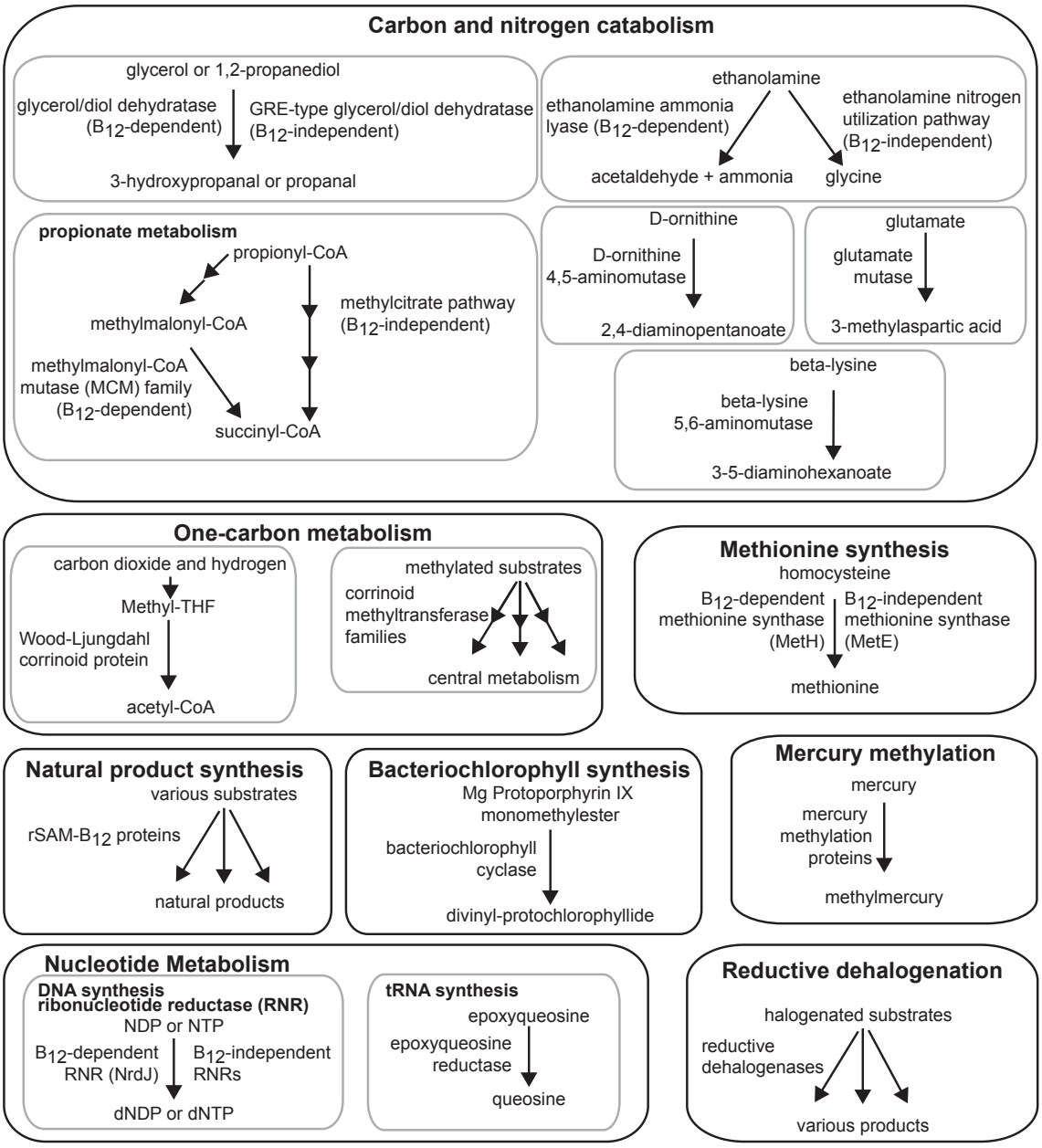



**Figure 1: Functions carried out by cobamide-dependent processes.** Reactions carried out by cobamide-dependent enzymes are shown on the left side of the arrows and cobamide-independent alternative processes, if known, on the right. Annotations or query genes used for searching for each function are listed in Table 1 and Table 2.

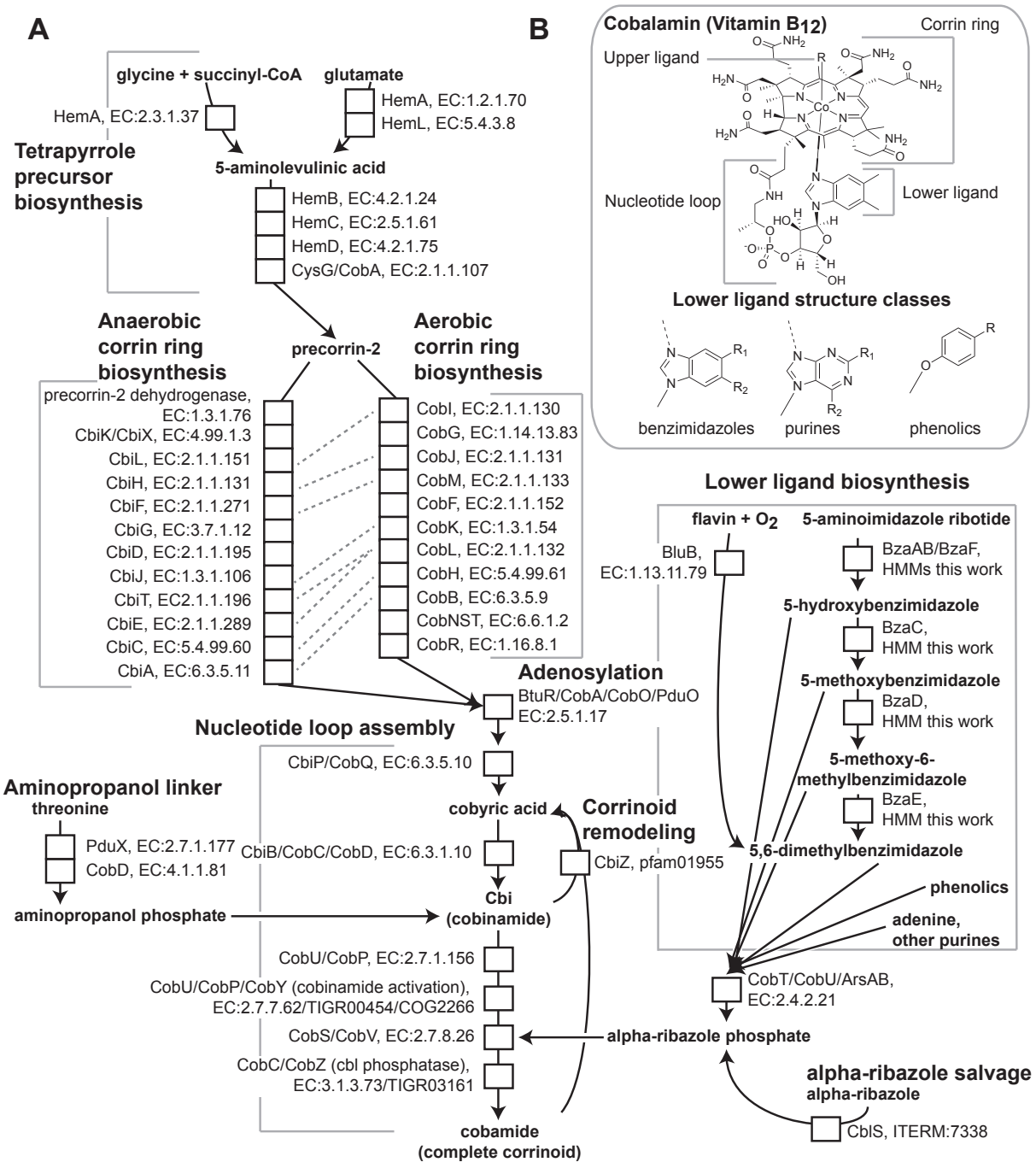



**Figure 2: Cobamide biosynthesis and structure. A.** The cobamide biosynthesis pathway is shown with each enzymatic step indicated by a white box labeled with the gene names and functional annotation. Subsections of the pathway and salvaging and remodeling pathways are bracketed or boxed with labels in bold. Orthologous enzymes that carry out similar reactions in aerobic and anaerobic corrin ring biosynthesis are indicated by dashed lines. **B.** Structure of cobalamin. The upper ligand R can be an adenosyl or methyl group. Classes of possible lower ligand structures are also shown. Benzimidazoles: $R_1$=H, OH, $CH_3$, $OCH_3$; $R_2$=H, OH, $CH_3$, $OCH_3$. Purines: $R_1$=H, $CH_3$, $NH_2$; $R_2$=H, $NH_2$, OH, O. Phenolics: R=H, $CH_3$.

## Materials and Methods

### Data set download and filtering

The names, unique identifiers, and metadata for 44,802 publicly available bacterial genomes on the Joint Genome Institute's Integrated Microbial Genomes with Expert Review database (JGI/IMGer, https://img.jgi.doe.gov/cgi-bin/mer/main.cgi)[129] classified as "finished" (accessed January 11, 2017) or "permanent draft" (accessed February 23, 2017) were downloaded. To assess genome completeness, we searched for 55 single copy gene annotations[130,131] using the "function profile: genomes vs functions" tool in each genome. Completeness was measured first based on the unique number of single copy gene annotation hits (55/55 was best) and second, by the average copy number of the annotations (values closest to 1 were considered most complete). We removed 2,776 genomes with fewer than 45 out of 55 unique single copy genes. To filter the remaining genomes to one genome per species, we used name-based matching to create species categories, in which each unique binomial name was considered a single species. The genome with the highest unique single copy gene number and had an average single copy gene number closest to 1 was chosen to represent a species. If both scores were identical the representative genome was chosen at random. For strains with genus assignments, but without species name assignments, we considered each genome to be a species. The list of species was manually curated for species duplicates caused by data entry errors.

### Detection of cobamide biosynthesis and dependence genes in genomes

Annotations from Enzyme Commission (EC) numbers (http://www.sbcs.qmul.ac.uk/iubmb/enzyme/), Pfam, TIGRFAM, Clusters of Orthologous Groups (COG), and IMG Terms[129,132–135] for cobamide biosynthesis, cobamide-dependent enzymes, and cobamide-independent alternative annotations were chosen. These included annotations used by Degnan *et al.*[9], but in other cases alternative annotations were chosen to improve specificity of the identified genes (Table 1). For example, EC: 4.2.1.30 for glycerol dehydratase identifies both cobamide-dependent and -independent isozymes, so pfam annotations specific to the cobamide-dependent version were used instead. These genes were identified in each genome using the "function profile: genomes vs functions" tool (Jan-May 2017).

For genes without functional annotations in the IMGer database, we chose sequences that were genetically or biochemically characterized[40,45,136,137] to use as the query genes in one-way BLASTP[138] against the filtered genomes using the IMGer "gene profile: genomes vs genes" tool, accessed Jan-May 2017 (Table 2).

Output files for the cobamide genes were combined into a master file in Microsoft Excel. This master file was used as input for custom python 2.7 code that interpreted the presence or absence of genes as predicted phenotypes. We used Microsoft Excel and python for further analysis. Genomes were scored for the presence or absence of cobamide-dependent enzymes and alternatives based on the annotations in Table 1 and Table 2. We then created criteria for seven cobamide biosynthesis phenotypes based on the presence of certain sets of cobamide biosynthesis genes (Table 3, 4): very likely cobamide producer, likely cobamide producer, possible cobamide producer, tetrapyrrole precursor salvager, cobinamide (Cbi) salvager, likely non-producer, and very likely non-producer and classified genomes accordingly. These are grouped into complete biosynthesis (very likely, likely, and possible cobamide producer),

partial biosynthesis (tetrapyrrole precursor salvager and Cbi salvager), and no biosynthesis (likely non-producer and very likely non-producer).

During cobamide biosynthesis, the lower ligand base is activated by CobT to allow attachment to the nucleotide loop. For phenolic lower ligands, this reaction is carried out by ArsA and ArsB, subfamilies of *cobT* homologs found in tandem[64,69]. To distinguish putative *arsAB* homologs from other *cobT* homologs that are not known to produce phenolyl cobamides, IMGer entries for all genes that were annotated as *cobT* homologs were downloaded. Tandem *cobT* homologs were defined as those with sequential IMG gene IDs. This list of tandem *cobT* genes was then filtered by size to eliminate genes encoding less than 300 or more than 800 amino acid residues, indicating annotation errors (CobT is approximately 350 AA residues). The remaining tandem *cobT* homologs were assigned as putative *arsAB* homologs.

To identify the anaerobic benzimidazole biosynthesis genes *bzaABCDEF*, four new hidden Markov model profiles (HMMs) were created and two preexisting ones (TIGR04386 and TIGR04385) were refined. Generally, the process for generating the new HMMs involved performing a Position-Specific Iterated (PSI) BLAST search using previously classified instances of the Bza proteins aligned in Jalview[138,139]. Due to their similarity, BzaA, BzaB, and BzaF were examined together, as were BzaD and BzaE. To help classify these sequences, Training Set Builder (TSB) was used[140]. All six HMMs have not been assigned TIGRFAM accessions at the time of publication, but will be included in the next TIGRFAM release. Protein sequences for 10591 of the filtered genomes were queried for each *bza* HMM using hmm3search (HMMER3.1)[141]. Hits are only reported above the trusted cutoff defined for each HMM. A hit for *bzaA* and *bzaB* or *bzaF* indicated that the genome had the potential to produce benzimidazole lower ligands. The specific lower ligand was predicted based on the *bza* genes present[68].

We used BLASTP on IMGer to search for tetrapyrrole precursor biosynthesis genes that appeared to be absent in the 201 species identified as tetrapyrrole precursor salvagers. Query sequences used were the following: *Rhodobacter sphaeroides* HemA (GenPept C49845); *Clostridium saccharobutylicum* DSM 13864 HemA, HemL, HemB, HemC, and HemD (GenBank: AGX44136.1, AGX44131.1, AGX44132.1, AGX44134.1, AGX4133.4, respectively). We additionally searched for the *Bacillus subtilis* HemD, which only has the UroIII synthase activity (UniProtKB P21248.2). We visually inspected the ORFs near any BLASTP hits in the IMGer genome browser. One hundred eighty species remained after this analysis. Genomes were classified as a particular type of tetrapyrrole precursor salvager only if they were missing all genes upstream of a precursor.

**Table 1. Annotations used for cobamide-dependent enzyme families and alternatives**.

| Process name | Gene symbols | Pathway | Annotation used | Reference |
|---|---|---|---|---|
| ribonucletotide reductase ($B_{12}$-independent) | NrdAB, NrdDG | deoxyribonucleotide synthesis | TIGR02487, TIGR02506, or TIGR02491 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |

| | | | | |
|---|---|---|---|---|
| ribonucletotide reductase ($B_{12}$-dependent) | NrdJ | deoxyribonucleotide synthesis | TIGR02504, TIGR02505 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| methionine synthase ($B_{12}$-dependent) | MetH | methionine synthesis | EC:2.1.1.13 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| methionine synthase ($B_{12}$-independent) | MetE | methionine synthesis | EC:2.1.1.14 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| glutamate mutase | GlmES | glutamate metabolism | EC:5.4.99.1 | Banerjee R, Ragsdale SW. (2003). THE MANY FACES OF VITAMIN B12: CATALYSIS BY COBALAMIN-DEPENDENT ENZYMES. *Annu Rev Biochem* **72**: 209-247.[13] |
| methylmalonyl-CoA mutase family | MCM, MutA, MeaA, IcmAB | propionate metabolism (MCM) | (TIGR0640 and TIGR0641) or EC:5.4.99.2 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| methylcitrate pathway ($B_{12}$-independent) | PrpBCDF, AcnBD, SucD | propionate metabolism (methylcitrate) | EC:2.3.3.5, EC:4.1.3.30, EC:4.2.1.3, EC:4.2.1.99, (methylcitrate dehydratase: either EC:4.2.1.79, EC:4.2.1.117), (succinyl-CoA synthase alph subunit: either EC:6.2.1.4, EC:6.2.1.5) -at least 4 steps | Textor S, Wendisch VF, De Graaf AA, Müller U, Linder MI, Linder D, et al. (1997). Propionate oxidation in Escherichia coli: Evidence for operation of a methylcitrate cycle in bacteria. *Arch Microbiol* **168**: 428-436.[142] |
| ethanolamine ammonia lyase | EutBC | ethanolamine utilization as nitrogen or carbon source | EC:4.3.1.7 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| $B_{12}$-dependent glycerol/diol dehydratase | PduCDE, DhaBCD | propanediol utilization, glycerol fermentation | pfam02286, pfam02287, pfam02288 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181. |

| | | | | |
|---|---|---|---|---|
| beta-lysine 5,6-aminomutase | KamED | lysine degradation | EC:5.4.3.3 | Banerjee R, Ragsdale SW. (2003). THE MANY FACES OF VITAMIN B12: CATALYSIS BY COBALAMIN-DEPENDENT ENZYMES. *Annu Rev Biochem* **72**: 209-247. [13] |
| D-ornithine 4,5-aminomutase | OraSE | ornithine degradation | EC:5.4.3.5 | Banerjee R, Ragsdale SW. (2003). THE MANY FACES OF VITAMIN B12: CATALYSIS BY COBALAMIN-DEPENDENT ENZYMES. *Annu Rev Biochem* **72**: 209-247. [13] |
| epoxyqueonsine reductase | QueG | queosine synthesis | TIGR00276 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| reductive dehalogenase | PceA | reductive dehalogenation | TIGR02486 | Banerjee R, Ragsdale SW. (2003). THE MANY FACES OF VITAMIN B12: CATALYSIS BY COBALAMIN-DEPENDENT ENZYMES. *Annu Rev Biochem* **72**: 209-247. [13] |
| Wood-Ljungdahl corrinoid protein | AcsCDE | acetogenesis | TIGR00381, TIGR00316, EC:2.1.1.258, EC:2.1.1.245 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| corrinoid methyltransferase families | MtaAB, MtbAB, MtmB, MttB, MtqAB | acetogenesis, methanogenesis, 1-C metabolism | EC:2.1.1.246, EC:2.1.1.247, EC:2.1.1.248, EC:2.1.1.249, EC:2.1.1.250, EC:2.1.1.252, EC:2.1.1.253, or EC:2.1.1.90 | Roth JR, Lawrence JG, Bobik TA. (1996). Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu Rev Microbiol* **50**: 137-181.[23] |
| bacteriochlorophyll cyclase | BchE | bacteriochlorophyll biosynthesis | TIGR02026 | Gough SP, Petersen BO, Duus JO. (2000). Anaerobic chlorophyll isocyclic ring formation in Rhodobacter capsulatus requires a cobalamin cofactor. *Proc Natl Acad Sci U S A* **97**: 6908-13.[143] |
| rSAM-$B_{12}$ proteins | | various | pfam02310 and pfam04055 in same ORF | Broderick JB, Du BR, Duschene KS, Shepard EM. (2014). Radical S-Adenosylmethionine Enzymes. *Chem Rev* **114**: 4229-4317.[128] |
| corrinoid binding domain | | various | pfam02310, pfam02607 | |

**Table 2. Query genes used for BLASTP-based search for enzyme families without annotations.**

| Process name | Gene symbols | Pathway | Query used in BLASTP | Settings for BLASTP | Reference |
|---|---|---|---|---|---|
| ethanolamine nitrogen utilization pathway ($B_{12}$-independent) | Csal_0675, Csal_0679, Csal_0680, Csal_0681 | ethanolamine utilization as nitrogen source | Chromohalobacter salexigens genes Csal_0675, Csal_0679, Csal_0680, Csal_0681 (IMG gene_oid:637965845, 637965849, 637965850, 637965851) | 1-way BLASTP with minimum evalue=1e-10, minimum percent identity = 40%, genome had to have all four genes | Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, Wichelecki DJ, et al. (2015). Experimental strategies for functional annotation and metabolism discovery: Targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**: 909-931.[40] |
| $B_{12}$-independent glycerol/diol dehydratase | DhaB, RiDD | propanediol utilization, glycerol fermentation | Clostiridum butryicum GRE (IMG gene_oid: 2688143584) or Roseburia inulinivorans GRE (IMG gene_oid: 644270222) or Rhodopseudomonas palustrus Bis188 GRE (IMG gene_oid:637924274) | 1-way BLASTP with minimum evalue=1e-10, minimum percent identity=50% | Brien JRO, Raynaud C, Croux C, Girbal L, Soucaille P, Lanzilotta WN, et al. (2004). Insight into the Mechanism of the B12-Independent Glycerol Dehydratase from Clostridium butryicum: Preliminary Biochemical and Structural Characterization. *Biochemistry* **43**: 4635-4645[137]<br>LaMattina JW, Keul ND, Reitzer P, Kapoor S, Galzerani F, Koch DJ, et al. (2016). 1,2-propanediol Dehydration in Roseburia inulinivorans; Structural Basis for Substrate and Enantiomer Selectivity. *J Biol Chem* **291**: 15515-15526[42]<br>Zarzycki J, Sutter M, Cortina NS, Erb TJ, Kerfeld CA. (2017). In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. *Sci Rep* **7**: 42757.[45] |
| mercury methylation | HgcAB | mercury methylation | Desufovibrio desulfuricans ND-132 HgcAB (IMG gene_oid: 2503785873 and 2503785874) | 1-way BLASTP with minimum evalue=1e-10, minimum percent identity=30% | Parks JM, Johs A, Podar M, Bridou R, Hurt RA, Smith SD, et al. (2013). The Genetic Basis for Bacterial Mercury Methylation. *Science* **339** (6215): 1332-1335.[136] |

**Table 3. Definitions of the cobamide biosynthesis pathway sections used in classifying genomes.**

| Pathway section | Steps |
|---|---|
| Complete aerobic biosynthesis (23 total) | ALA synthesis (either EC:2.3.1.37 or both EC:1.2.1.70 and EC:5.4.3.8), EC:4.2.1.24, EC:2.5.1.61, EC:4.2.1.75, EC:2.1.1.107, EC:2.1.1.130, EC: 1.14.13.83, EC:2.1.1.131, EC:2.1.1.133, EC:2.1.1.152, EC:1.3.1.54, EC:2.1.1.132, EC:5.4.99.61, EC:6.3.5.9, EC:6.6.1.2, EC:2.5.1.17, EC:6.3.5.10, EC:6.3.1.10, EC:2.7.1.156, cobinamide activation (EC:2.7.7.62 or TIGR00454 or COG2266), cobalamin phosphatase (EC:3.1.3.73 or TIGR03161), EC:2.7.8.26 |
| Complete anaerobic biosynthesis (25 total) | ALA synthesis (either EC:2.3.1.37 or both EC:1.2.1.70 and EC:5.4.3.8), EC:4.2.1.24, EC:2.5.1.61, EC:4.2.1.75, EC:2.1.1.107, EC:1.3.1.76, EC:4.99.1.3, EC:2.1.1.151, EC:2.1.1.131, EC:2.1.1.271, EC:3.7.1.12, EC:2.1.195, EC:1.3.1.106, EC:2.1.1.196, EC:2.1.1.289, EC:5.4.99.60, EC:6.3.5.11, EC:2.5.1.17, EC:6.3.5.10, EC:6.3.1.10, EC:2.7.1.156, cobinamide activation (EC:2.7.7.62 or TIGR00454 or COG2266), cobalamin phosphatase (EC:3.1.3.73 or TIGR03161), EC:2.7.8.26 |
| Tetrapyrrole precursor biosynthesis (5 total) | ALA synthesis (either EC:2.3.1.37 or both EC:1.2.1.70 and EC:5.4.3.8), EC:4.2.1.24, EC:2.5.1.61, EC:4.2.1.75, EC:2.1.1.107 |
| Combined corrin ring biosynthesis (9 total) | (EC:2.1.1.130 or EC:2.1.1.151), (EC:2.1.1.131 or both EC:2.1.1.131 and EC:1.14.13.54), (EC:2.1.1.133 or EC:2.1.1.271), (EC:2.1.1.152 or both EC:3.7.1.12 and EC:2.1.1.195), (EC:1.3.1.106 or EC:1.3.1.54 ), (EC:2.1.1.132 or both EC:2.1.1.289 and EC:2.1.1.196), (EC:5.4.99.61 or EC:5.4.99.60), (EC:6.3.5.9 or EC:6.3.5.11), and (EC:6.6.1.2 or EC:4.99.1.3). |
| Aminopropanol Linker | EC:2.7.1.177, EC:4.1.1.81 |
| Adenosylation | EC:2.5.1.17 |
| Nucleotide loop assembly (7 total) | EC:6.3.5.10, EC:6.3.1.10, EC:2.7.1.156, cobinamide activation (EC:2.7.7.62 or TIGR00454 or COG2266), cobalamin phosphatase (EC:3.1.3.73 or TIGR03161), EC:2.7.8.26 |
| Core biosynthesis genes (8 total) | (EC:2.1.1.130 or EC:2.1.1.151), (EC:2.1.1.133 or EC:2.1.1.271), (EC:5.4.99.61 or EC:5.4.99.60), EC:6.3.5.10, EC:6.3.1.10, EC:2.7.1.156, cobinamide activation (EC:2.7.7.62 or TIGR00454 or COG2266), EC:2.7.8.26 |

**Table 4. Definitions of cobamide biosynthesis categories.**

| Classification | Criteria |
|---|---|
| very likely cobamide producer | 25/25 anaerobic steps or 23/23 aerobic steps |
| likely cobamide producer | ≥4/5 tetrapyrrole precursor biosynthesis steps and (either ≥90% anaerobic steps or ≥90% aerobic steps) |
| possible cobamide producer | ≥6/9 corrin ring biosynthesis steps and (either ≥18/25 anaerobic steps or ≥16/23 aerobic steps) or ≥16/21 (tetrapyrrole precursor + corrin ring + end biosynthesis steps) |
| tetrapyrrole precursor salvager | missing ALA synthesis (either EC:2.3.1.37 or both Ec:1.2.1.70 and EC:5.4.3.8) and has ≥5/9 corrin ring steps and (any of ≥18/25 anaerobic steps, ≥16/23 aerobic steps or ≥14/16 (corrin ring + end biosynthesis steps)) |
| cobinamide (cbi) salvage | not any of the producer categories nor tetrapyrrole precursor salvager and has ≥5 end biosynthesis steps |
| likely non-producer | did not fit any other category |
| very likely non-producer | not any of the producer categories or partial biosynthesis categories and has ≤5/9 corrin ring steps |

## Strains and growth conditions

*Clostridium scindens* ATCC 35704, *Clostridium sporogenes* ATCC 15579, and *Treponema primitia* ZAS-2 (gift from Jared Leadbetter) were grown anaerobically with and without added 5-aminolevulinic acid (1 mM for *C. sporogenes* and *T. primitia*, 0.5 mM for *C. scindens*).

*Desulfotomaculum reducens* MI-1 (gift from Rizlan Bernier-Latmani), *Listeria monocytogenes* (gift from Daniel Portnoy), *Blautia hydrogenotrophica* DSM 10507, *Clostridium kluyveri* DSM 555 (gift from Rolf Thauer), and *Clostridium phytofermentans* ISDg (gift from Susan Leschine) were grown anaerobically.

*Clostridium scindens* ATCC35704 was grown at 37ºC under 80% $N_2$, 20% $CO_2$ in an anaerobic defined mineral salts medium with the following composition (g/L): NaCl, 1; $MgCl_2 \cdot 6H_2O$, 0.5; $KH_2PO_4$, 0.2; $NH_4Cl$, 0.3; KCl, 0.3; $CaCl_2 \cdot 2\ H_2O$, 0.015. In addition, 2.29 g of N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES, free acid), 2 ml of a trace element solution[144], 1 ml of a $Na_2SeO_3$-$Na_2WO_4$ solution[145],10 mg of resazurin, and 40 mg each of the amino acids arginine, cysteine, glycine, histidine, isoleucine, leucine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine were added per liter[146]. After the medium was boiled and cooled under $N_2$, the gas was switched to an 80% $N_2$, 20% $CO_2$ mix, and the reductants $Na_2S \cdot 9\ H_2O$ and L-cysteine were added to final concentrations of 0.2 mM each. Next, 2.52 g $NaHCO_3$ (30 mM final concentration) was added to the medium, and the pH was adjusted to 7.0. The medium was dispensed under 80% $N_2$, 20% $CO_2$ in 10 ml aliquots in 25 ml Balch tubes, or for large volumes, 1 L in 2 L pyrex bottles. Tubes and bottles were sealed with butyl stoppers and aluminum crimp seals, autoclaved for 30 min, and cooled to room

temperature. Glucose was subsequently added to a final concentration of 25 mM, and Wolin vitamin solution[147] (prepared without cobalamin) was added to a final concentration of 1% (v/v). Where indicated, 1 mM ALA was added after autoclaving. The trace element stock solution contained (g/L) Nitriloacetic acid, 1.11; $MnSO_4 \cdot H_2O$, 0.5; $FeSO_4 \cdot 7H_2O$, 0.1; $CoCl_2 \cdot 6H_2O$, 0.1; $ZnCl_2$, 0.1, $NiCl_2 \cdot 6H_2O$, 0.05; $CuSO_4 \cdot 5H_2O$, 0.01; $AlK(SO_4)_2 \cdot 12\ H_2O$, 0.01; $H_3BO_3$, 0.01; $Na_2MoO_4 \cdot 2H_2O$, 0.01. The $Na_2SeO_3 \cdot Na_2WO_4$ solution contained (g/L): $Na_2SeO_3 \cdot 5H_2O$, 0.006; $Na_2WO_4 \cdot 2H_2O$, 0.008; NaOH, 0.5.

*Clostridium sporogenes* ATCC 15579 was grown in the same medium and conditions as *C. scindens* with the following changes: cysteine, serine, and threonine were omitted, and 1 mL of a vitamin solution containing 500 mg/L nicotinic acid, 50 mg/L thiamine HCl, 5 mg/L biotin, and 5 mg/L p-aminobenzoic acid was added per liter of medium[146].

*Treponema primitia* ZAS-2 was grown at room temperature in anaerobic 4YACo medium with a headspace of 80% $H_2$, 20% $CO_2$ as previously described[148], with the following changes. For cobalamin added cultures, the final concentration of cyanocobalamin was reduced from 4.42 μM to 37 nM. For testing no addition and ALA addition to cultures, cobalamin-supplemented cultures were serially passaged three times in cobalamin-free medium or in cobalamin-free medium containing 1 mM ALA before being used as inocula for growth experiments. Growth was monitored spectrophotometrically ($O.D._{.650}$). All growth experiments were performed in triplicate.

*Desulfotomaculum reducens* MI-1 was grown anaerobically at 37°C under an $N_2$ atmosphere in a modified Widdel low phosphate medium with the following per liter: $NH_4Cl$, 0.25 g; $CaCl_2 \cdot 2\ H_2O$, 0.1 g; $MgCl_2 \cdot 6\ H_2O$, 0.5 g; NaCl, 5 g; KCl, 0.5 g; $KH_2PO_4$, 0.03 g; TES, 2.292 g; yeast extract, 0.5 g; $Na_2SO_4$, 2.84 g; $NaHCO_3$, 2.52 g; cysteine-sulfide solution (2.5%), 4 ml; Se/WO solution, 1 ml[145]; Trace Elements Solution SL-10, 1 ml[149]; 1000X Wolin's Vitamin solution (without cobalamin), 2 ml. Na-lactate was added as the electron donor (20 mM) and the pH of the medium was adjusted to 7.3 with KOH.

*Listeria monocytogenes* was grown anaerobically at 25°C under an $N_2$ atmosphere in medium containing the following per liter: NaCl, 1 g; $MgCl_2 \cdot 6\ H_2O$, 0.5 g; $KH_2PO_4$, 0.2 g; $NH_4Cl$, 0.3 g; KCl, 0.3 g; $CaCl_2 \cdot 2\ H_2O$, 15 mg; Na-pyruvate, 2.2 g; yeast extract, 1 g; 1,2-propanediol, 5.9 ml; Trace Elements Solution SL-10, 1 ml; 1000X Wolin's Vitamin solution (without cobalamin), 1 ml; Se/WO solution, 1 ml. The pH of the medium was adjusted to 7.

*Blautia hydrogenotrophica* DSM 10507 was grown anaerobically in Anaerobic Basal Broth (Hi-Media M1636) under 5% $CO_2$, 10% $H_2$, 85% $N_2$ headspace at 37ºC.

*Clostridium kluyveri* DSM 555 was grown anaerobically at 33°C under an atmosphere of 80% $N_2$, 20% $CO_2$ in medium (pH 6.8-7) with the following per liter: Potassium acetate, 10 g; $K_2HPO_4$, 0.31 g; $KH_2PO_4$, 0.23 g; $NH_4Cl$, 0.25 g; $MgSO_4 \cdot 7\ H_2O$, 0.2 g; yeast extract, 1 g; ethanol, 20 ml; $NaHCO_3$ (9.1% solution), 27 ml; cysteine-sulfide solution (2.5%), 4 ml; Se/WO solution, 0.5 ml; Trace Elements Solution SL-10, 1ml; 1000X Wolin's Vitamin solution (without cobalamin), 1 ml.

*Clostridium phytofermentans* ISDg (ATCC 700394) was grown anaerobically at 25°C under an $N_2$ atmosphere in GS-2CB medium[150].

**Corrinoid extraction and analysis**

Corrinoid extractions were performed as described[107]. For corrinoids extracted from 1 L cultures of *C. sporogenes, C. scindens*, and *T. primitia*, high performance liquid chromatography (HPLC) analysis was performed with an Agilent Series 1200 system (Agilent

Technologies, Santa Clara, CA) equipped with a diode array detector with detection wavelengths set at 362 and 525 nm. Samples were injected onto an Agilent Eclipse XDB C18 column (5 µm, 4.6 x 150 mm) at 35 °C, with 0.5 ml/min flow rate. Compounds in the samples were separated using acidified water and methanol (0.1% formic acid) with a linear gradient of 18 to 30% acidified methanol over 20 min.

For all other bacteria excluding *B. hydrogenotrophica*, extracted corrinoids were analyzed as above, except with a 1.5 ml/min flow rate and a 40°C column. Corrinoids were eluted with the following method: 2% acidified (0.1% formic acid) methanol for 2 min, 2% to 10% acidified methanol in 0.1 min, and 10 to 40% acidified methanol over 9 min.

For *B. hydrogenotrophica,* corrinoids were analyzed as above with the following changes. Samples were injected onto an Agilent Zorbax SB-Aq column (5 µm, 4.6 x 150 mm) with 1 ml/min flow rate at 30 ºC. The samples were separated with a gradient of 25 to 34% acidified (0.1% formic acid) methanol over 11 minutes, followed by 34 to 50% over 2 min and 50 to 75% over 9 minutes.

**Results**

**Most bacteria are predicted to have at least one cobamide-dependent enzyme**

We surveyed publicly available bacterial genomes for 51 functions involved in cobamide biosynthesis, modification and salvage, as well as 15 cobamide-dependent enzyme families and five cobamide-independent alternative enzymes and pathways. To make generalizations about the abundances of bacteria with cobamide-dependent metabolisms and biosynthesis, the data set was reduced to representative strains for 11,436 species from approximately 45,000 available genomes. Our results indicate that the capability to use cobamides is widespread in bacteria. Eighty-six percent of species in the filtered data set have at least one of the 15 cobamide-dependent enzyme families shown in Fig. 1, Table 1, and Table 2, and 88% of these species have more than one family (Fig. 3A). This is consistent with previous analyses of smaller data sets[7–9]. The four major phyla in the data set have different distributions of the number of cobamide-dependent enzyme families per genome, with the Proteobacteria and Bacteroidetes having higher mean numbers of enzyme families than the Firmicutes and Actinobacteria (Fig. 3A). The most abundant cobamide-dependent enzymes are involved in core metabolic processes such as methionine synthesis and nucleotide metabolism, whereas processes such as reductive dehalogenation and mercury methylation are less abundant (Fig. 3B). We also observe phylum-level differences in the relative abundance of cobamide-dependent enzyme families (Fig. 3B), most notably the nearly complete absence of epoxyqueuosine reductase in Actinobacteria. Nonetheless, the cobamide-dependent methionine synthase (MetH) and, to a lesser extent, methylmalonyl-CoA mutase (MCM) and the cobamide-dependent ribonucleotide reductase (RNR), are the most abundant cobamide-dependent enzyme families in all of the four phyla (Fig. 3B).

For some cobamide-dependent processes, cobamide-independent alternative enzymes or pathways also exist (Fig. 1, right side of arrows). For example, we find that the occurrence of MetH is more common than the cobamide-independent methionine synthase, MetE, but that most bacteria have both enzymes (Fig. 3C). In contrast, cobamide-independent RNRs are found more commonly than the cobamide-dependent versions, and 30% of genomes have both cobamide-dependent and -independent RNRs (Fig. 3C). The cobamide-dependent propionate (which uses MCM), ethanolamine, and glycerol/propanediol metabolisms appear more abundant than the cobamide-independent alternatives (Fig. 3C). However, the abundance of the cobamide-

dependent propionate metabolism is overestimated because the MCM annotation used in this analysis includes mutases for which cobamide-independent versions have not been found. The abundance of both the ethanolamine and glycerol/propanediol cobamide-independent functions may be underestimated, as they were identified based on similarity to a limited number of sequences. We did not observe dramatic phylum-level differences in the relative abundances of cobamide-dependent and –independent processes (Fig. 4).

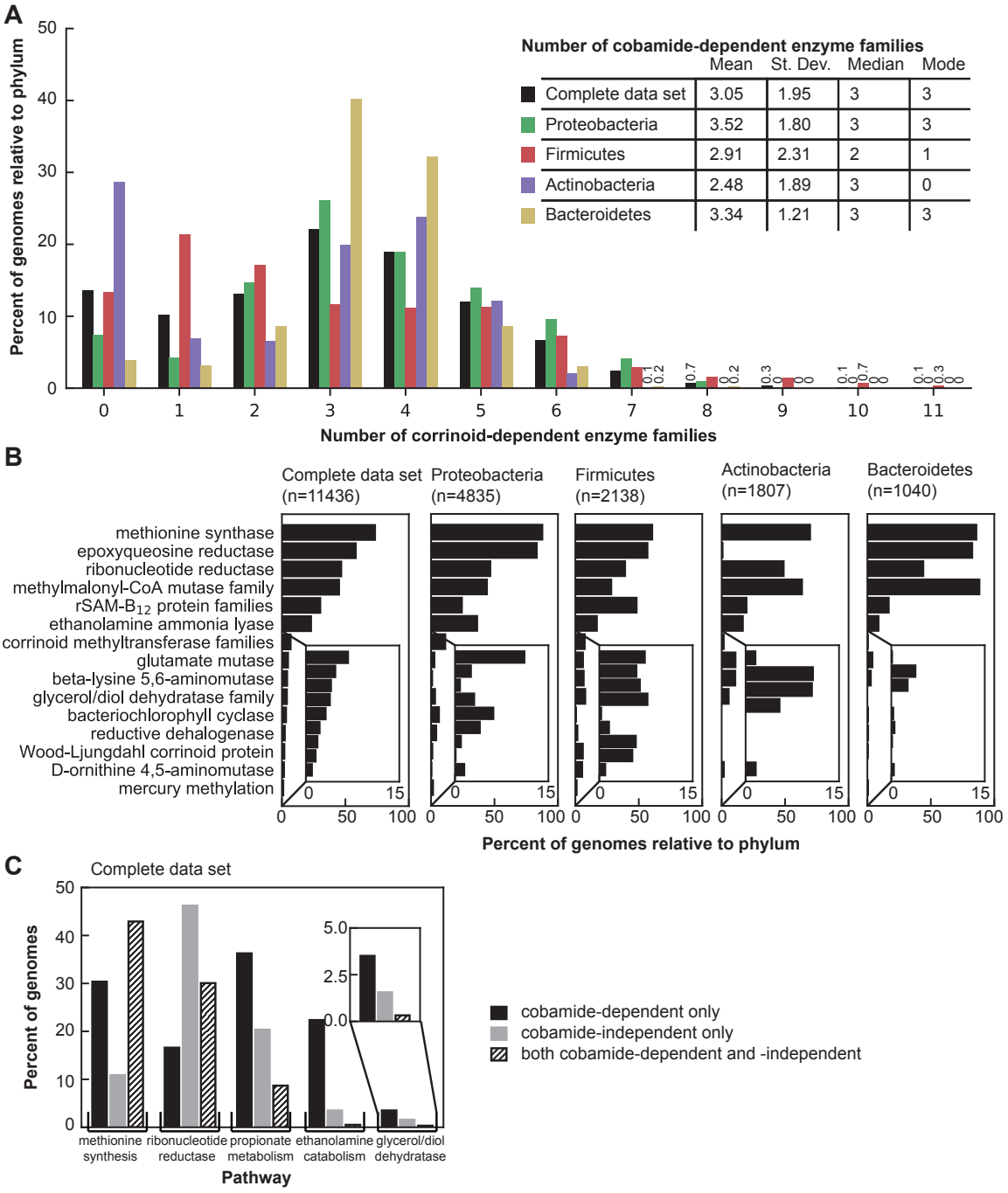



**Figure 3: Cobamide dependence in bacteria. A.** Histogram of the number of cobamide-dependent enzyme families (shown in Fig. 1, Table 1, 2) per genome in the complete filtered data set and the four most abundant phyla in the data set. The numbers are given for bars with values less than 1%. The

inset lists the mean, standard deviation, median, and mode of cobamide-dependent enzyme families for each phylum. **B.** Rank abundance of cobamide-dependent enzyme families in the filtered data set and the four most abundant phyla. The inset shows an expanded view of the nine less abundant functions. **C.** Abundance of five cobamide-dependent processes and cobamide-independent alternatives in the complete filtered data set. Genomes with only the cobamide-dependent, only the cobamide-independent, or both pathways are shown for each process.

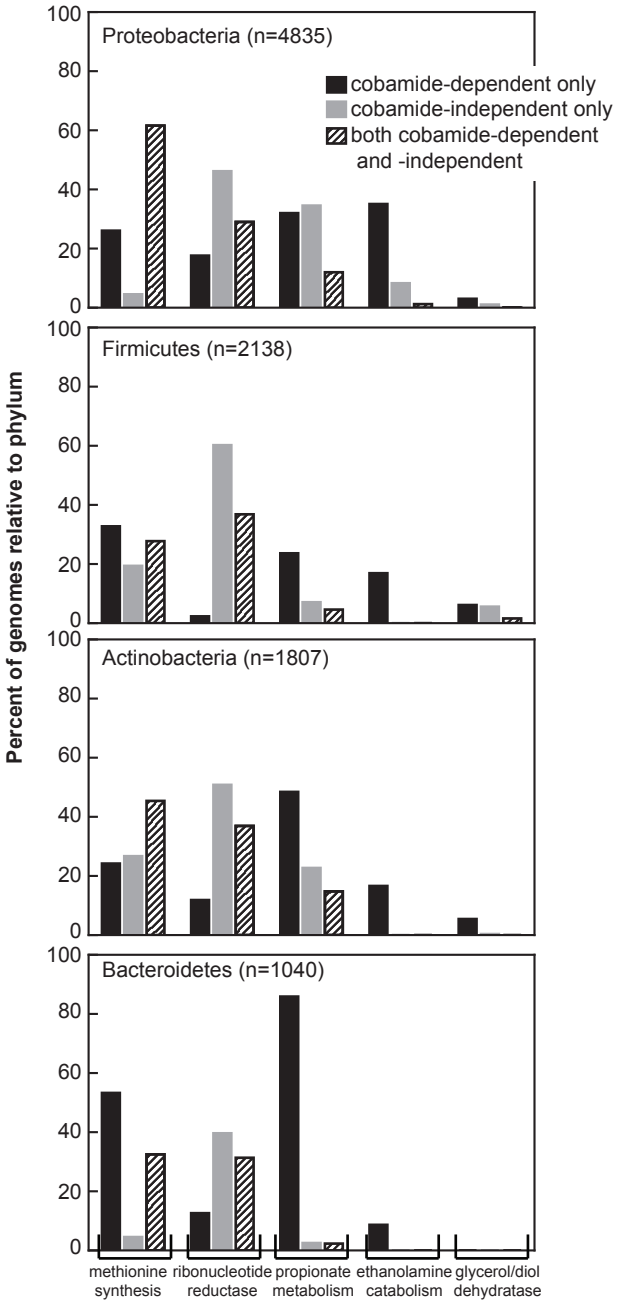



**Figure 4. Cobamide-dependent enzyme families with cobamide-independent alternatives by phylum.** The same analysis of Fig. 3C was separated by phylum.

**Thirty-seven percent of bacterial species are predicted to produce cobamides *de novo***

We analyzed the filtered data set to make informed predictions of cobamide biosynthesis to determine the extent of cobamide biosynthesis in bacteria and to identify marker genes predictive of cobamide biosynthesis. A search for genomes containing the complete pathways for anaerobic or aerobic cobamide biosynthesis, as defined in the model bacteria *Salmonella enterica* serovar Typhimurium and *Pseudomonas denitrificans*, respectively[46], revealed that few genomes contain all annotations for the complete pathway, but many contain nearly all. Some bacteria that appear to have an incomplete pathway might nonetheless be capable of cobamide biosynthesis because of poor annotation, non-homologous replacement of certain genes[50,100], or functional overlap of some of the enzymes. We therefore relied on experimental data on cobamide biosynthesis in diverse bacteria to inform our predictions, using 63 bacteria that are known to produce cobamides (Table 5, 6), including 5 tested in this study (Table 5, 6, bold names, Fig. 5). We identified a core set of eight annotations shared by all or all except one of the genomes of cobamide-producing bacteria (Table 6, gray highlight). These core annotations include three required for corrin ring biosynthesis: *cbiL*, *cbiF* and *cbiC* in the anaerobic pathway, which are orthologous to *cobI*, *cobM* and *cobH*, respectively, in the aerobic pathway (Table 6, Fig. 2A). An additional five nucleotide loop assembly annotations were also highly abundant in these genomes (Table 6).

**Table 5: References for experimentally confirmed cobamide producers examined in this study**

| Organism name | Pathway phenotype | Alternate genome Used in Analysis | Reference |
|---|---|---|---|
| *Acetobacterium woodii* DSM 1030 | anaerobic | | [151] |
| *Agrobacterium tumefaciens* C58 | aerobic | C58-UWASH | [152] |
| *Aphanizomenon flos-aquae* | unknown | NIES-81 | [153] |
| *Bacillus megaterium* | anaerobic | ATCC 14581/ATCC 10778 | [154] |
| ***Blautia hydrogenotrophica* DSM 10507** | anaerobic | | **this study** |
| *Clostridium cochlearium* DSM 1285 | anaerobic | NLAE-zl-C224 | [155] |
| ***Clostridium kluyveri* DSM 555** | anaerobic | | **this study** |
| ***Clostridium phytofermentans* ISDg** | anaerobic | | **this study** |
| *Clostridium tetanomorphum* | unknown | DSM 665 | [156] |
| *Crocosphaera watsonii* WH 8501 | unknown | | [157] |
| *Dehalobacter restrictus* CF | anaerobic | | [158] |
| *Desulfitobacterium hafniense* Y51 | anaerobic | | [69] |
| *Desulfitobacterium* sp. PCE1 (DSM 10344) | anaerobic | | [69] |
| *Desulfobacterium autotrophicum* HRM2 | anaerobic | | [159] |
| *Desulfobulbus propionicus* DSM 2032 | anaerobic | | [159] |
| ***Desulfotomaculum reducens* MI-1** | anaerobic | | **this study** |

| | | | |
|---|---|---|---|
| *Desulfovibrio desulfuricans* LS | anaerobic | ND132 | [160] |
| *Desulfovibrio vulgaris* DSM 644 (strain Hildenborough) | anaerobic | | [161] |
| *Dinoroseobacter shibae* DFL12T | aerobic | | [11] |
| *Eubacterium (Clostridium) barkeri* | anaerobic | VPI 5359 | [162] |
| *Eubacterium hallii* DSM 3353 | anaerobic | | [163,164] |
| *Eubacterium limosum* DSM 20517 | anaerobic | KIST612 | [165] |
| *Geobacter lovleyi* SZ | anaerobic | | [117] |
| *Geobacter sulfurreducens* PCA | anaerobic | | [68] |
| *Lactobacillus coryniformis* CRL 1001 | anaerobic | DSM 20001 | [166] |
| *Lactobacillus reuteri* DSM 20016 | anaerobic | | [167] |
| *Lactobacillus rossiae* DSM 15814 | anaerobic | | [168] |
| ***Listeria monocytogenes*** | anaerobic | v. 1/2a 10403S | **this study** |
| *Methylobacter luteus* | aerobic | IMV-B-3098 | [169] |
| *Methylobacterium dichloromethanicum (extorquens)* DM4 VKM B-2191T (DSM 6343) | aerobic | | [170] |
| *Methylobacterium extorquens* AM1 VKM B-2067 (NCIMB 9133) | aerobic | | [170] |
| *Methylococcus capsulatus* Bath | aerobic | | [169] |
| *Methylophilus methylotrophus* VKM B-162 (ATCC 53528) | aerobic | | [170] |
| *Methylosinus trichosporium* OB3b VKM B-2117T (ATCC 35070) | aerobic | | [170] |
| *Moorella thermoacetica* | anaerobic | ATCC 39073 | [171] |
| *Pelobacter propionicus* DSM 2379 | anaerobic | | [172] |
| *Pelosinus fermentans* R7 | anaerobic | | [108] |
| *Prochlorococcus sp.* MIT9313 | anaerobic | | [173] |
| *Propionibacterium acidipropionici* DSM 20273 | anaerobic | ATCC 4875 | [174] |
| *Propionibacterium freudenreichii* DSM4902 | anaerobic | DSM 20271 | [175] |
| *Propionibacterium freudenreichii* subspecies shermanii | anaerobic | JS | [47] |
| *Pseudomonas denitrificans* SC510 | aerobic | ATCC 13867 | [152] |
| *Pseudomonas putida* KT2440 | aerobic | | [152] |
| *Rhodobacter capsulatus* SB1003 | aerobic | | [176] |
| *Rhodobacter sphaeroides* 2.4.1 | aerobic | | [104] |
| *Rhodospirillum rubrum* | aerobic | ATCC 11170 | [70] |
| *Ruegeria pomeryoi* DSS-3 | aerobic | | [173] |
| *Salmonella enterica* serovar Typhimurium LT2 | anaerobic | | [177] |
| *Sinorhizobium meliloti* RM1021 | aerobic | | [67] |
| *Sporomusa ovata* DSM 2662 | anaerobic | | [151] |
| *Streptomyces coelicolor* A3 (2) | aerobic | | [178] |

| | | | |
|---|---|---|---|
| *Streptomyces griseus* ATCC 11009 | aerobic | DSM 40236 | [179] |
| *Sulfuorospirillum multivorans* DSM 12446 | anaerobic | | [26] |
| *Synechococcus elongatus* PCC7942 | anaerobic | | [11] |
| *Synechococcus* sp. CC9311 | anaerobic | | [11] |
| *Synechococcus* sp. WH5701 | anaerobic | | [11] |
| *Synechococcus* sp. WH7803 | anaerobic | | [173] |
| *Synechococcus* sp. WH7805 | anaerobic | | [11] |
| *Synechococcus* sp. WH8102 | anaerobic | | [173] |
| *Synechocystis* sp. PCC6803 | anaerobic | | [11] |
| *Thermosipho africanus* H1760334 | anaerobic | | [98] |
| *Thermosipho africanus* TCF52B | anaerobic | | [98] |
| *Veillonella parvula* DSM 2008 | anaerobic | | [74] |
| *Yersinia enterocolitica* 8081 | anaerobic | | [180] |

**Table 6. Experimentally-verified cobamide producers and their cobamide biosynthesis annotation content.**

| Biosynthesis phenotype | Phylum | Class | Order | Family | Organism name | Tetrapyrrole precursor biosynthesis: ALA synthesis (HemA or HemAL) | Tetrapyrrole precursor biosynthesis: EC:4.2.1.24 (HemB) | Tetrapyrrole precursor biosynthesis: EC:2.5.1.61 (HemC) | Tetrapyrrole precursor biosynthesis: EC:4.2.1.75 (HemD) | Tetrapyrrole precursor biosynthesis: EC:2.1.1.107 (CysG/CobA) | Corrin ring biosynthesis: Anaerobic: EC:1.3.1.76 (CysG) | Corrin ring biosynthesis: Anaerobic: EC:4.99.1.3 (CbiK/CbiX) | Corrin ring biosynthesis: Anaerobic: **EC:2.1.1.151 (CbiL)** | Corrin ring biosynthesis: Anaerobic: EC:2.1.1.131 (CbiH) | Corrin ring biosynthesis: Anaerobic: **EC:2.1.1.271 (CbiF)** | Corrin ring biosynthesis: Anaerobic: EC:3.7.1.12 (CbiG) | Corrin ring biosynthesis: Anaerobic: EC:2.1.1.195 (CbiD) | Corrin ring biosynthesis: Anaerobic: EC:1.3.1.106 (CbiJ) | Corrin ring biosynthesis: Anaerobic: EC:2.1.1.196 (CbiT) | Corrin ring biosynthesis: Anaerobic: EC:2.1.1.289 (CbiE) | Corrin ring biosynthesis: Anaerobic: **EC:5.4.99.60 (CbiC)** | Corrin ring biosynthesis: Anaerobic: EC:6.3.5.11 (CbiA) | Corrin ring biosynthesis: Aerobic: **EC:2.1.1.130 (CobI)** | Corrin ring biosynthesis: Aerobic: EC:1.14.13.83 (CobG) | Corrin ring biosynthesis: Aerobic: EC:2.1.1.131 (CobJ) | Corrin ring biosynthesis: Aerobic: **EC:2.1.1.133 (CobM)** | Corrin ring biosynthesis: Aerobic: EC:2.1.1.152 (CobF) | Corrin ring biosynthesis: Aerobic: EC:1.3.1.54 (CobK) | Corrin ring biosynthesis: Aerobic: EC:2.1.1.132 (CobL) | Corrin ring biosynthesis: Aerobic: **EC:5.4.99.61 (CobH)** | Corrin ring biosynthesis: Aerobic: EC:6.3.5.9 (CobB) | Corrin ring biosynthesis: Aerobic: EC:6.6.1.2 (CobNST) | Corrin ring biosynthesis: Aerobic: EC:1.16.8.1 (CobR) | Adenosylation: EC:2.5.1.17 (CobA/BtuR/CobO/PduO) | Aminopropanol linker: EC:2.7.1.177 (PduX) | Aminopropanol linker: EC:4.1.1.81 (CobD) | Nucleotide loop assembly: **EC:6.3.5.10 (CbiP/CobQ)** | Nucleotide loop assembly: **EC:6.3.1.10 (CbiB/CobC/CobD)** | Nucleotide loop assembly: **EC:2.7.1.156 (CobU/CobP)** | Nucleotide loop assembly: **cobinamide activation (CobU/CobP/CobY)** | Nucleotide loop assembly: EC:2.4.2.21 (CobT/CobU/ArsAB) | Nucleotide loop assembly: cbl phosphatase (CobS/CobV) | Nucleotide loop assembly: **EC:2.7.8.26 (CobC/CobZ)** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Total genomes with step | | 63 | 63 | 60 | 63 | 63 | 48 | 43 | 62 | 58 | 63 | 58 | 53 | 54 | 23 | 18 | 63 | 61 | 63 | 9 | 58 | 63 | 12 | 54 | 48 | 63 | 61 | 36 | 1 | 59 | 17 | 29 | 62 | 63 | 63 | 63 | 51 | 33 | 63 |
| Anaerobic | Actinobacteria | | | | *Propionibacterium acidipropionici* | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Propionibacterium freudenreichii* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Propionibacterium* shermanii | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | Cyanobacteria | | | | *Prochlorococcus sp.* MIT9313 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus elongatus* | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus* sp. CC9311 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus* sp. WH5701 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus* sp. WH7803 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus* sp. WH7805 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechococcus* sp. WH8102 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Synechocystis* sp. PCC6803 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | Bacilli | Bacillales | | *Bacillus megaterium* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | ***Listeria monocytogenes*** | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | Lactobacillales | | *Lactobacillus coryniformis* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Lactobacillus reuteri* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Lactobacillus rossiae* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | Clostridia | Clostridiales | Clostridiaceae | *Clostridium cochlearium* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | ***Clostridium kluyveri*** | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | Eubacteriaceae | *Acetobacterium woodii* | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Eubacterium barkeri* | ■ | ■ | | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Eubacterium hallii* | ■ | ■ | | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Eubacterium limosum* | ■ | ■ | | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | Lachnospiraceae | ***Blautia hydrogenotrophica*** | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | ***Clostridium phytofermentans*** | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | Peptococcaceae | *Dehalobacter restrictus* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Desulfitobacterium hafniense* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Desulfitobacterium* sp. PCE1 | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | ***Desulfotomaculum reducens*** | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | Thermoanaerobacterale | | *Moorella thermoacetica* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | Negativicutes | | | *Pelosinus fermentans* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Sporomusa ovata* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Veillonella parvula* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | Proteobacteria | δ | Desulfobacterales | | *Desulfobacterium autotrophicum* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Desulfobulbus propionicus* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | Desulfovibrionales | | *Desulfovibrio desulfuricans* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Desulfovibrio vulgaris* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | Desulfuromonadales | | *Geobacter lovleyi* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Geobacter sulfurreducens* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Pelobacter propionicus* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | ε | | | *Sulfurospirillum multivorans* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | γ | | | *Salmonella typhimurium* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Yersinia enterocolitica* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | Thermotogae | | | | *Thermosipho africanus* H1760334 | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Thermosipho africanus* TCF52B | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| Aerobic | Actinobacteria | | | | *Streptomyces coelicolor* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Streptomyces griseus* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | Proteobacteria | α | Rhizobiales | | *Agrobacterium tumefaciens* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Methylobacterium dichloromethanicum*[a] | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Methylobacterium extorquens*[a] | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Methylosinus trichosporium*[a] | ■ | ■ | ■ | ■ | ■ | | | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Sinorhizobium meliloti* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Dinoroseobacter shibae* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Rhodobacter capsulatus* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Rhodobacter sphaeroides* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | Rhodobacterales | | *Ruegeria pomeryoi* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | Rhodospirillales | | *Rhodospirillum rubrum* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | γ | Methylococcales | | *Methylobacter luteus* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | | | *Methylococcus capsulatus* | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | ■ | ■ |
| | | | Pseudomonadales | | *Pseudomonas denitrificans* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Pseudomonas putida* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| Unk. | Cyanobacteria | | | | *Aphanizomenon flos-aquae* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | | | | | *Crocosphaera watsonii* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |
| | Firmicutes | | | | *Clostridium tetanomorphum* | ■ | ■ | ■ | ■ | ■ | | ■ | ■ | ■ | | ■ | | ■ | | | ■ | ■ | ■ | | ■ | ■ | | | | | | | | | | | ■ | ■ | ■ | ■ | | | ■ |

[a]*Methylophilus methylotrophus* was also reported to be an aerobic corrinoid producer, but its genome only has 1 corrin ring biosynthesis annotation (CobH). The reported concentration of corrinoid it produced is 6-fold less than other strains in the study by Ivanova et al. (2006). We do not think this strain can actually produce corrinoids.

Bold species names were identified as cobamide producers in this study.

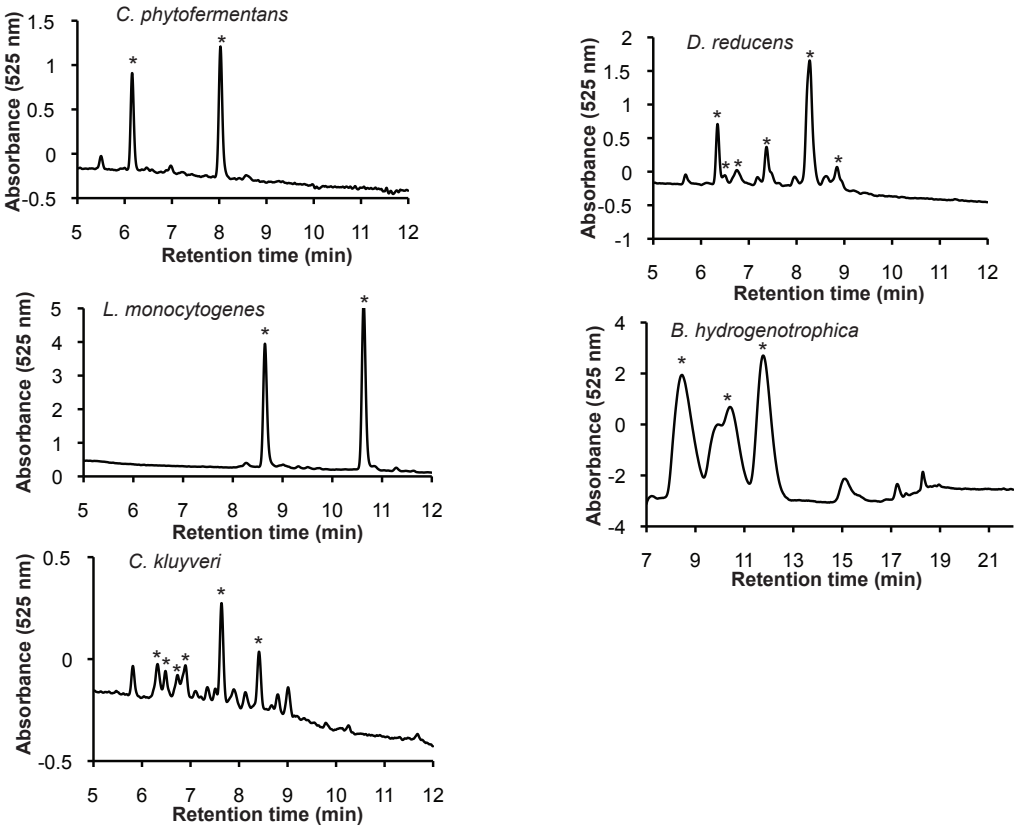



**Figure 5. Corrinoid extractions from the organisms tested in Table 5 and Table 6.** Corrinoids separated on HPLC as described in the methods. Asterisks indicate peaks consistent with a corrinoid based on the UV-Vis spectrum.

Our analysis additionally showed that the anaerobic and aerobic corrin ring biosynthesis pathways cannot be distinguished based on their annotated gene content, presumably because portions of the two pathways share orthologous genes (Table 6; Fig. 2A, dashed lines). Even the cobalt chelatases, *cobNST* and *cbiX/cbiK*, are not exclusive to genomes with the aerobic or anaerobic pathways, respectively (Table 6). Cobalt chelatase annotations are also found in some bacteria that lack most of the corrin ring and nucleotide loop assembly genes, suggesting that there is overlap in annotations with other metal chelatases[181].

We next sought to predict cobamide biosynthesis capability across bacteria by analyzing the filtered genome data set by defining different levels of confidence for predicting cobamide biosynthesis (Table 3, 4). Annotations that are absent from the majority of genomes of experimentally verified cobamide producers (*cobR*, *pduX*, and *cobD*) (Table 6, Fig. 2A), as well as one whose role in cobamide biosynthesis has not been determined (*cobW*)[182] were excluded from these threshold-based definitions. We did not exclusively use the small set of core annotations identified in Table 6 because a correlation between the absence of these genes and lack of cobamide biosynthesis ability has not been established. Using these threshold-based definitions, we predict that 37% of bacteria in the data set have the potential to produce cobamides (Fig. 6, black bars). Forty-nine percent of species in the data set have at least one

cobamide-dependent enzyme but lack a complete cobamide biosynthetic pathway. Genomes in the latter category can be further divided into non-producers, which contain fewer than five corrin ring biosynthesis genes, and precursor salvagers, which contain distinct portions of the pathway (described in a later section). The distribution of cobamide-dependent enzyme families also varies based on predicted cobamide biosynthesis, with predicted cobamide producers having more cobamide-dependent enzyme families per genome than non-producers (Fig. 7).

To assess whether the core corrin ring annotations (Table 6, gray highlight) identified in the experimentally verified cobamide producers could be used as markers, the threshold-based assignments of cobamide biosynthesis categories were compared to the frequency of the three annotations. The presence of each core annotation alone is largely consistent with the threshold-based category assignments, as each is present in 99% of genomes in the producer categories and in less than 1% of the non-producers (Table 7). The presence of two or all three marker annotations matches the threshold-based predictions even more closely (Table 7). The corrin ring markers chosen in Table 7 are slightly more predictive of our threshold-based cobamide biosynthesis classifications than *cbiA/cobB* (EC:6.3.5.11/EC:6.3.5.9), a previously selected marker used in environmental DNA analysis[183]; although *cbiA/cobB* was found in 99% of predicted cobamide producers, is it also present in 2.6% of predicted non-producers and 46% of precursor salvagers.

**Table 7. Presence of corrin ring marker annotations in predicted cobamide biosynthesis categories**

| | Cobamide biosynthesis category | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Cobamide producers | | | | Partial biosynthesis | | Non-producers | | |
| | Very likely n=1016 | Likely n=2361 | Possible n=832 | All cobamide producers n=4209 | Tetrapyrrole precursor salvager n=201 | Cbi salvage n=1734 | Likely n=29 | Very likely n=5263 | All non-producers n=5292 |
| CbiL/CobI | 100.0* | 99.2 | 98.0 | 99.2 | 87.1 | 8.9 | 62.1 | 0.1 | 0.5 |
| CbiF/CobM | 100.0 | 99.8 | 96.8 | 99.2 | 93.5 | 4.7 | 65.5 | 0.4 | 0.6 |
| CbiC/CobH | 100.0 | 99.7 | 96.4 | 99.1 | 99.5 | 5.0 | 69.0 | 0.5 | 0.9 |
| CbiL/CobI and CbiF/CobM | 100.0 | 99.0 | 94.9 | 98.4 | 80.6 | 1.3 | 37.9 | 0.0 | 0.2 |
| CbiL/CobI and CbiC/CobH | 100.0 | 98.9 | 94.5 | 98.3 | 89.6 | 4.5 | 41.4 | 0.0 | 0.2 |
| CbiF/CobM and CbiC/CobH | 100.0 | 99.4 | 93.4 | 98.4 | 89.6 | 8.2 | 41.4 | 0.0 | 0.3 |
| CbiL/CobI and CbiF/CobM and CbiC/CobH | 100.0 | 98.6 | 91.6 | 97.6 | 76.6 | 0.5 | 24.1 | 0.0 | 0.1 |

*Numbers represent the percent of genomes containing each marker annotation and combinations of annotations within each cobamide biosynthesis category.

As with the cobamide-dependent enzyme families, the four major phyla in the data set have major differences in their predicted cobamide biosynthesis phenotypes (Fig. 6). Around half of Actinobacteria (57%) and Proteobacteria (45%) and 30% of Firmicutes are predicted to be cobamide producers. In contrast, only 0.6% of Bacteroidetes are predicted to produce cobamides *de novo*, yet 96% have at least one cobamide-dependent enzyme, suggesting that most members of this phylum must acquire cobamides from other organisms in their environment. In addition, Bacteroidetes have the highest relative proportion of species predicted to salvage Cbi via a partial cobamide biosynthesis pathway, and most of the tetrapyrrole precursor salvagers are Firmicutes (see later section), whereas very few Actinobacterial species are predicted to salvage precursors (Fig. 6). These divisions reveal potential cobamide and cobamide precursor requirements across phyla.

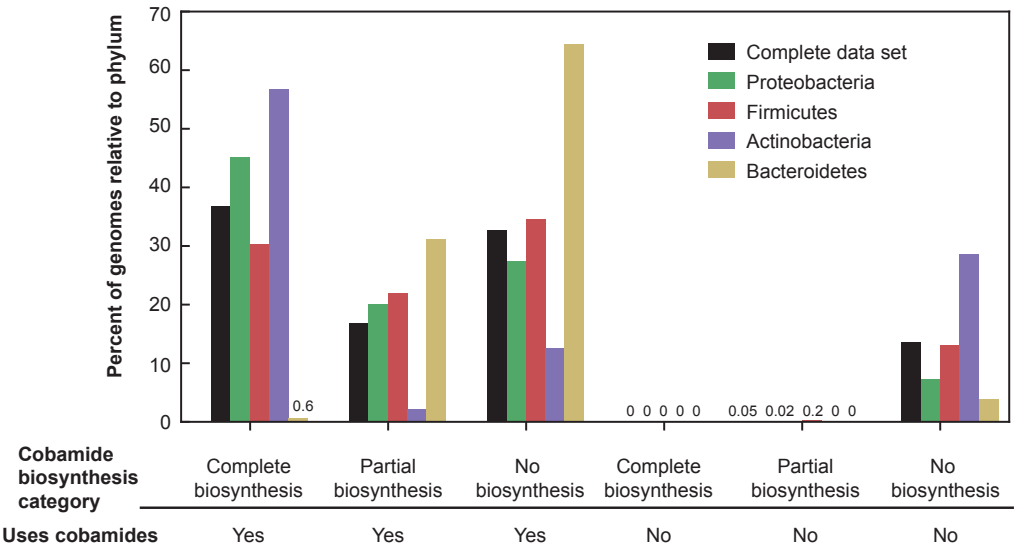



**Figure 6: Predicted cobamide biosynthesis phenotypes in the complete filtered data set and the four most abundant phyla in the data set.** Genomes were classified into predicted corrinoid biosynthesis phenotypes based on the criteria listed in Table 3 and Table 4. The "Partial biosynthesis" category includes cobinamide salvagers and tetrapyrrole precursor auxotrophs. The "Uses cobamides" category is defined as having one or more of the cobamide-dependent enzyme families shown in Figure 1. The numbers are given for bars that are not visible.

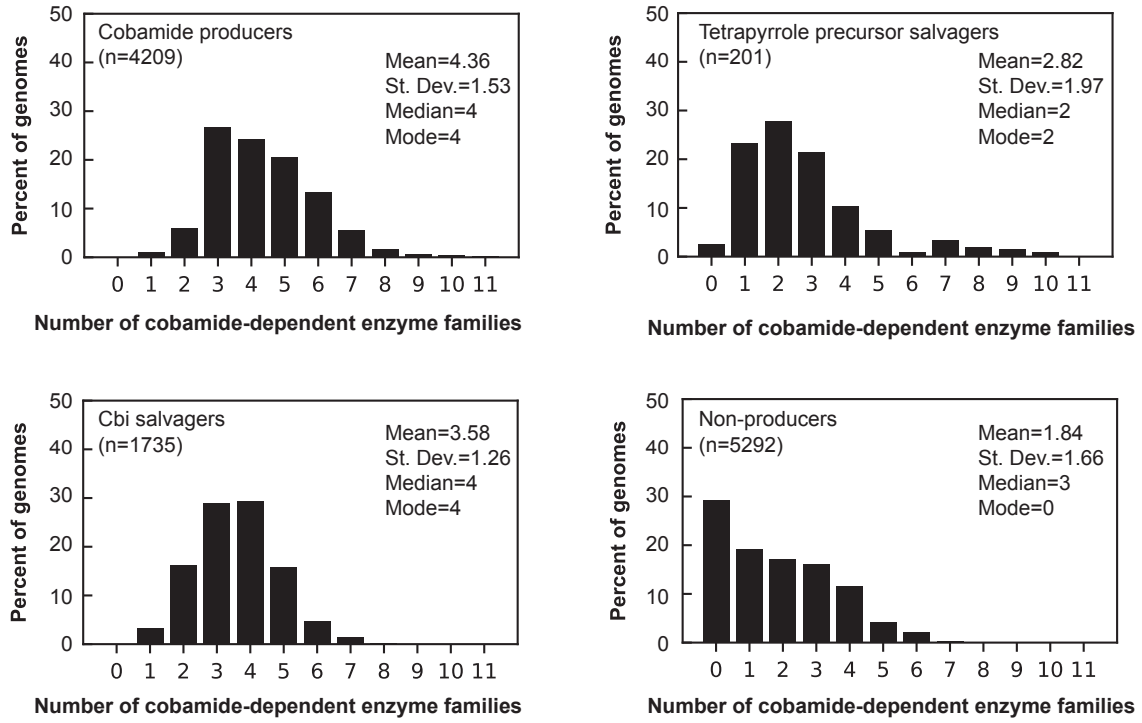



**Figure 7. Number of cobamide-dependent enzyme families per genome in four cobamide**

**biosynthesis classification categories.** The data in Fig. 3A is separated by cobamide biosynthesis category (Fig 6, Table 4).

## Predicting cobamide structure

Lower ligand structure is determined by the intracellular production of lower ligand bases as well as specific features of the lower ligand attachment genes *cobT* or *arsAB*[64,66–70,75,184]. We first defined predictions for the biosynthesis of the class of cobamides containing benzimidazole lower ligands (benzimidazolyl cobamides), based on the presence of genes for the biosynthesis of benzimidazoles. We used the presence of *bluB*, the aerobic synthase for the lower ligand of cobalamin, 5,6-dimethylbenzimidazole (DMB), as a marker for cobalamin production[66,67,71] and found it in 25% of genomes in the data set, including those without complete cobamide biosynthesis pathways. *bluB* is most abundant in predicted cobamide-producing bacteria (Fig. 8A), particularly in Proteobacteria (Fig. 8B).

Anaerobic biosynthesis of DMB and three other benzimidazoles requires different combinations of the *bza* genes as shown in Figures 2A and 5C[68,72]. Because annotations for the majority of the *bza* genes were not available, we developed profile HMMs to search for them. Ninety-six genomes contain one or more *bza* genes, and 88 of these contain either *bzaF* or both *bzaA* and *bzaB*, the first step necessary for the anaerobic biosynthesis of all four benzimidazoles (Fig. 8C). As seen with *bluB*, anaerobic benzimidazole biosynthesis genes are highly enriched in cobamide producers (Fig. 8A). Examining the set of *bza* genes in each genome allowed us to predict the structures of cobamides produced in 86 out of the 96 genomes (Fig. 8C). Based on the frequency of *bluB* and the *bza* genes, 24% of bacteria are predicted to produce cobalamin, the cobamide required by humans.

To predict the biosynthesis of phenolyl cobamides, we searched for genomes containing two adjacent *cobT* annotations, since the *cobT* homologs *arsA* and *arsB*, which together are necessary for activation of phenolic compounds for incorporation into a cobamide, are encoded in tandem[64]. Using this definition, *arsAB* was found in only 27 species, and is almost entirely restricted to the class Negativicutes in the phylum Firmicutes, which are the only bacteria reported to produce phenolyl cobamides[73,185] (Fig. 8A, B).

Forty-two percent of predicted cobamide producers in the data set do not have any of the benzimidazole biosynthesis or phenolic attachment genes (Fig. 8A). However, bacteria that have the α-ribazole kinase CblS (Fig. 8A, B, inner rings) and the transporter CblT (not included) are predicted to use activated forms of lower ligand bases found in the environment (Fig. 2A, α-ribazole salvaging); we found CblS in 363 species (3.2%), mostly in the Firmicutes phylum (Fig. 8A, B, inner rings)[99,100]. A higher proportion of bacteria, 1,041 species (9.1%), have a CbiZ annotation (Fig. 8A, B, outer rings), an amidohydrolase that cleaves the nucleotide loop, allowing cells to rebuild a cobamide with a different lower ligand[105] (Fig. 2A, corrinoid remodeling). CbiZ is found in genomes of predicted cobamide producers and Cbi auxotrophs (see following section) (Fig. 8A), as expected based on experimental studies[104,106–108]. The reliance of some bacteria on exogenous lower ligands or α-ribazoles produced by other organisms precludes prediction of cobamide structure in all cases.

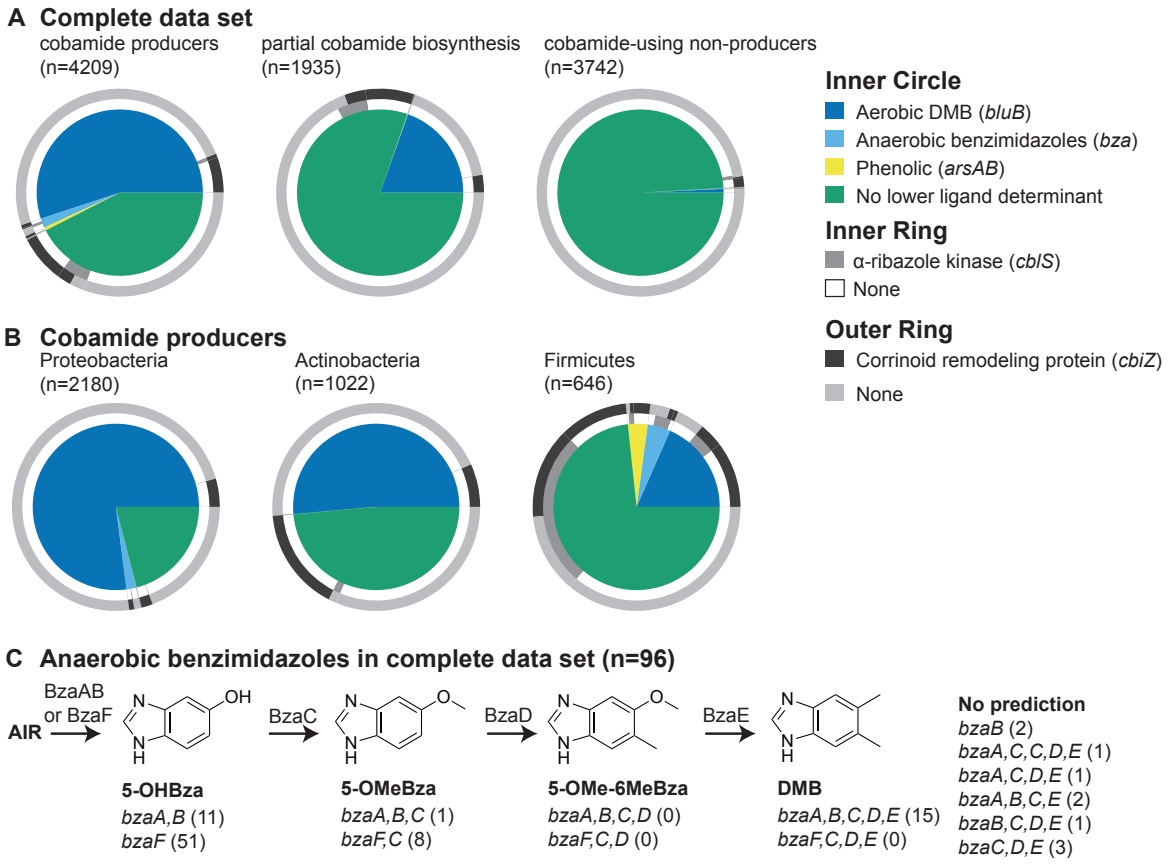



**Figure 8: Lower ligand structure predictions. A, B.** Proportion of genomes containing the indicated lower ligand structure determinants (inner circle), α-ribazole salvaging gene (inner ring), and corrinoid remodeling gene (outer ring) in the complete filtered data set separated by cobamide producer category (A) and in cobamide producers separated by phylum (B). **C.** The anaerobic benzimidazole biosynthesis pathway is shown with the functions that catalyze each step above the arrows. The genes required to produce each benzimidazole are shown below each structure, with the number of genomes in the complete filtered data set containing each combination of genes in parentheses. The sets of *bza* genes that do not have a predicted structure are listed on the right. Abbreviations: AIR, aminoimidazole ribotide; 5-OHBza, 5-hydroxybenzimidazole; 5-OMeBza, 5-methoxybenzimidazole; 5-OMe-6-MeBza, 5-methoxy-6-methylbenzimidazole.

## Seventeen percent of bacteria have partial cobamide biosynthetic pathways

Our analysis of the cobamide biosynthesis pathway revealed two categories of genomes that lack some or most genes in the pathway, but retain contiguous portions of the pathway. Genomes in one category, the Cbi (cobinamide) -salvaging bacteria (15% of genomes), contain the nucleotide loop assembly steps but lack all or most of the corrin ring biosynthesis annotations (Fig. 9A). As demonstrated in *Escherichia coli*[79], *Thermotoga lettingae*[97], and *Dehalococcoides mccartyi*[107], and predicted in human gut microbes[9], Cbi salvagers can take up the late intermediate Cbi, assemble the nucleotide loop and attach a lower ligand.

We observed an additional 201 genomes (1.7%) that lack one or more initial steps in tetrapyrrole precursor biosynthesis but have complete corrin ring biosynthesis and nucleotide loop assembly pathways, primarily in the Firmicutes (Table 3, 4). After searching these

genomes manually for genes missing from the pathway, we designated 180 of these species as tetrapyrrole precursor salvagers, a new classification of cobamide intermediate auxotrophs (Fig. 9A). These organisms are predicted to produce cobamides only when provided with a tetrapyrrole precursor or a later intermediate in the pathway.

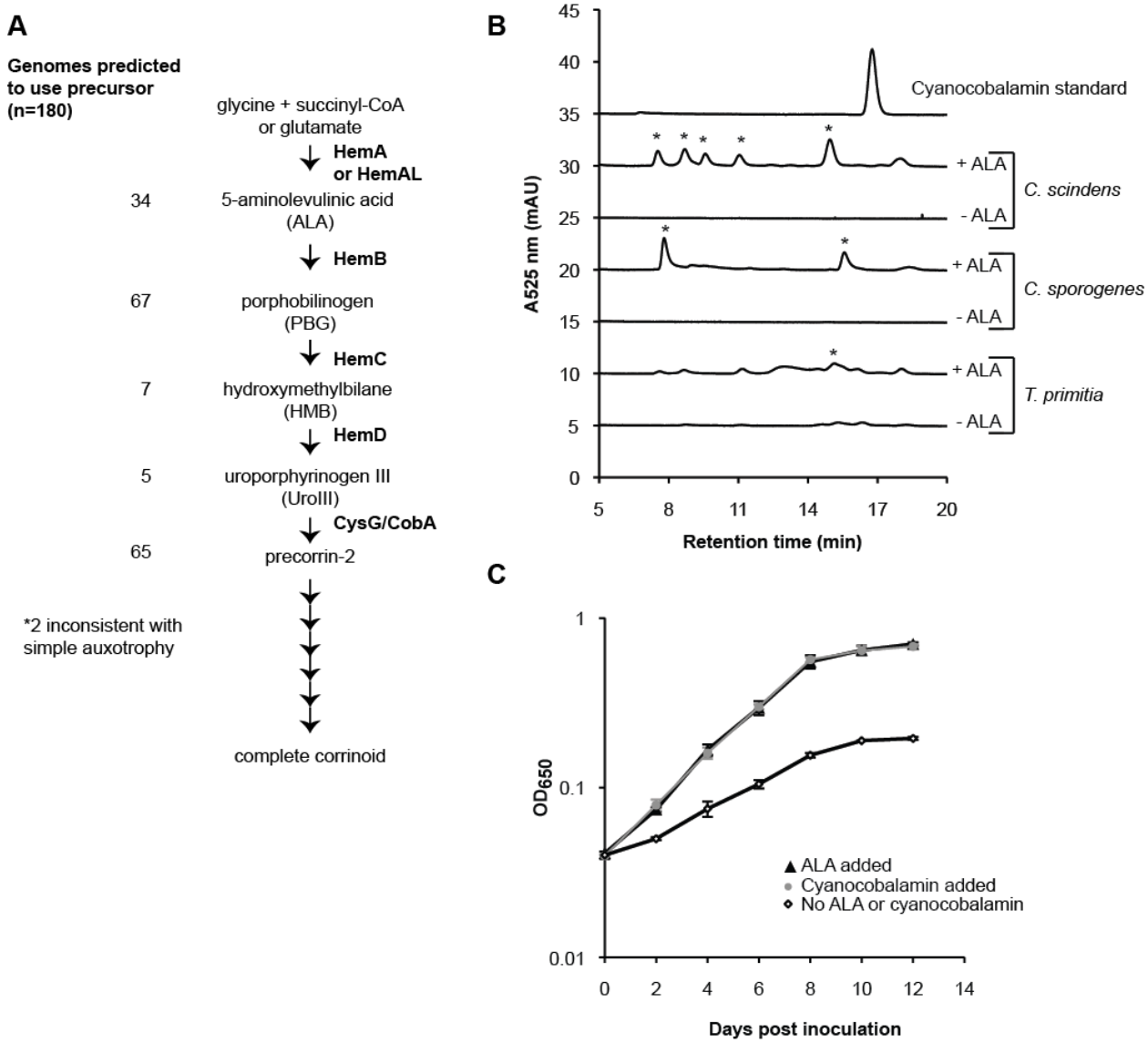



**Figure 9: Characterization of putative tetrapyrrole precursor salvagers A.** Steps in cobamide biosynthesis. The functions that catalyze each step are indicated to the right of each arrow. The number of genomes in the complete filtered data set in each precursor salvage category is on the left. Two genomes had cobamide biosynthesis pathways inconsistent with simple auxotrophy.. **B.** HPLC analysis of corrinoid extracts from *Clostridium scindens*, *Clostridium sporogenes*, and *Treponema primitia* grown with and without added ALA. A cyanocobalamin standard (10 μM) is shown for comparison. Asterisks denote peaks with UV-Vis spectra consistent with that of a corrinoid. **C.** *T. primitia* ZAS-2 growth in 4YACo medium with and without added cyanocobalamin (37 nM) or ALA (1 mM). Each point represents the average of three biological replicates. Error bars are the standard deviation.

**Experimental validation of 5-aminolevulinic acid (ALA) dependence**

The identification of putative tetrapyrrole precursor salvagers suggests either that these bacteria are capable of taking up a tetrapyrrole precursor from the environment to produce a cobamide or that they synthesize the precursors through a novel pathway. We therefore tested three putative tetrapyrrole precursor salvagers for their ability to produce corrinoids (cobamides and other corrin ring-containing compounds) in the presence and absence of a tetrapyrrole precursor. *Clostridium scindens* and *Clostridium sporogenes*, which are predicted to require 5-aminolevulinic acid (ALA), produced corrinoids in defined media only when ALA was supplied, suggesting that they do not have a novel ALA biosynthesis pathway (Fig. 9B). We tested an additional predicted ALA salvager, the termite gut bacterium *Treponema primitia* ZAS-2, for which a defined medium has not been developed. When cultured in medium containing yeast autolysate, *T. primitia* produced trace amounts of corrinoids, and corrinoid production was increased by supplementing this medium with ALA (Fig. 9B). The ability of *T. primitia* to use externally supplied ALA was further shown by its increased growth rate and cell density at stationary phase when ALA was added (Fig. 9C). Together, these results support the hypothesis that predicted ALA salvagers synthesize cobamides by taking up ALA from the environment.

**Discussion**

Vitamin $B_{12}$ and other cobamides have long been appreciated as a required nutrient for humans, bacteria, and other organisms due to their critical function as enzyme cofactors. The availability of tens of thousands of genome sequences afforded us the opportunity to conduct a comprehensive analysis of cobamide metabolism across over 11,000 bacterial species. This analysis gives an overview of cobamide dependence and cobamide biosynthesis across bacteria, allowing for the generation of hypotheses for cobamide and cobamide precursor interactions in bacterial communities. Our work shows that cobamide use is much more widespread than cobamide biosynthesis, consistent with the majority of previous studies of smaller data sets [7–9]. The prevalence of cobamide-dependent enzymes in bacteria, coupled with the relative paucity of *de novo* cobamide producers, underscores the ubiquity of both cobamide-dependent metabolism and cobamide salvaging in microbial communities. Here, we additionally find that cobamide production and use are unevenly distributed across the major phyla represented in the data set, identify bacteria dependent on cobamide precursors, and predict cobamide structure. These results underscore the widespread nutritional dependence of bacteria.

The most abundant types of cobamide-dependent enzymes in our data set are methionine synthase, epoxyqueuosine reductase, RNR, and MCM. For all of these enzymes, cobamide-independent alternative enzymes or pathways exist. (Note that the newly discovered alternative to epoxyqueuosine reductase, QueH[41], was not included in our analysis.) The prevalence of cobamide-dependent enzymes for which cobamide-independent counterparts exist, particularly in the same genome, suggests that cobamide-dependent enzymes confer distinct advantages. This view is supported by the observations that MetE is sensitive to stress and has a 100-fold lower turnover number than MetH[36,37,186] and that cobamide-independent RNRs are limited in the oxygen concentrations in which they are active[33,39].

In our analysis of cobamide biosynthesis, it was not possible to use a single definition of the complete *de novo* cobamide biosynthesis pathway across all bacterial genomes because of divergence in sequence and function. Similarly, while Archaea are known to produce and use cobamides, the archaeal cobamide biosynthetic pathway differs in key steps from the bacterial pathways, making annotation-based assignment of biosynthesis predictions difficult without further experimental characterization of non-homologous replacements[48]. The use of experimental data gives confidence to our predictions and allowed identification of marker genes for cobamide biosynthesis. Nevertheless, our predictions likely overestimate the extent of cobamide biosynthesis *in situ*, as genome predictions do not account for differences in gene expression. For example, cobamide production in *S. typhimurium* is repressed in environments containing oxygen or lacking propanediol[23], and cobamide biosynthesis operons are commonly subjected to negative regulation by riboswitches[7,86]. The abundance of cobamide importers[7–9], even in bacteria capable of cobamide biosynthesis, reinforces the possibility that many bacteria may repress expression of cobamide biosynthesis genes in favor of cobamide uptake in some environments.

A comparison of genomes containing one or more cobamide-dependent annotations to those with none revealed an absence of bacteria that produce cobamides but do not use them. This finding suggests that altruistic bacteria that produce cobamides exclusively for others do not exist. Metabolically coupled organisms that crossfeed cobalamin in exchange for another nutrient have been described in the mutualistic relationships between algae and cobalamin-producing bacteria[111,114], yet it remains unclear if such intimate partnerships are widespread. Notably, our results show that cobamide biosynthesis is unevenly distributed across bacteria, with Actinobacteria enriched in and Bacteroidetes lacking in *de novo* cobamide biosynthesis. Such phylogenetic comparisons can be used to make crude predictions of cobamide-based nutritional interactions among different taxa.

The reliance of many bacteria on environmental cobamides, coupled with the fact that structurally different cobamides are not functionally equivalent in bacteria[9,11,20,21,93,107,117], underscores the importance of cobamide lower ligand structure in microbial interactions. Additional variation in the nucleotide loop was not considered here because of the absence of genes specific to norcobamide biosynthesis[26,61]. We were able to predict lower ligand structure for 58% of predicted cobamide producers. The remaining bacteria may produce purinyl cobamides, the class of cobamides containing purine bases as lower ligands, which are abundant in some bacterial taxa and microbial communities[11,69,187]. Further analysis of substrate specificity in CobT and other lower ligand attachment enzymes could lead to improved strategies for predicting production of cobamides with purinyl lower ligands, as some CobT homologs appear to segregate into different clades based on lower ligand structure[69,74,75]. The presence of free benzimidazoles and α-ribazoles in microbial communities[95,101,102] and the ability of bacteria to take up and incorporate these compounds into cobamides[19,74,93,94] suggest that it will not be possible to predict the structures of cobamides produced by all bacteria *in situ* solely from genomic analysis.

We predict that 32% of bacteria that have cobamide-dependent enzymes are unable to synthesize cobamides, attach a preferred lower ligand to Cbi, or remodel corrinoids. This group of bacteria must take up cobamides from their environment for use in their cobamide-dependent metabolisms. Given the variable use of structurally different cobamides by different bacteria, the availability of specific cobamides is likely critical to bacteria that are unable to synthesize cobamides or alter their structure. The availability of preferred cobamides may limit the range

of environments that these organisms can inhabit. Variation in the abundance of different cobamides has been observed in different environments. For example, in a TCE-contaminated groundwater enrichment culture, 5-hydroxybenzimidazolyl cobamide and *p*-cresolyl cobamide were the most abundant cobamides[185], compared to cobalamin in bovine rumen[188] and 2-methyladeninyl cobamide in human stool[187]. One strategy for acquiring preferred cobamides could be selective cobamide import, as suggested by the ability of two cobamide transporters in *Bacteroides thetaiotaomicron* to distinguish between different cobamides[9].

Dependence on biosynthetic precursors has been observed or predicted for amino acids, nucleotides, and the cofactors thiamin and folate[189–192]. Here, we describe genomic evidence for dependence on cobamide precursors, namely Cbi or tetrapyrrole precursors. The prevalence of Cbi-salvaging bacteria (Fig. 9A) suggests that it is common for bacteria to fulfill their cobamide requirements by importing Cbi from the environment and assembling the nucleotide loop intracellularly. Consistent with this, Cbi represented up to 9% of total corrinoids in TCE-contaminated groundwater enrichments[185], and represented up to 12.8% of the total corrinoids detected in human stool samples[187].

Our analysis defined five types of tetrapyrrole precursor salvagers and experimentally verified the ALA salvager phenotype in three species. It was observed previously that *Porphyromonas gingivalis* lacks the steps to synthesize precorrin-2[193]. However little additional work has explored tetrapyrrole precursor salvagers. This biosynthesis category was overlooked in previous genomic studies of cobamide biosynthesis because these studies considered only the corrin ring biosynthesis and nucleotide loop assembly portions of the pathway[7–11]. Tetrapyrrole precursors have been detected in biological samples, suggesting that they are available for uptake in some environments. For example, uroporphyrin III, a derivative of the tetrapyrrole precursor uroporphyrinogen III (UroIII), was detected in human stool[194,195] and ALA has been found in swine manure extract[196]. Although we confirmed experimentally the ALA dependence phenotype, we were unable to detect ALA in several biological samples using a standard chemical assay via a fluorometric derivatization[197] or bioassay with *Rhodobacter sphaeroides hemAT1*[52], which lacks ALA synthase, suggesting either that ALA is not freely available in these environments or is present at concentrations lower than the 100 nM detection limit of these assays (data not shown). Based on the ecosystem assignment information available for 48% of the genomes, 78% of tetrapyrrole precursor salvagers are categorized as host-associated bacteria compared to 41% in the complete filtered dataset. One interpretation of this finding is that tetrapyrrole precursors are provided by the host, either from host cells that produce them as intermediates in heme biosynthesis[198,199] or, for gut-associated microbes, as part of the host's diet. Alternatively, these precursors may be provided by other microbes, as was observed in a coculture of *Fibrobacter* species[200]. Genome analysis suggests that Candidatus *Hodgkinia cicadicola*, a predicted UroIII salvager[201], may acquire a tetrapyrrole precursor from its insect host or other endosymbionts to be able to provide methionine for itself and its host via the cobamide-dependent methionine synthase. Seventeen percent of cobamide-requiring human gut bacteria lacked genes to make UroIII *de novo* from glutamate, suggesting they could be UroIII salvagers[9].

Nutritional dependence is nearly universal in bacteria. Auxotrophy for B vitamins, amino acids, and nucleic acids is so common that these nutrients are standard components of bacterial growth media. We speculate that the availability of cobamides in the environment, coupled with the relative metabolic cost of cobamide biosynthesis, has driven selection for loss of the cobamide biosynthesis pathway[119]. The large number of genomes with partial cobamide

biosynthesis pathways, namely in the “possible cobamide biosynthesis”, “likely non-producer”, and “Cbi salvager” classifications, suggests that some of these genomes are in the process of losing the cobamide biosynthesis pathway. At the same time, evidence for horizontal acquisition of the cobamide biosynthesis pathway suggests an adaptive advantage for nutritional independence for some bacteria[202,203]. Such advantages could include early colonization of an environmental niche, ability to synthesize cobamides with lower ligands that are not commonly available, or association with hosts that do not produce cobamides. The analysis of the genomic potential of bacteria for cobamide use and production presented here could provide a foundation for future studies of the evolution and ecology of cobamide interdependence.

# Chapter 3

## The flexible cobamide requirements of *Clostridioides* (*Clostridium) difficile* 630 Δ*erm*

**Abstract**

*Clostridioides (Clostridium) difficile* is an opportunistic pathogen whose ability to cause disease is closely linked to its ability to take advantage of changing nutrient conditions during dysbiosis of the human gut microbiota[204,205]. Better understanding of its metabolism may provide additional strategies for *C. difficile* infection prevention and treatment. *C. difficile* has an unusually high number of cobamide-dependent metabolisms encoded in its genome, most of which have not been characterized in their function *in vitro* or *in vivo*, but it does not encode the complete *de novo* cobamide biosynthesis pathway. Cobamides vary in their lower ligand structure. Those lower ligands are not functionally equivalent, and seven different cobamides have been found in human stool. To address the importance of cobamides to *C. difficile* and to understand how *C. difficile* acquires cobamides, we studied wild type and targeted mutants of *C. difficile* 630 Δ*erm in vitro* under cobamide-dependent conditions. Like other aspects of its physiology, *C. difficile* 630 Δ*erm*'s cobamide metabolism is versatile, as it can use diverse cobamides for methionine and deoxyribonucleotide synthesis, and can use alternative nutrient sources or enzymes to fulfill its metabolic needs. We found that *C. difficile* 630 Δ*erm* can produce the cobamide pseudocobalamin when provided the early precursor 5-aminolevulinc acid (ALA) or the late intermediate cobinamide (Cbi), and can produce other cobamides if provided an alternative lower ligand. The ability of *C. difficile* 630 Δ*erm* to take up cobamides and Cbi at physiological concentrations requires the transporter BtuFCD, but interestingly, other *C. difficile* strains including the hypervirulent strain R20291 contain insertions in *btuC*, that may restrict their ability to use cobamides produced by other microbes. *C. difficile* strains are expected to vary in their cobamide acquisition strategies based on the gene content and adapt their metabolic networks as nutrient conditions change.

### Introduction

The human gut microbiota is composed of hundreds to thousands of species of bacteria and archaea, forming a complex community[206]. Members of the community compete for many of the same nutrients, such as carbon sources, but also provide benefits to other members by releasing waste products that these others can use. One example of such trophic interactions is carbon mineralization. First, primary fermenters breakdown carbohydrates from the host mucin or the diet, producing short chain fatty acids (SCFA)[207]. These SFCA can be used by the host, and also metabolized by other microbes into methane and carbon dioxide. Other modes of nutrient sharing include production and auxotrophy of vitamins and amino acids, as many gut bacteria are auxotrophic for one or more vitamins and amino acids[10], and must acquire them from the host diet or other microbes. In one example, when the mucin-degrading bacterium *Akkermansia muciniphila* is co-cultured *in vitro* with *Eubacterium hallii*, *E. hallii* can provide vitamins that allow *A. muciniphila* to access metabolic pathways than it cannot use when co-cultured with *Anaerostipes caccae,* creating a different SCFA profile[164]. The production of these metabolites can affect not only growth of members of the microbiota, but also host health[208].

*Clostridioides (Clostridium) difficile* is a human intestinal pathogen that is among the most common nosocomial infections in the United States, with nearly 300,000 healthcare-associated cases per year[209]. *C. difficile* colonization of the gut is correlated with dysbiosis of the gut microbiota, often after antibiotic treatment[210]. Its abilities to germinate from spores, proliferate in the gut, and cause disease are impacted by ecological and metabolic factors[205,211]. Antibiotic-induced changes in microbiota species richness and abundance results in changes in niche availability that allow *C. difficile* spores to germinate and proliferate in the gut[208]. The global alteration of the gut metabolome following antibiotic treatment is correlated with increased susceptibility to *C. difficile* infection, and recent work has linked changes in specific metabolites to changes in the microbiome using model systems[210,212–214]. For example, succinate availability increases after disturbance of the microbiota, allowing *C. difficile* expansion in a gnotobiotic mouse model[215]. Additionally, specific commensal bacteria have been shown to produce compounds that stimulate *C. difficile* metabolism. In a bi-association, *Bacteroides thetaiotaomicron* can break down host mucin and produce sialic acid that can be used by *C. difficile* for expansion in the gut[216]. *C. difficile* can also induce other members of the microbiota to produce indole, which is thought to create a more favorable environment for the pathogen by inhibiting competing microbes[217].

Some interactions with the microbiota have been shown to be inhibitory to *C. difficile*. Co-culturing with certain *Bifidobacterium* spp. with particular carbon sources reduces *C. difficile* toxin production relative to monoculture[218]. While primary bile acids produced by the host promote *C. difficile* spore germination, *Clostridium scindens* and other 7α-dehydroxylating Clostridia transform these compounds into secondary bile acids, which are inhibitory to *C. difficile*[219,220]. This example in particular shows that compounds in the same class can have different effects on the disease state. Understanding how specific metabolites affect the physiology of *C. difficile* and its virulence will improve the prevention and treatment of *C. difficile* associated disease.

One class of metabolites that has not been explored for their ability to affect *C. difficile* growth and virulence are cobamides, the vitamin $B_{12}$ family of cofactors. Cobamides are used in many microbial metabolisms in the gut including methionine synthesis, deoxyribonucleotide synthesis, acetogenesis, and other carbon catabolism pathways. Over 80% of all sequenced bacteria[221] and of sequenced human gut bacteria[9] have one or more cobamide-dependent enzymes, suggesting that cobamides are widely used cofactors across microbial ecosystems. Strikingly, fewer than 40% of bacterial species are predicted to produce cobamides *de novo*[9,221], and therefore over half of bacteria that use cobamides must acquire them from their environment. Cobamides vary in the structure of the lower ligand (Fig. 1), and organisms studied to date are selective in which cobamides they can use[9,11,20,21,93,107,117]. Seven cobamides in addition to the cobamide precursor cobinamide (Cbi) have been detected in the human gut at concentrations over 10 ng per gram of feces[187]. As discussed in Chapter 1, in an environment with plentiful, diverse cobamides and cobamide precursors, if a microbial species requires a particular cobamide, it could either synthesize it *de novo,* remodel cobamides to the preferred structure, import select cobamides, or regulate its cobamide-dependent enzymes to prevent expression when inhibitory cobamides are present in the cell.

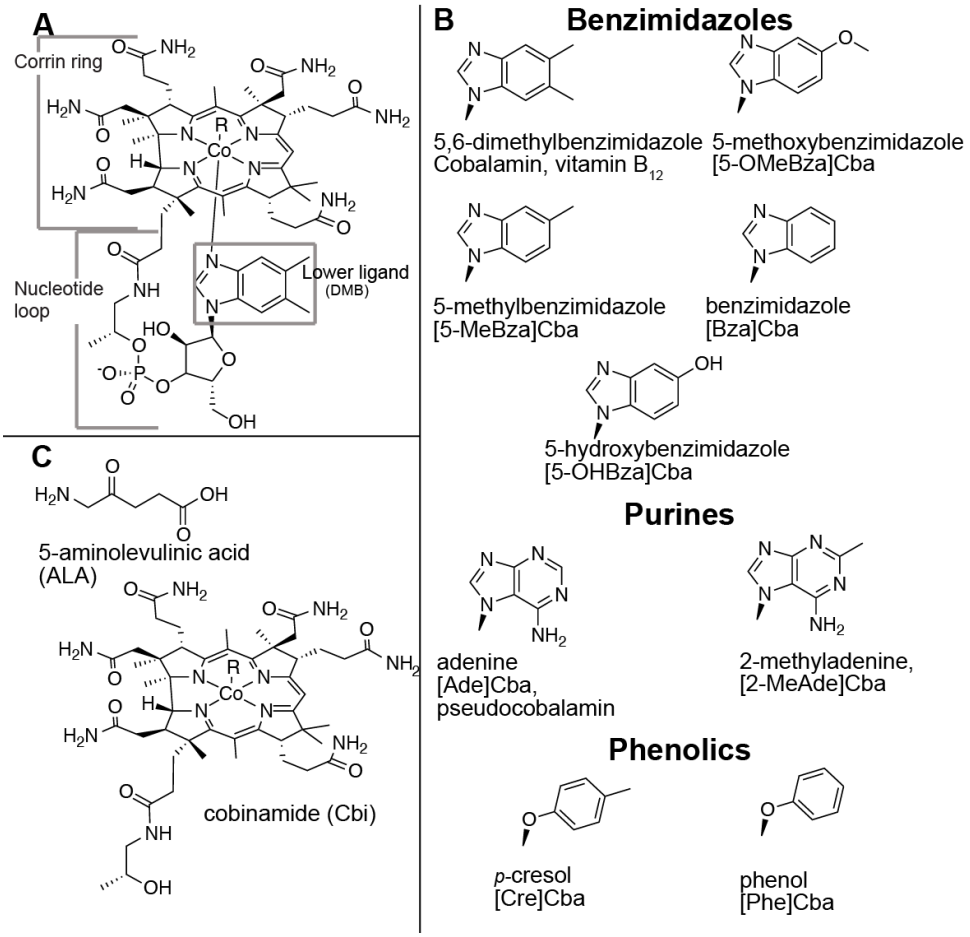



**Figure 1. Cobamide and cobamide precursor structures. A.** Structure of cobalamin (vitamin $B_{12}$), a cobamide **B.** Lower ligands of cobamides analyzed in this study, with the three classes of lower ligand structure labeled. For each structure, the lower ligand name, abbreviation for the cobamide containing the lower ligand, and alternate names of the cobamide (when applicable) are indicated. **C**. Cobamide precursors used in this study.

*C. difficile* genomes contain most of the approximately 30 genes required for *de novo* cobamide biosynthesis, but they are missing *hemA* and *hemL*, the first two enzymes in the pathway required for synthesis of the precursor 5-aminolevulinic acid (ALA) (Fig. 1C). Therefore, *C. difficile* is predicted to be able to produce cobamides only when ALA is available[221] (Fig. 2). *C. difficile* has seven predicted cobamide-dependent enzymes encoded in its genome that are involved in methionine synthesis, nucleotide metabolism, and carbon metabolism (Fig. 2). Having access to cobamides therefore may be important to *C. difficile*, but little is known about *C. difficile*'s use of these pathways *in vitro* or *in vivo*. Deletion of EutA, the reactivating factor required for ethanolamine ammonia lyase (EutBC) activity, in *C. difficile* strain 630 Δ*erm* reduces the mean time to morbidity in a hamster model, suggesting that access to ethanolamine is important in modulating virulence[222]. Metabolic models and transcriptomics[223,224] suggest that the Wood-Ljungdahl pathway is an important electron sink, and there is experimental evidence that it may be used for autotrophic growth by some strains[225]. To be able to use the Wood-Ljungdahl pathway, catabolize ethanolamine, and perform other cobamide-dependent metabolisms, we predict that *C. difficile* requires

cobamides or cobamide precursors such as ALA present in the gut. While ALA is an intermediate made in all tetrapyrrole-producing organisms, including the host, cobamides are only produced by some bacteria and archaea.

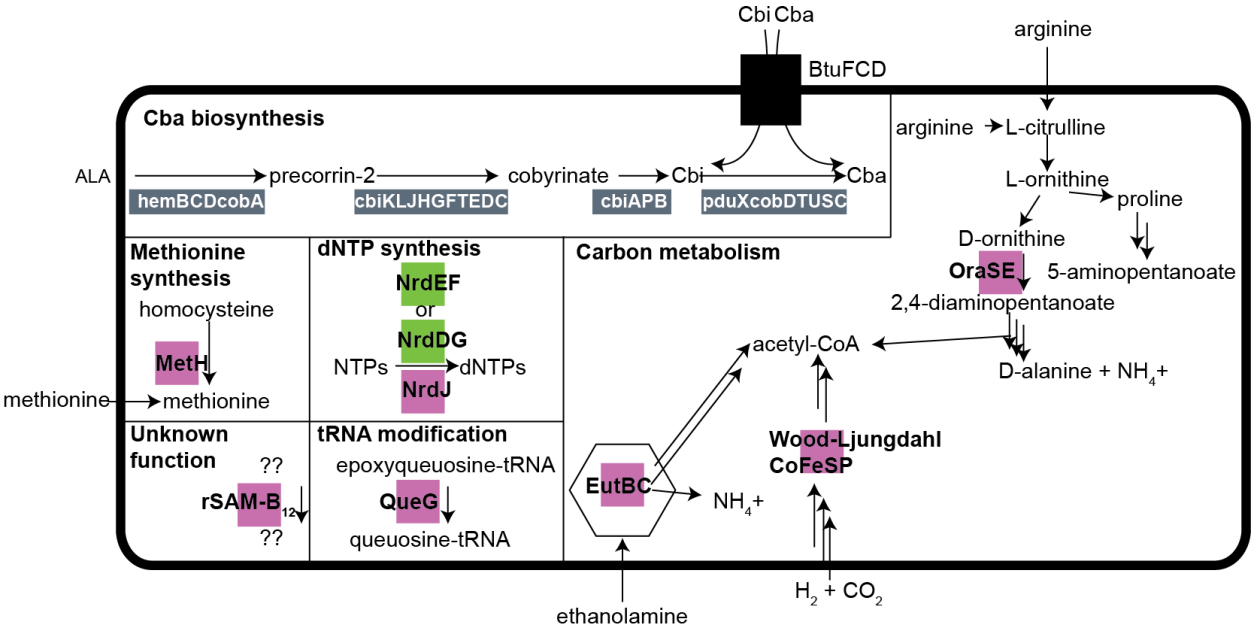



**Figure 2. Cobamide-dependent processes in *C. difficile*.** Diagram of cobamide metabolism in *C. difficile* 630 Δ*erm*. The cobamide biosynthesis pathway is in gray. In purple are cobamide-dependent enzymes, in green are cobamide-independent isozymes. The transporter BtuFCD is in black. Abbreviations: Cba, cobamide; Cbi, cobinamide; ALA, 5-aminolevulinic acid; rSAM, radical S-adenosylmethionine; NTPs, ribonucleotides; dNTPs, deoxyribonucleotides. Enzymes: MetH, methionine synthase; NrdEF, aerobic (oxygen-requiring) cobamide-independent ribonucleotide reductase (RNR); NrdDG, anaerobic (oxygen-sensitive) cobamide-independent RNR; NrdJ, cobamide-dependent RNR; QueG, epoxyqueuosine reductase; EutBC, ethanolamine ammonia lyase; CoFeSP, corrinoid iron-sulfur protein; OraSE, D-ornithine 4,5-aminomutase.

To address the importance of cobamides to *C. difficile* and to understand how *C. difficile* acquires cobamides, we studied wild type and targeted mutants of *C. difficile* 630 Δ*erm in vitro* under cobamide-dependent conditions. We found that, like other aspects of its physiology, *C. difficile* 630 Δ*erm*'s cobamide metabolism is versatile, as it can use a surprising diversity of cobamides for methionine and deoxyribonucleotide synthesis, and can use alternative nutrient sources or enzymes to fulfill its metabolic needs. We found that *C. difficile* 630 Δ*erm* can produce the cobamide pseudocobalamin when provided ALA or the late intermediate Cbi, and can produce other cobamides if provided an alternative lower ligand. The ability of *C. difficile* 630 Δ*erm* to take up cobamides and Cbi at physiological concentrations requires the transporter BtuFCD, but interestingly, other *C. difficile* strains including the hypervirulent strain R20291 contain insertions in *btuC*, which may restrict their ability to use cobamides produced by other microbes.

## Materials and Methods
### Strains and growth conditions

**Table 1: Bacterial strains and plasmids**

| Strain or plasmid | Description | Source |
|---|---|---|
| **Strains** | | |
| *Escherichia coli* XL1-Blue | | QB3 MacroLab |
| *Escherichia coli* CA434 | hsd20(rB-, mB-, recA13, rpsL20, leu, proA2, with IncPb conjugative plasmid R702 | Chain Biotech[226] |
| *Clostridioides difficile* | | |
| 630 Δ*erm* | Erythromycin sensitive strain | [227] |
| 630 Δ*erm* Δ*pyrE* | | [228] |
| 630 Δ*erm* Δ*pyrE* Δ*btuFCD* | | This study |
| 630 Δ*erm* Δ*pyrE* Δ*cbiKLJHGFTEDC* | | This study |
| 630 Δ*erm* Δ*pyrE* Δ*nrdDG* | | This study |
| **Plasmids** | | |
| R702 | Conjugation plasmid | [226] |
| pMTL-YN3 | Allelic exchange vector | [228] |
| pXL001 | pMTL-YN3 directed to *btuFCD* | This study |
| pXL002 | pMTL-YN3 directed to *cbiKLJHGFTEDC* | This study |
| pXL003 | pMTL-YN3 directed to *nrdDG* | This study |

*C. difficile* 630 Δ*erm* and *C. difficile* 630 Δ*erm* Δ*pyrE* were streaked from frozen stocks onto BHIS agar[229] before transferring to *Clostridium difficile* defined medium (CDDM), containing casamino acids,[230] with 8 g/L glucose. Agar plates and 96-well plate cultures were incubated at 37 ºC in an anaerobic chamber (Coy Labs) under 10% $H_2$ 10% $CO_2$ 80% $N_2$ headspace. For *C. difficile* 630 Δ*erm* Δ*pyrE* and derived strains, 5 µg/ml uracil was included in all defined media. For corrinoid extractions and NrdJ phenotype experiments, strains pre-cultured in CDDM plus 8 g/L glucose and inoculated into the same medium. For MetH phenotype experiments, CDDMK medium plus 8 g/L glucose without methionine was used. CDDMK contains the same salts, trace metals, and vitamins as CDDM, but the casamino acids, tryptophan and cysteine are replaced with the individual amino acids as follows: 100 mg/L histidine, 100 mg/L tryptophan, 100 mg/L glycine, 100 mg/L tyrosine, 200 mg/L arginine, 200 mg/L phenylalanine, 200 mg/L threonine, 200 mg/L alanine, 300 mg/L lysine, 300 mg/L serine, 300 mg/L valine, 300 mg/L isoleucine, 300 mg/L aspartic acid, 400 mg/L leucine, 500 mg/L cysteine, 600 mg/L proline, 900 mg/L glutamic acid[231]. All defined liquid media were prepared by boiling under 80% $N_2$/20% $CO_2$ gas. After the pH stabilized between 6.8 and 7.2, the medium was dispensed into stoppered tubes and autoclaved. Sterile glucose and vitamins were added after autoclaving.

For preculturing for MetH phenotype assays, *C. difficile* 630 Δ*erm* was grown in CDDM medium, then washed twice in CDDMK without methionine prior to inoculation at OD = 0.01 in a 96 well plate. *C. difficile* 630 Δ*erm* Δ*pyrE* Δ*btuFCD* was prepared in the same manner, but the medium contained 5 µg/L added uracil. For NrdJ phenotype assays, *C. difficile* 630 Δ*erm* Δ*pyrE* Δ*nrdDG* was grown in CDDM plus 5 µg/ml uracil plus 10 nM cobalamin, and washed three times in medium without cobalamin prior to inoculation at OD = 0.01 in a 96 well plate.

$OD_{600}$ was read on a Biotek Synergy 2 plate reader after 23 to 24 hours. Blank medium wells were subtracted from the measurement and corrected for pathlength.

ALA**,** Cbi and cobalamin were purchased from Sigma Aldrich. Other cobamides were produced as described in Men *et al*.[108]

**Strain and plasmid construction**

The allelic coupled exchange (ACE) system of Ng *et al*. was used for construction of *C. difficile* mutant strains (Table 1)[228]. Briefly, arms of homology flanking the target gene(s) in the *C. difficile* 630 Δ*erm* genome (CP016318) were cloned into pMTL-YN3 (Chain Biotech) via PCR (Table 2) and Gibson assembly[232] in *E. coli* XL1- Blue. Plasmid inserts were sequenced by Sanger sequencing before transformation into *E. coli* CA434 (Chain Biotech). Conjugation of *E. coli* CA434 and *C. difficile* 630 Δ*erm* Δ*pyrE* was performed as described[233], except that *C. difficile* and *E. coli* cultures were cultured for 5-8 hours prior to pelleting *E. coli* and mixing with the *C. difficile* recipient. After 16 hours growth on BHIS agar, the mixed cells were resuspended in 1 ml PBS and 100 µl of the suspension was plated on each of 5-7 plates of BHIS agar with added 10 µg/ml thiamphenicol, 250 µg/ml D-cycloserine, and 16 µg/ml cefoxitin. After colonies were visible, they were purified by streaking onto BHIS with 15 µg/ml thiamphenicol, 250 µg/ml D-cycloserine, and 16 µg/ml cefoxitin, at least twice before counterselection on CDDM agar plus 2 mg/L 5-fluoroorotic acid (5-FOA) plus 5 µg/ml uracil. The resulting colonies were purified by streaking at least twice prior to screening by colony PCR for the deletion and the presence of *tcdB* (Table 2)[233]. For the deletion of *nrdDG,* 10 nM cobalamin was added to all media during the ACE procedure.

**Table 2. Primers used in this study**

| Primer ID | Sequence | Purpose |
|---|---|---|
| P2188 | CAT AAT ATG TCA GAG AAT ACT GTA GTC | tcdB presence |
| P2189 | GTT CTG AGG TAT ATT CTG GTA TAT ATT | tcdB presence |
| P2269 | GCT TGA TGT GTT GGT AGC AC | Checking pMTL-YN3 for insert |
| P2270 | AAG TAC ATC ACC GAC GAG CA | Checking pMTL-YN3 for insert |
| P2271 | AAC AGC TAT GAC CGC GGC CGC ATT TTA ATG AAA ACT ATT TC | Gibson assembly primer for pMTL-YN3 and btuFCD arms of homology |
| P2272 | TTC AAA AAA ATT ATA ATC TAT ACT CTA ATT TAT TGT TGA CCT CTT TGC AAG G | Gibson assembly primer for pMTL-YN3 and btuFCD arms of homology |
| P2273 | CCT TGC AAA GAG GTC AAC AAT AAA TTA GAG TAT AGA TTA TAA TTT TTT TGA A | Gibson assembly primer for pMTL-YN3 and btuFCD arms of homology |
| P2274 | CAG GCC TCG AGA TCT CCA TGG TAT GGA TAT GCA AAA AGA AC | Gibson assembly primer for pMTL-YN3 and btuFCD arms of homology |
| P2282 | ACA GCT ATG ACC GCG GCC GCC TAA TTT CTA TAG CTA AAG C | Gibson assembly primer for pMTL-YN3 and cbiKLJHGFTEDC arms of homology |
| P2286 | AGT CTC CTT TAA ATA TTG CTT TCA CTT ATG TAT TCC ATT CTA TTT CCC CCT TAA T | Gibson assembly primer for pMTL-YN3 and cbiKLJHGFTEDC arms of homology |
| P2287 | ATT AAG GGG GAA ATA GAA TGG AAT ACA TAA GTG AAA GCA ATA TTT AAA GGA GAC T | Gibson assembly primer for pMTL-YN3 and cbiKLJHGFTEDC arms of homology |
| P2288 | AGG CCT CGA GAT CTC CAT GGA TAC CTG TTG GAA AAG GAA T | Gibson assembly primer for pMTL-YN3 and cbiKLJHGFTEDC arms of homology |

| | | |
|---|---|---|
| P2289 | ACA GCT ATG ACC GCG GCC GCC AAT AAG TTT TTT ACA GAT T | Gibson assembly primer for pMTL-YN3 and nrdDG arms of homology |
| P2290 | CAA TAT ATA GTA ACA GGA GGT TTT TTT AAA ATA TAA ATA AAC AGG ATT AAA TAT ATG C | Gibson assembly primer for pMTL-YN3 and nrdDG arms of homology |
| P2291 | GCA TAT ATT TAA TCC TGT TAA TTT ATA TTT TAA AAA AAC CTC CTG TTA CTA TAT ATT G | Gibson assembly primer for pMTL-YN3 and nrdDG arms of homology |
| P2292 | TCT GCA GGC CTC GAG ATC TCC ATG GTA TTA CTA TAC CAA CTT TTT CTT TTA GAG T | Gibson assembly primer for pMTL-YN3 and nrdDG arms of homology |
| P2404 | GAA GGT GAT TTT AAT GAA AAC TAT TTC TAT TTC TAA ACA AG | Flanking primer for checking btuFCD knockout |
| P2405 | TAT GCA AAA AGA ACT TAT AGA TTT AGT AAC TAG TC | Flanking primer for checking btuFCD knockout |
| P2406 | CGT CCG TCT TAT CTA CTG ATT GAT AAT AAT C | Internal primer for checking btuFCD knockout |
| P2663 | GCA AAA TAT GAT TAC TTG ATG CCT TG | Flanking primer for checking cbiKLJHGFTEDC knockout |
| P2664 | TCT TAC AAG CAA CAC TGA AAT TAT G | Flanking primer for checking cbiKLJHGFTEDC knockout |
| P2451 | CAG GAT AAC TAA CCC AAT AAG GCT TTG GTA ATA AGA CTT C | Flanking primer for checking nrdDG knockout |
| P2452 | TGC TTT ATT TGT TTG CCC TTT TCT TGA GGA C | Flanking primer for checking nrdDG knockout |
| P2458 | GAA GAT ATA CAA GAT TCT GTA GTT AAG GTT C | Internal primer for checking nrdDG knockout |
| P2459 | AGA AGT ATC TGT TCC GAA GTT TAC ACT TG | Internal primer for checking nrdDG knockout |

**Corrinoid extraction and analysis**

*C. difficile* was grown in 50 ml CDDM plus 8 g/L glucose under 80% $N_2$/ 20% $CO_2$ headspace for 16-22 hours at 37ºC prior to extraction. Two cultures were combined for each condition for a total volume of 100 ml. Corrinoid extractions were performed as described[107], except that cell pellets were autoclaved for 35 minutes prior to addition of methanol and potassium cyanide.

High-performance liquid chromatography (HPLC) analysis was performed with an Agilent Series 1200 system (Agilent Technologies, Santa Clara, CA) equipped with a diode array detector with detection wavelengths set at 362 and 525 nm. For Fig. 6B and 7A, samples were injected onto an Agilent Zorbax SB-Aq column (5 μm, 4.6 × 150 mm) at 30 °C, with 1 mL/min flow rate. Compounds in the samples were separated with a gradient of 25 to 34% methanol containing 0.1% formic acid in acidified water (containing 0.1% formic acid) over 11 min, followed by a 34 to 50% gradient over 2 min, and 50 to 75% over 9 min. For Fig. 6A, samples were injected onto an Agilent Eclipse Plus C18 column (5 μm, 9.6 × 250 mm) at 30 ºC, with 2 mL/min flow rate. Compounds in the samples were separated with a gradient of 10 to 42% methanol with 0.1% formic acid in acidified water (containing 0.1% formic acid) over 20 min, then 100% methanol for 2 minutes. Injection volumes were 5 μL of standards for total amounts injected: Cbi (1), 200 pmoles; pseudocobalamin (2), 225 pmoles; cobalamin (3), 50 pmoles; [Bza]Cba (4), 114 pmoles; [2-MeAde]Cba (5), 114 pmoles; [Cre]Cba (6), 151 pmoles. 10 μl *C. difficile* samples, and 40 μl *C. difficile* Δ*pyrE* mutant samples were injected.

***C. difficile* MLST tree construction**

For the 248 *C. difficile* genomes classified as "finished" or "permanent draft" on JGI/IMG[234] (accessed March 2019), the seven MLST gene sequences, *adk, atpA, dxr, glyA, recA, sodA,* and *tpi* [235], were downloaded and aligned individually using MUSCLE[236]. The

alignments were concatenated and genomes missing one or more MLST genes or having duplicate genes were removed from the analysis for a total of 79 strains analyzed. The concatenated alignment was manually trimmed in UGENE[237], then columns with 95% or greater gaps were removed with trimAL[238]. This alignment was used as input for RAxML 8.2.12[239] on the CIPRES webserver[240] with 100 bootstraps, using the GTRCAT model. The tree was visualized and annotated in iTOL[241].

## Results

### *C. difficile* requires methionine or a cobamide for growth

*C. difficile* encodes seven cobamide dependent enzymes in its genome (Fig. 2), but whether all of these enzymes are functional and which cobamides can be used for each pathway are unknown. Common *in vitro* growth conditions using casamino acid medium or brain heart infusion broth allow *C. difficile* access to many carbon and energy sources that may not require cobamides for use. Like some other Clostridia we have tested including *Clostridium scindens* and *Clostridium sporogenes* (Chapter 2), the genome of *C. difficile* 630 Δ*erm* encodes the complete anaerobic cobamide biosynthesis pathway with the exception of *hemA* and *hemL*, two genes required for the synthesis of the first committed tetrapyrrole precursor, 5-aminolevulinc acid (ALA)[53,54,221]. Therefore, we predicted that, like *C. scindens* and *C. sporogenes*, *C. difficile* would be able to produce a cobamide only if ALA or a downstream cobamide precursor is provided[221], and otherwise be reliant on exogenous cobamides for cobamide-dependent growth.

To interrogate the cobamide-dependent enzymes of *C. difficile*, we sought to grow the organism in conditions that functionally isolate each enzyme. *C. difficile* encodes the cobalamin-dependent methionine synthase MetH, but does not contain the cobamide-independent alternative enzyme MetE, suggesting that it requires a cobamide or methionine for growth. Previously, methionine has been classified as a "growth-enhancing," but not essential, amino acid in a medium containing cobalamin (vitamin $B_{12}$)[231]. We modified the CDDM casamino acid medium by replacing the casamino acids with individually added amino acids, and omitted methionine. To test whether *C. difficile* can use cobamides for methionine synthesis and to identify the specific cobamides that support its MetH-dependent growth, we cultured *C. difficile* with a range of concentrations of cobalamin, Cbi, and eight other cobamides that we purified. Remarkably, unlike other bacteria that have been reported to use only certain cobamides for methionine synthase activity[9,19,64], all of the cobamides and Cbi were able to confer high growth yields to *C. difficile* at concentrations as low as 1 nM (Fig. 3). Higher concentrations of ALA and methionine were required to support growth (Fig. 3). The observation that *C. difficile* could grow in a MetH-dependent condition with ALA and Cbi suggested that its encoded cobamide biosynthesis pathway lacking the ALA biosynthesis genes was indeed functional.

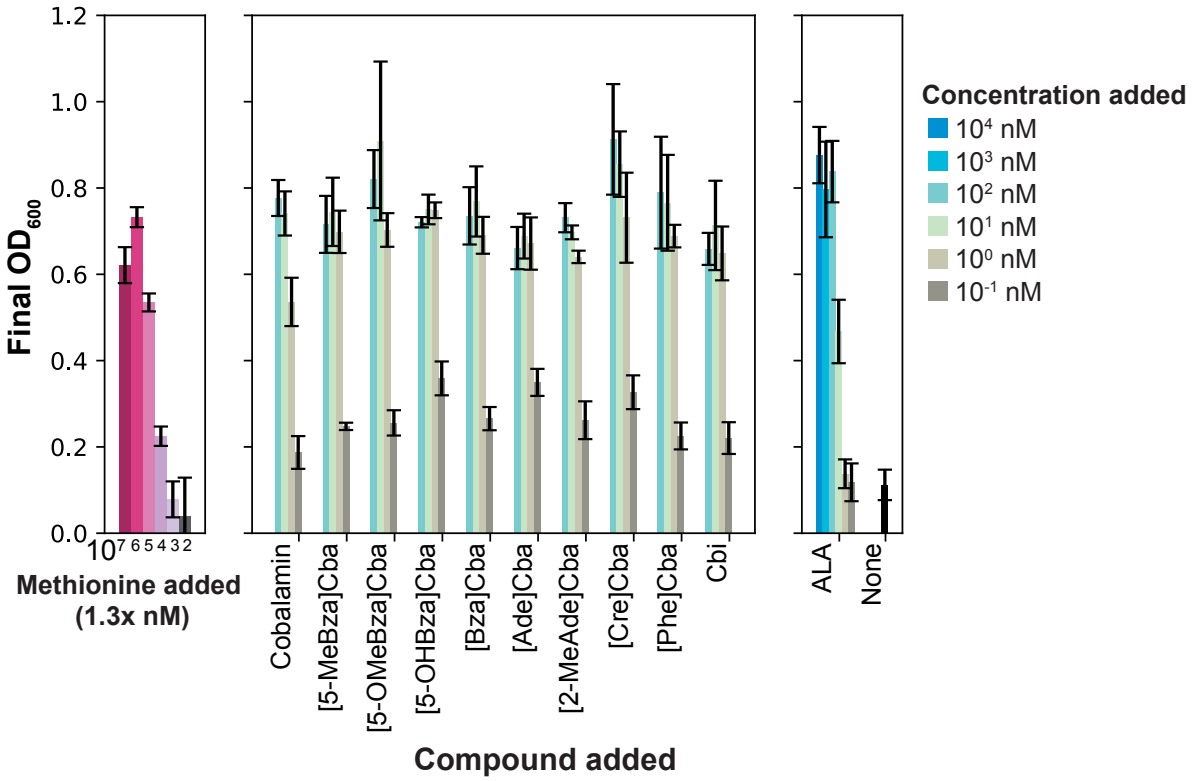



**Figure 3. Growth of *C. difficile* 630 Δ*erm* under MetH-requiring conditions.** The mean $OD_{600}$ of four stationary phase cultures grown for 22.5 hours in CDDMK medium plus glucose without methionine is plotted as a function of the amount of compound added in nM. Error bars are the standard deviation of four biological replicates.

### *C. difficile* ribonucleotide reductase NrdJ requires a more restricted set of cobamides

*C. difficile* has three ribonucleotide reductases (RNRs)[33] encoded in its genome: an oxygen-dependent class I (*nrdEF,* CDIF630erm_RS16320, CDIF630erm_RS16325), a cobalamin-dependent class II (*nrdJ,* CDIF630erm_RS07280), and an oxygen-sensitive, cobamide-independent class III (*nrdDG*, CDIF630erm_RS00990, CDIF630erm_RS00995). In principle, any of these three isozymes could be used for deoxyribonucleotide synthesis from ribonucleotides. When grown anaerobically in CDDM with glucose, adding cobamides or cobamide precursors did not affect growth yield of the parent strain *C. difficile* 630 Δ*erm* Δ*pyrE* (Fig. 4), suggesting that the class III RNR, NrdDG, is functional under these growth conditions. To test whether the class II RNR, NrdJ, is functional, and if so, to study its cobamide preference, we deleted the *nrdD* and *nrdG* genes while providing exogenous cobalamin. This strain could grow upon cobamide addition, indicating that NrdJ is functional. In contrast to MetH, NrdJ is more selective in the cobamides it can use (Fig. 5). Since the cobamide cofactor used by NrdJ has a "base-on" configuration[13], we expected that only benzimidazolyl and purinyl cobamides, which can form a coordination bond between the cobalt and lower ligand, would function in *C. difficile*'s NrdJ. As predicted, there was little growth with [Cre]Cba and [Phe]Cba. [5-OHBza]Cba also did not support growth, despite its ability to adopt a base-on configuration.

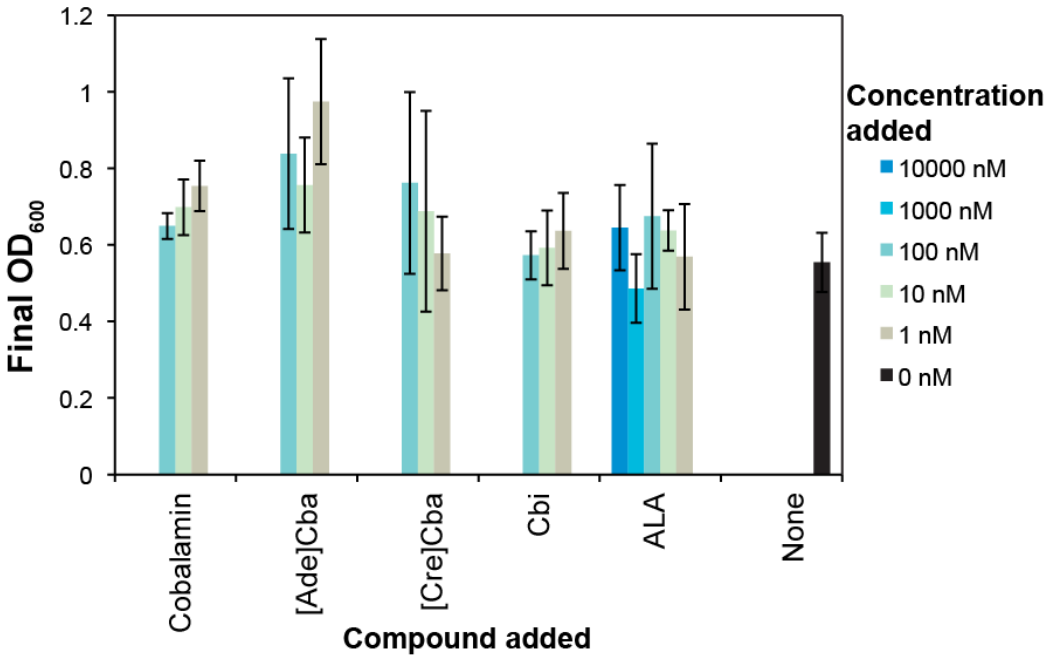



**Figure 4. *C. difficile* 630 Δ*erm* Δ*pyrE* does not change its growth yield in response to external cobamides if methionine is present**. *C. difficile* 630 Δ*erm* Δ*pyrE* was grown in CDDM casamino acid medium plus uracil and glucose for 23.5 hours. The mean $OD_{600}$ of three biological replicates is plotted. Error bars are the standard deviation.

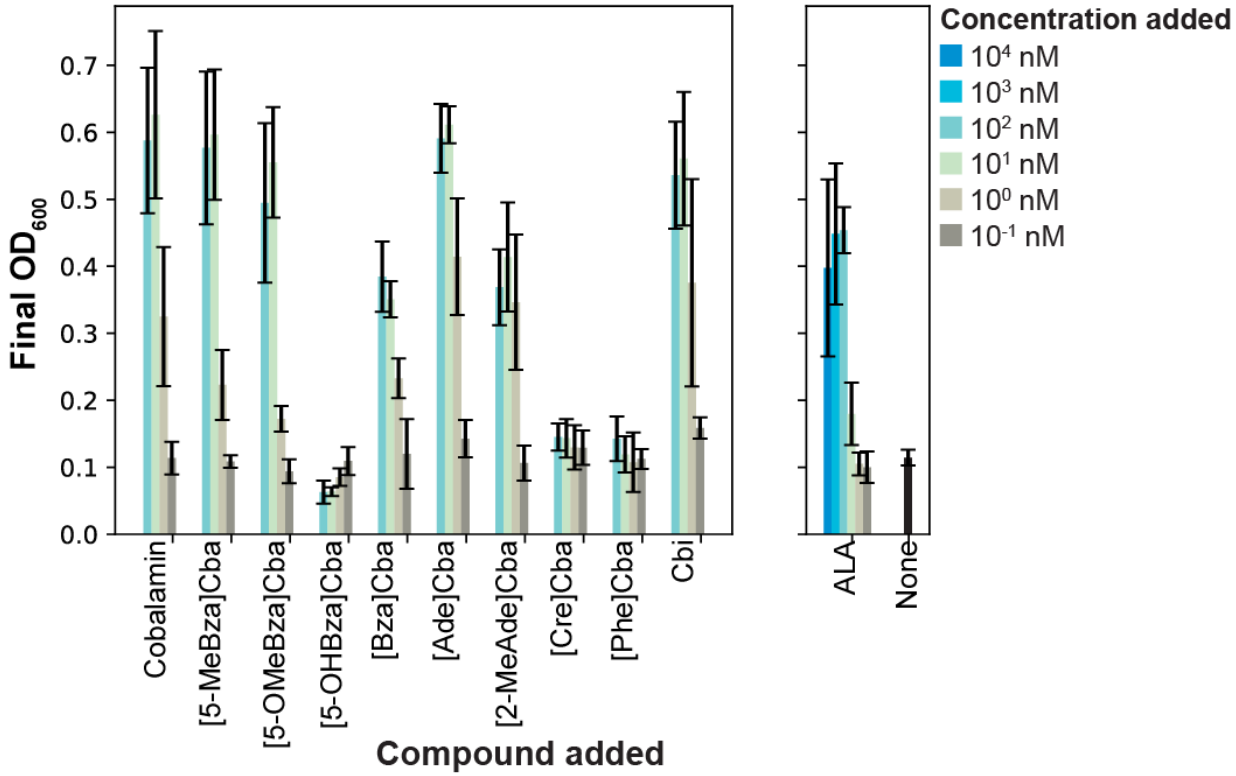



**Figure 5. *C. difficile* NrdJ is selective in which cobamides it can use.** The mean $OD_{600}$ of three stationary phase cultures of *C. difficile* 630 Δ*erm* Δ*pyrE* Δ*nrdDG* grown in CDDM with added uracil and glucose for 22.5 hours is plotted as a function of the amount of compound added in nM. Error bars are the standard deviation of three biological replicates.

### *C. difficile* can produce pseudocobalamin from the precursor ALA via the *cbi* genes

The observation that *C. difficile* could grow under cobamide-dependent conditions with ALA or Cbi (Fig. 3, 5) suggested that it can produce a cobamide from these precursors using the cobamide biosynthetic genes encoded in its genome[221]. Because *C. difficile* lacks all known genes for biosynthesis of benzimidazoles and attachment of phenolic lower ligands, it is predicted to be incapable of producing benzimidazolyl or phenolyl cobamides, but may produce a purinyl cobamide[64,66–70,75,184]. To test these predictions, *C. difficile* was grown in CDDM medium with glucose with ALA or Cbi, a condition that does not require cobamides. The corrinoid fraction, which includes cobamides and cobamide precursors including Cbi, was extracted from the cell pellets following growth to saturation. Consistent with our predictions, HPLC analysis of the extracted corrinoids showed that *C. difficile* produces a cobamide only when ALA or Cbi was added (Fig. 6A). Moreover, the major cobamide produced co-eluted with the purinyl cobamide pseudocobalamin (Fig. 6A).

We expect that pseudocobalamin synthesis from the early precursor ALA, but not from the late precursor Cbi, relies on the corrin ring biosynthesis genes *cbiKLJHGFTEDC* (Fig. 2). We constructed a *cbiKLJHGFTEDC* deletion mutant, and as expected, this strain could synthesize pseudocobalamin with added Cbi, but not with ALA (Fig. 6A).

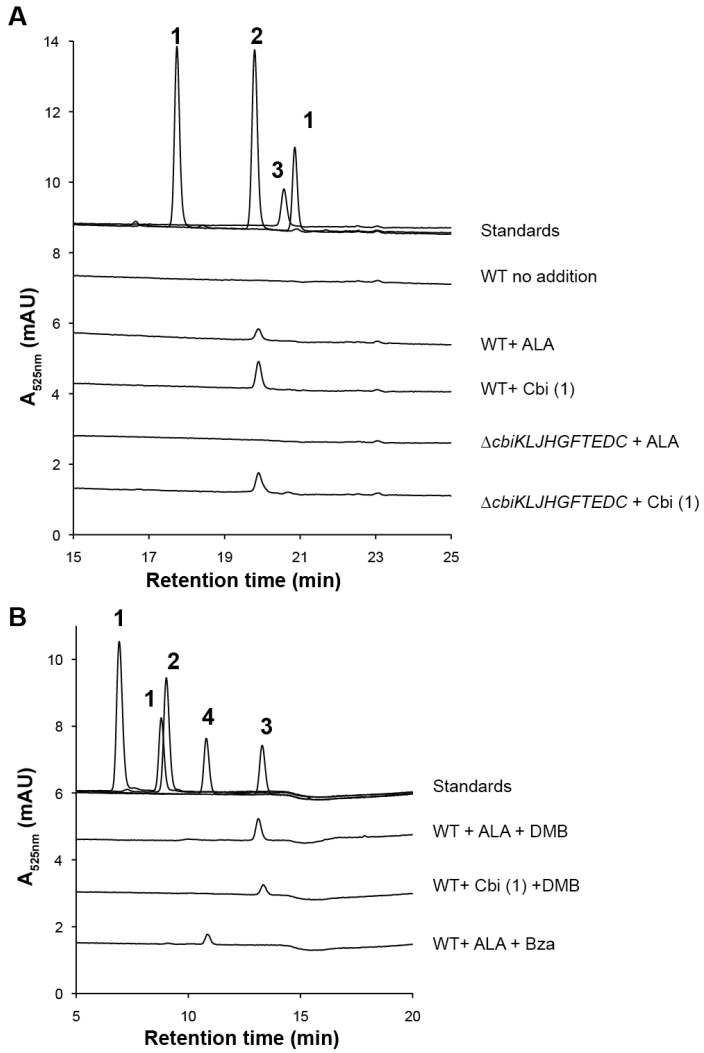



**Figure 6. HPLC analysis of corrinoid extracts from *C. difficile* strains. A.** Production of cobamides with added cobamide precursors is shown for *C. difficile* 630 Δ*erm* and 630 Δ*erm* Δ*pyrE* Δ*cbiKLJHGFTEDC* grown with added ALA (100 nM) or Cbi (10 nM). Cbi **(1)**, pseudocobalamin **(2)**, and cobalamin **(3)** are shown as standards. **B**. Guided biosynthesis with added benzimidazoles. *C. difficile* 630 Δ*erm* with added ALA (100 nM) or Cbi (10 nM) and lower ligands DMB or Bza (100 nM). Cbi **(1)**, pseudocobalamin **(2)**, cobalamin **(3)**, and [Bza]Cba **(4)** are shown as standards.

### *C. difficile* can perform guided biosynthesis but does not remodel cobamides

To test whether *C. difficile* can perform guided biosynthesis, a process in which an exogenously provided lower ligand base is incorporated into a cobamide [19,74,91,93,94], *C. difficile* was cultured with ALA or Cbi and DMB (the lower ligand of cobalamin, Fig. 1A) or a related compound, benzimidazole (Bza). Analysis of corrinoid extracts showed that *C. difficile* could attach both exogenous lower ligands to form cobalamin and [Bza]Cba, respectively (Fig. 6B).

Some bacteria and archaea are able to remodel cobamides by removing the lower ligand and nucleotide loop with the amidohydrolase enzyme CbiZ and rebuild the cobamide with a different lower ligand [104–107]. We did not find *cbiZ* in the *C. difficile* genome, and as

expected, we did not observe evidence of remodeling; when cobalamin, [2-MeAde]Cba, or [Cre]Cba was provided to *C. difficile*, the same cobamides were recovered from the cells (Fig. 7A).

### *C. difficile* requires *btuFCD* for efficient uptake of cobamides and Cbi

Recovery of added cobamides and cobamides made from Cbi from the cellular fraction (Fig. 6, 7A) suggests that *C. difficile* takes up cobamides via an active transporter. We identified a candidate vitamin $B_{12}$ uptake operon (*btuFCD*) downstream of a sequence annotated as a cobalamin riboswitch, suggesting that these genes function in corrinoid import[8,9,76,77,83,84,86,88]. We constructed a deletion of this operon and found that, unlike wild type *C. difficile*, when grown with 10 nM cobalamin, no cobalamin could be recovered in corrinoid extracts from the cellular fraction of the Δ*btuFCD* mutant (Fig. 7A). In contrast, ALA uptake is unaffected in the Δ*btuFCD* mutant, as this strain can produce pseudocobalamin when ALA is provided (Fig. 7A).

Consistent with the requirement for the corrinoid transporter BtuFCD, the Δ*btuFCD* mutant grew poorly in medium without methionine even with $10^4$ to $10^5$ nM added cobinamide or cobalamin, while the ability of methionine or ALA to support growth remained unaffected by the Δ*btuFCD* mutation (Fig. 7B).

Some strains of *C. difficile* in the PCR-ribotype 027 (RT027) clade have a *tlpB* transposon insertion in *btuC*, likely rendering the BtuFCD transporter nonfunctional (Fig. 8)[242]. This insertion is also found in the ancestral strain CD196[242]. In *Escherichia coli* and *Synecoccocus* sp. PCC 7002, *btuC* deletion mutants are impaired in cobalamin uptake[88,243]. Based on the absence of corrinoids in the cell fraction of a 630 Δ*erm* Δ*btuFCD* strain (Fig. 7), we infer that strains with an insertion in *btuC*, including the hypervirulent R20291, would require cobamides or Cbi at concentrations higher than 0.1 mM in the environment if relying on cobamide-dependent enzymes. In stool samples of individuals not taking cobalamin supplements at high doses, corrinoids are present at tens to hundreds of ng per gram; the most abundant cobamide was [2-MeAde]Cba at an average of 794 ng/g but cobalamin was found at concentrations averaging 19 ng/g and was among the lowest concentration detected[187]. For [2-MeAde]Cba, this is an equivalent of approximately 500 nM, suggesting that strains without a functional BtuFCD transporter may not be able to use the corrinoids present in the gut. (Fig. 7B). However, they would be able to use ALA if it was available.

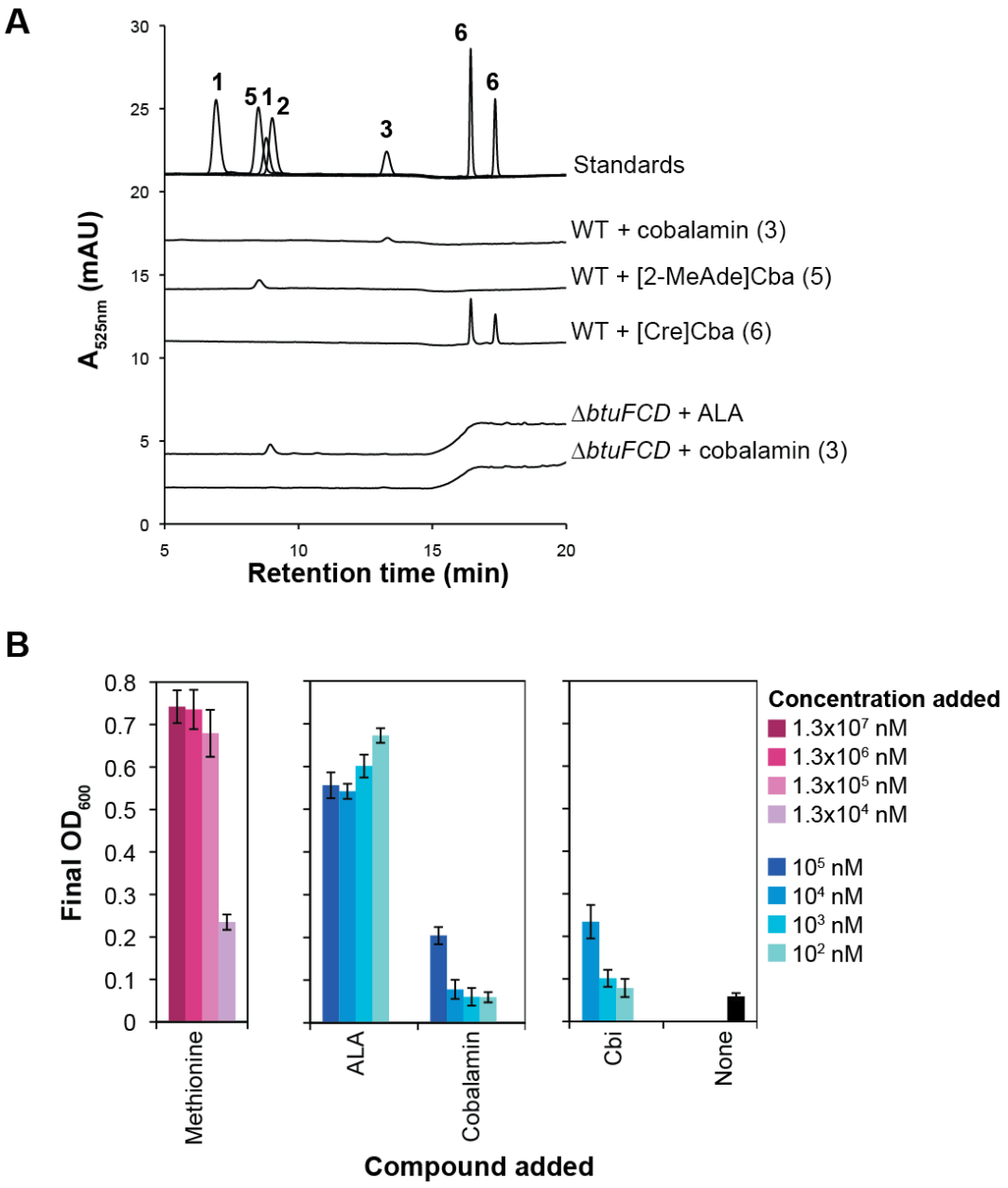



**Figure 7. Δ*btuFCD* mutants are impaired in cobamide and Cbi uptake. A.** HPLC analysis of corrinoid extracts from cell pellets of *C. difficile* 630 Δ*erm* and *C. difficile* 630 Δ*erm* Δ*pyrE* Δ*btuFCD* were grown with 10 nM cobamides or 100 nM ALA. Cbi **(1)**, pseudocobalamin **(2)**, cobalamin **(3)**, [2-MeAde]Cba **(5),** [Cre]Cba **(6)** shown as standards. **B.** *C. difficile* 630 Δ*erm* Δ*pyrE* Δ*btuFCD* grown in MetH-dependent conditions. The mean $OD_{600}$ of three biological replicates stationary phase cultures in CDDMK with 5 ug/ml uracil without methionine for 23.5 hours is plotted as a function of the amount of compound added in nM. Error bars are the standard deviation of three biological replicates.

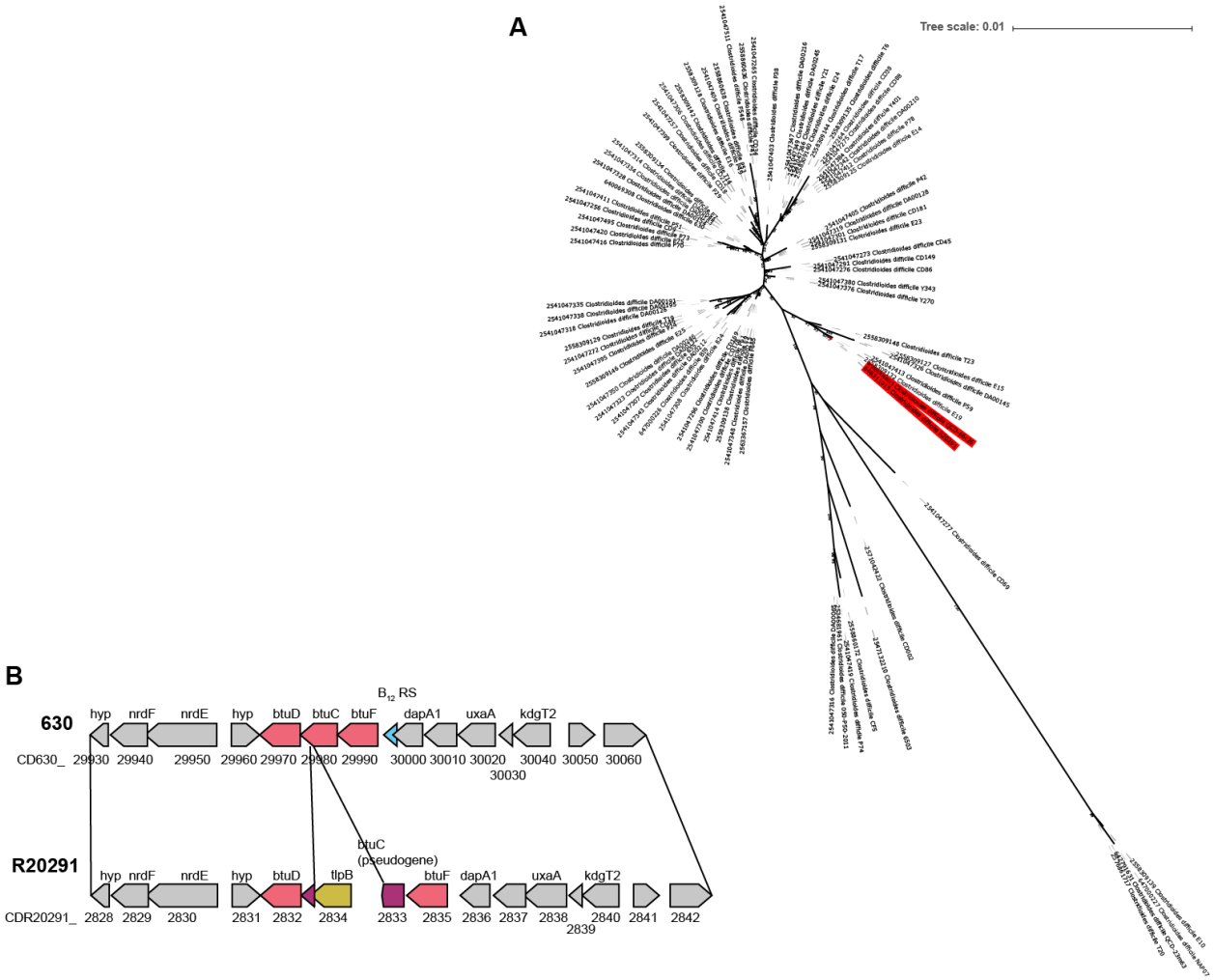



**Figure 8. Distribution of *tlpB* insertion in *btuC* in C. *difficile* strains A.** Maximum likelihood MLST tree of *C. difficile* strains. Branch labels are bootstrap support from 100 bootstraps. Genomes with the transposon insertion into *btuC* are highlighted in red, other branches have intact *btuFCD* ORFs. **B.** The gene neighborhood of selected genomes at the *btuFCD* locus shows the *tlpB* insertion. Neighborhoods were generated using JGI/IMG's best COG matches.

## Discussion

*C. difficile's* ability to cause disease is closely linked to its ability to take advantage of changing nutrient conditions during gut microbiota dysbiosis[204,205]. Better understanding of its metabolism alone and in response to changes in the community may provide additional strategies for prevention and treatment of *C. difficile* infection. *C. difficile* has an unusually high number of cobamide-dependent metabolisms encoded in its genome[221], and its predicted ALA salvaging phenotype suggests a dependence on host, diet, or microbiota-derived cobamides and cobamide precursors. This study showed that *C. difficile* 630 Δ*erm* is an ALA-salvaging bacterium, and in the absence of exogenous lower ligand precursors, produces the cobamide pseudocobalamin when provided with the cobamide precursors ALA or Cbi. It is able to use its own "native" cobamide to carry out methionine synthesis through MetH and deoxyribonucleotide synthesis through NrdJ, but can also use other cobamides for both processes or bypass the cobamide-

dependent enzyme. Specifically, addition of methionine could substitute for MetH, and the cobamide-independent class III RNR can be used by wild type *C. difficile* in the absence of cobamides. This redundancy makes *C. difficile* metabolically versatile in that it can use cobamides but does not have an obligate cobamide requirement.

We tested a suite of nine cobamides to support the growth of *C. difficile* in the absence of exogenous methionine, many of which can be found in the human gut. While there were minor differences in concentration required for robust growth yield, *C. difficile's* MetH appears to be promiscuous in its cobamide selectivity. The cobamide selectivity of MetH orthologs from other organisms has previously been studied with a smaller panel of cobamides, and the amount of selectivity varied. These studies show that a subset of the cobamides could be used by each MetH homolog, however, lack of comprehensive testing of cobamide use in many cases makes it difficult to determine which cobamides are able to support catalysis. While eukaryotic algae showed no growth with pseudocobalamin under MetH-dependent conditions[11], a *Salmonella enterica metE* mutant was able to use cobalamin, pseudocobalamin, [Phe]Cba, and [Cre]Cba to grow without added methionine, although other cobamides were not tested[19,64]. *Bacteroides thetaiotaomicron* can use both benzimidazolyl cobamides and purinyl cobamides for MetH-dependent growth, but not phenolyl cobamides[9]. Other MetH homologs have been shown to bind cobalamin in a base-off, His-on configuration[89]. In *Spirulina platensis,* MetH bound pseudocobalamin with a higher affinity than cobalamin[22]. Further experimental testing of cobamide selectivity of MetH homologs will help determine which residues are important for cobamide selectivity. In contrast to MetH, *C. difficile*'s NrdJ was more selective, which was expected because of its requirement for the base-on configuration[244,245]. Growth assays in a *Sinorhizobium meliloti* cobamide-biosynthesis mutant found that it could not grow with [Cre]Cba and had reduced growth with pseudocobalamin relative to its native cobamide, cobalamin, under conditions where RNR was the only cobamide-dependent enzyme required for growth[74]. Our results in *C. difficile* show the same inability to use [Cre]Cba. Comprehensive testing of cobamide selectivity of other NrdJ homologs will provide additional context for the promiscuity or selectivity of *C. difficile* NrdJ.

*C. difficile*'s versatility in cobamide use is notable given the diversity of cobamides likely available to it in the gut. Seven different cobamides are found in human feces in quantities of tens to thousands of ng per g stool[187]. *C. difficile* thus potentially has access to many cobamide types. [2-MeAde]Cba and [Cre]Cba were typically the most abundant cobamides in the individuals sampled[187], and both are able to support *C. difficile's* growth with MetH (Fig. 3).

Not only do our data show that *C. difficile* is able to use multiple cobamides to support its metabolism, but it can also use the early precursor ALA to produce a cobamide and thus does not strictly rely on cobamide or Cbi import. This could be important for RT027 strains that may lack the ability to take up complete cobamides to conduct cobamide-dependent metabolism. ALA, and possibly other early cobamide precursors, could be provided by either the microbiota, or by the host, both of which produce ALA as a precursor to heme and other tetrapyrroles. ALA concentrations in the human gut have not been measured. The late intermediate Cbi was found to be present at tens of ng per gram of feces and up to 600 ng/g in patients receiving milligram daily doses of cobalamin[187]. How the cobamide and cobamide precursor profile in the gut changes during *C. difficile* infection is unknown, but since much of the cobamide content in the lower GI tract is produced by resident gut microbes[125,246], it is interesting to speculate that the corrinoid profile may change during dysbiosis.

In addition to scavenging available complete cobamides, *C. difficile* is also able to attach non-native lower ligands to cobamides it produces from cobamide precursors. Benzimidazolyl cobamides did not provide a growth advantage over pseudocobalamin, the cobamide that *C. difficile* makes without added lower ligand, in the two metabolisms we investigated in this study. Benzimidazoles can be found in many environments including host associated fluids such as rumen fluid and termite guts[95], but the amount of cobamide lower ligands has not been determined in the human gut.

*C. difficile* 630 Δ*erm* has flexibility to acquire cobamides and cobamide precursors to carry out its cobamide dependent metabolisms. What form of cobamide or cobamide precursor it prefers to use in the host remains to be discovered. Evidence from transcriptomics is mixed on whether genes encoding cobamide-dependent enzymes or the cobamide biosynthesis pathway are expressed during infection, but Fletcher *et al*.[224] reported that *hemB*, which encodes the enzyme that converts ALA to the cobamide intermediate porphobilinogen, was among the most highly expressed genes in the mouse gut[212,224,247–249]. *C. difficile* infection studies use a variety of animal models with mono-associations of *C. difficile*, gnotobiotic animals with a consortium of commensal strains, and antibiotic-treated conventional animals. Since both diet and the microbiota contribute to the cobamide profile in the gut[125,187,246], the availability of cobamides may vary significantly in these systems and affect the expression and use of cobamide biosynthesis and dependent pathways by *C. difficile*. Additionally, these studies use different *C. difficile* strains, and as noted above, those in the RT027 clade may not have a functional BtuFCD transporter, and therefore would be defective in taking up exogenous cobamides, which could influence cobamide-dependent metabolism.

*C. difficile* strains also have the option to forgo cobamides altogether in the case of methionine synthesis by taking up methionine from the environment, or by using a different RNR. However, some cobamide-dependent pathways do not have other alternatives, including the Wood-Ljungdahl and the ethanolamine catabolism pathways, of which there is some evidence about their importance to *C. difficile* in providing an electron sink[223,224] and modulating virulence[222], respectively. The cobamide availability to different *C. difficile* strains may influence which cobamide-dependent pathways are accessible and could affect virulence of the strain, but further *in vivo* studies are needed to determine to what extent cobamides and cobamide-dependent pathways are important to colonization and disease.

# Chapter 4

## Phylogenetic analysis of cobamide biosynthesis and dependence in the Clostridiales

**Abstract**
Cobamide cofactors are only produced by a subset of bacteria that require them. The heterogeneous distribution of the cobamide biosynthesis pathway across the domain Bacteria suggests multiple independent gains and losses of the pathway, and a few examples of known horizontal gene transfer and independent loss have been observed. Additionally, there are many genomes that contain partial cobamide biosynthesis pathways, suggesting gain or loss of individual genes in these lineages. To further explore the evolutionary history of cobamide biosynthesis and dependence genes, I analyzed their presence and absence in the order Clostridiales in the phylum Firmicutes to predict gain and loss events within the lineage. I observed that loss of some or the entire cobamide biosynthetic pathway has occurred multiple times in the Clostridiales. In some cases, the presence of cobamide-dependent genes in the clade relative to sister groups may explain the retention of cobamide biosynthesis in this clade. Based on gene neighborhoods of the corrin ring synthesis genes in a subset of species, the gene order of the cluster is largely conserved. However, tetrapyrrole precursor genes, cobalt uptake genes, and some corrin ring methyltransferases may rearrange frequently. These observations provide a starting point for more in depth study of the evolutionary history of cobamide biosynthesis genes in the Clostridiales.

### Introduction

Cobamides are a class of enzyme cofactors including vitamin $B_{12}$ that function in diverse metabolisms. While over 80% of bacteria have cobamide-dependent enzymes, fewer than 40% encode the complete pathway for *de novo* synthesis[221]. As described in Chapter 1, the biosynthetic pathway requires approximately 30 genes, with the first 5-7 genes being shared with other tetrapyrrole synthesis pathways including heme, chlorophyll, and siroheme[23]. The corrin ring assembly part of the pathway requires 10-12 genes and has two distinct branches, the oxygen-requiring late cobalt insertion pathway (aerobic), and the oxygen-sensitive early cobalt insertion pathway (anaerobic)[46]. Genomes tend to have only one of these alternative pathways for synthesizing the corrin ring, and in this chapter the analysis will focus on the anaerobic branch. The nucleotide loop assembly and lower ligand attachment genes are common to both branches. Variation in lower ligand structure is known to be encoded by certain synthesis and attachment genes[64–69], but the lower ligand identity of cobamides produced is not a focus of this chapter.

We and others have observed that the biosynthetic pathway is heterogeneously distributed across prokaryotes, with some clades being enriched for and some depleted in cobamide biosynthesis capability. For example, fewer than 1% of sequenced species in the phylum Bacteroidetes have the complete cobamide biosynthesis pathway, while 30% of Firmicutes have the pathway[221]. In archaea, the phylum Thaumarchaeota's basal clade lacks cobamide biosynthesis, but the derived ammonia-oxidizing clades contain the complete biosynthesis pathway, suggesting an acquisition of the pathway[250]. This begs the question: how does cobamide biosynthesis evolve to be present in some clades and not in others?

There is evidence that the cobamide biosynthetic pathway and, in some cases, cobamide-dependent enzymes and pathways were horizontally transferred. For example, comparative genomics of *Lactobacillus reuteri* JCM 1112T shows a gain of the cobalamin (vitamin $B_{12}$) biosynthesis and propanediol utilization pathways[251]. The authors of this study compare the *L. reuteri* pathway to that of other diverse bacteria with both pathways, and find that the relative positioning of the genes varies, suggesting the possibility that acquisition of the propanediol and cobamide pathways may have been independent events. Other examples of gain of cobamide biosynthesis by probable horizontal gene transfer include *Listeria monocytogenes* and *Listeria innocua*[251,252], *Lactobacillus rossiae*[168], and *Salmonella enterica*[203].

To complicate the evolutionary analysis of the cobamide biosynthesis pathway, many bacteria have some, but not all cobamide biosynthesis genes. In fact, up to 17% of bacteria have the nucleotide loop assembly genes, and in some cases, the corrin ring synthesis genes required to salvage cobamide intermediates including tetrapyrrole precursors and the late intermediate cobinamide[9,193,221]. Since complete and partial cobamide biosynthesis pathways were found in genomes of the same clade in several cases, we hypothesized that there have been multiple independent loss events of cobamide biosynthesis genes, and this is supported by analysis of specific clades of interest. In the *Mycobacterium* complex, most species retain the ability to produce cobamides, but *M. tuberculosis* has a truncated *cobF* gene, rendering it a cobamide auxotroph[253]. In the genus *Yersinia*, *Y. pseudotuberculosis* appears to have lost the cobamide biosynthesis pathway entirely[254]. These are just two examples of loss of cobamide biosynthesis genes in diverse bacteria[203].

Here I focus on the order Clostridiales in the phylum Firmicutes and examine the presence and absence of cobamide biosynthesis and dependent genes in these genomes in the context of their phylogenetic relationships. I use this to create hypotheses about gain and loss of the cobamide biosynthetic genes, including when they occurred and speculation on the process of relaxed selection based on the predicted cobamide dependence phenotype. This order has a large number of genomes with partial cobamide biosynthesis pathways, and also a relatively large number of species that have been experimentally confirmed to produce cobamides. This chapter provides observations of phylogenetically associated cobamide biosynthesis content that may lead to additional studies of the mechanism of gene gain and loss in these clades.

## Materials and methods

### Phylogenetic trees

Sequences were downloaded from the Joint Genome Institute's Integrated Microbial Genomes (IMG) database[234] in April 2018, and April-June 2019. Genomes were selected by taxonomy and placed in the IMG genome cart. If possible, genomes used were the same used in Chapter 2, those classified as "finished" or "permanent draft" on IMG and meeting same completeness threshold using the 55 single copy genes in Chapter 2. The genomes were not all completely circular, many consisted of multiple scaffolds. This genome cart was searched using the gene search tool by "gene product name" using "16S ribosomal" and "bacterial SSU" and the resulting gene cart was downloaded. For these 16S rRNA sequences, sequences less than 600 bp and greater than 1800 bp were discarded. Since many genomes have more than one 16S rRNA sequence, a single sequence was chosen to represent each genome at random using custom python code. For the Clostridiales tree, vsearch "cluster_fast" was run to choose one sequence at 97% identity using centroid clustering[255]. The sequences were aligned using the SINA

aligner webserver [256]. Then columns of the alignment with 95% or greater gaps were trimmed using trimAl[238]. Finally columns at the ends were manually trimmed. The alignment was used as input for RAxML v8.2.12[239] maximum likelihood phylogeny using the GTRCAT model on the CIPRES webserver[240]. The tree was visualized in iTOL[241] and Adobe Illustrator.

To annotate the trees with cobamide biosynthesis and dependence genes, the dataset of Shelton *et al.*[221] was used to assess the presence and absence of genes. For genomes absent from that dataset, the cobamide biosynthesis and dependence functions were used in a function profile against the genomes and processed as described in Chapter 2.

**Synteny analysis**

Synteny analysis was performed using IMG's gene neighborhood tool, using a "best hits" search. For the *Lachnospiraceae* (Fig. 4), *cbiD* from *Acetobacterium woodii* was used the query. For the *Peptostreptococcaceae* (Fig. 5), *cbiD* of *Clostridioides (Clostridium) difficile* was used and returned scaffolds containing hits to *cbiD.* Images of the scaffolds were saved, and manually annotated in Adobe Illustrator with the annotations from the gene pages in IMG, found via links in the gene neighborhood tool.

**Results**

To study the phylogenetic relationship of cobamide biosynthesis genes, I mapped the anaerobic cobamide biosynthesis and dependence annotations from Shelton *et al.* 2019 (Chapter 2) onto 16S rRNA species trees. These trees were composed of a single, randomly chosen 16S rRNA gene sequence as a representative for each genome. My tree of representative Firmicutes species (Fig. 1) agrees well with the analysis of Zhang and Lu[257], which used the same strains plus some additional species. Additionally, using the annotation-based approach of Chapter 2 to find the cobamide biosynthesis and dependence genes may lead to some artifacts, and I will explore some specific cases further.

The cobamide-dependent genes examined included methionine synthase, ribonucleotide reductase (RNR), methylmalonyl-CoA mutase, glutamate mutase, ethanolamine ammonia lyase, 5,6-beta-lysine aminomutase, D-ornithine 4,5-aminomutase, epoxyqueuosine reductase, reductive dehalogenases, the acetogenesis corrinoid-iron-sulfur protein, and methyltransferases proteins. Many of the carbon and nitrogen catabolism genes were characterized in members of class Clostridia, and are expected to be found in genomes of Clostridiales[156,258–261]. I did not include some known cobamide-dependent families in this analysis that are included in Chapter 2 since these were not identified with a simple annotation search. However, this approach allows an overview of cobamide biosynthesis with a more specific phylogenetic signal than was explored in Chapter 2[221].

**Representative Firmicutes show multiple instances of cobamide biosynthesis gain or loss**

To gain an overview of cobamide biosynthesis in the phylum Firmicutes, a 16S rRNA phylogeny of 86 representative strains and one outgroup[257] was annotated with the anaerobic cobamide biosynthesis genes and some cobamide-dependent genes[221] (Fig. 1). Regardless of whether or not the last common ancestor had a complete cobamide biosynthesis pathway, analysis of this tree suggests that gains and/or losses of the complete pathway have occurred multiple times in the Firmicutes lineage (Fig. 1). When starting with the assumption that the ancestor did not have the complete pathway (Fig. 1, red symbols), I hypothesize that there

were thirteen independent gains of the pathway, which then was lost in some lineages nine independent times. If the ancestor had the complete pathway (Fig. 1, cyan symbols), there were eighteen independent losses of the pathway and one gain in the known case of *L. monocytogenes*[251,252]. The reference genomes sample one genome per genus, and including additional genomes in the phylogenetic tree may reveal additional gain and loss events. As observed with the transposon insertion in *btuFCD* cobamide transporter operon in some *Clostridioides difficile* strains in Chapter 3, strain-level differences in cobamide biology do occur, which would not be reflected in the reference genome tree.. Gene trees of the cobamide biosynthesis genes may be able to provide additional support for the proposed gain and loss events. For example, if the gene trees largely match the species tree, these better support the hypothesis that the ancestor had a complete biosynthetic pathway that was later lost, but incongruence between the trees may suggest horizontal gene transfer events. It is possible, but improbable, for instances of horizontal gene transfer to give the same signal in the gene trees.

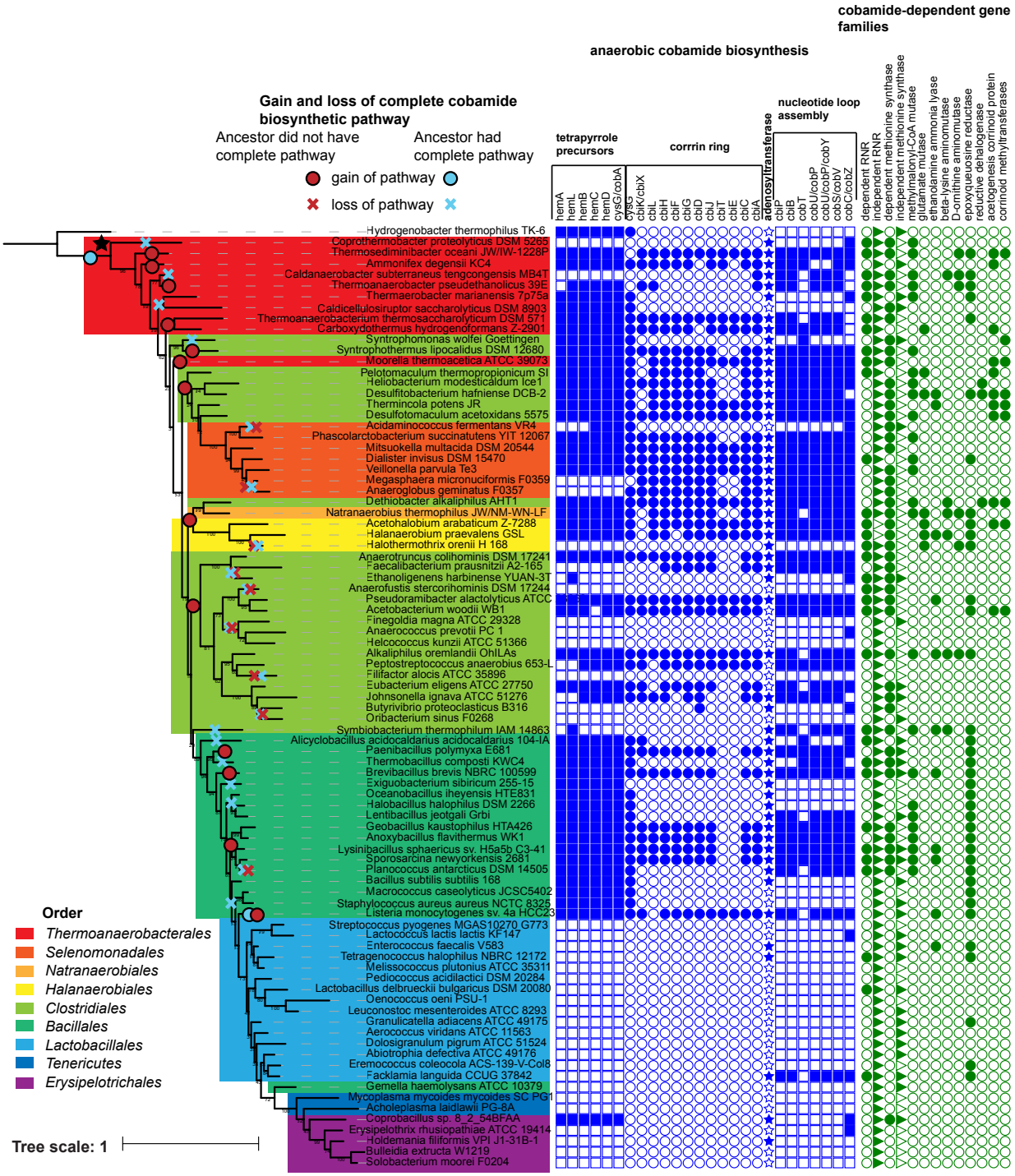



**Figure 1**. **Anaerobic cobamide biosynthesis and cobamide-dependent genes in representative Firmicutes.** A 16S tree of a subset of representative Firmicutes and outgroup species from Zhang and Lu[257] was made using RAxML, bootstraps of 100 replicates are shown. Annotations for cobamide biosynthesis and dependent genes from Chapter 2[221] are mapped onto the tree. Filled shapes indicate the presence of the annotation. For cobamide biosynthesis, the different shapes are used to visually separate parts of the pathway. Green circles and triangles represent cobamide-dependent and -independent alternative genes. The star indicates where *Coprothermobacter proteolyticus* has been

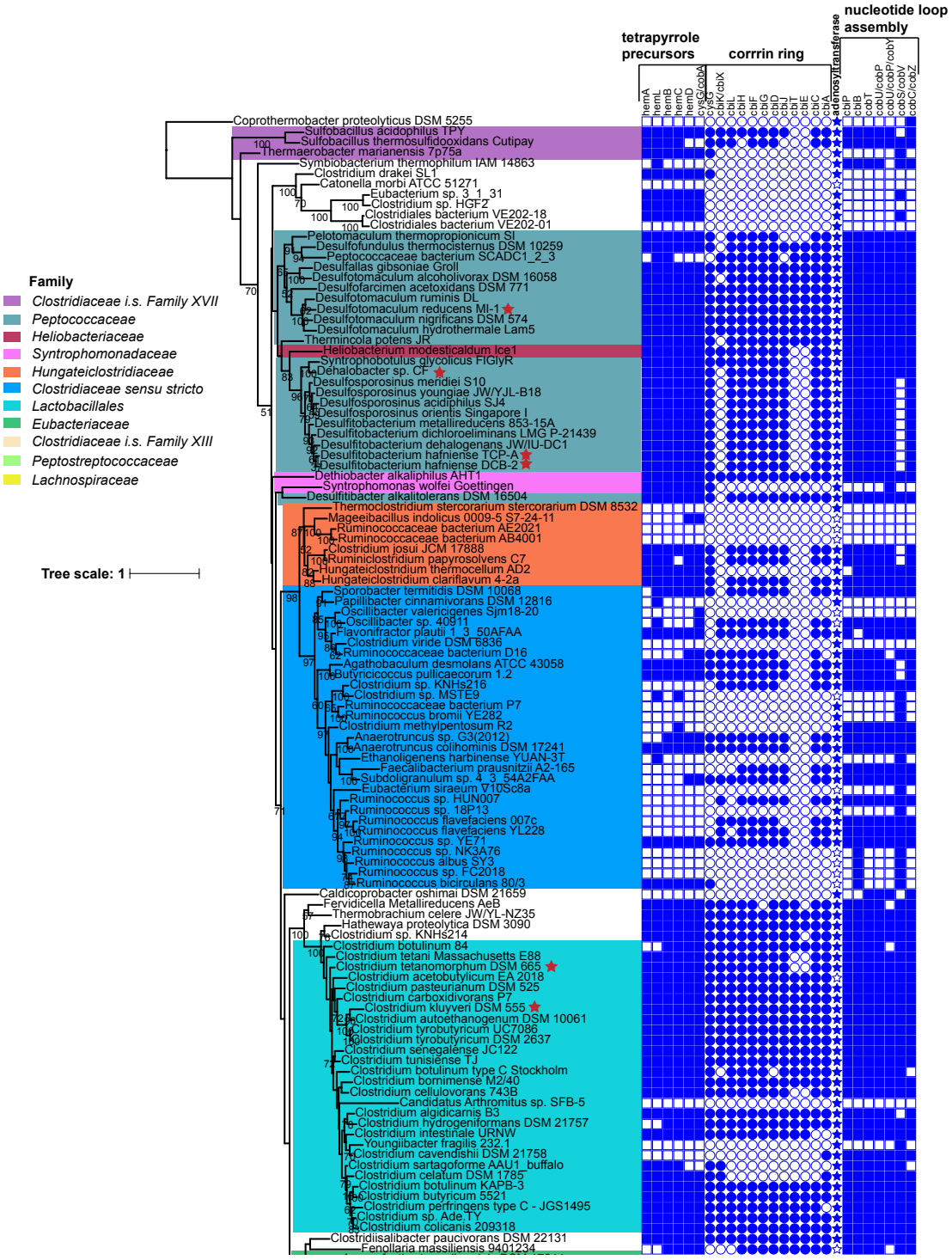
since transferred to a new phylum[262]. Branch length is number of substitutions per site. Colored symbols on the branches indicate the most parsimonious putative gains and losses of the cobamide biosynthetic pathway. Red symbols represent the hypothesis that the last common ancestor did not have a complete cobamide biosynthesis pathway, and cyan is the hypothesis where it did.

**Cobamide biosynthesis genes have been lost independently in the order Clostridiales**

To explore species-level differences in cobamide biosynthesis, the order Clostridiales was chosen for further analysis. A large number of organisms from this clade relative to other Firmicutes orders have been shown experimentally to produce cobamides, and some have been shown to be able to salvage 5-aminolevulinic acid (ALA) and produce a cobamide (Table 1). Because genome annotation only provides the potential for performing certain metabolisms, experimental information confirming the phenotype of closely related strains can improve confidence in genomic prediction. In a 16S rRNA species tree of a subset of the Clostridiales (Fig. 2), there are multiple apparent complete or partial losses of the cobamide biosynthesis pathway.

**Table 1. List of Clostridiales experimentally confirmed to produce cobamides or salvage ALA**

| Genome Name | Phenotype | Reference |
|---|---|---|
| *Desulfitobacterium hafniense* | cobamide producer | 69,87,92 |
| *Dehalobacter sp. CF* | cobamide producer | 158 |
| *Desulfotomaculum reducens* | cobamide producer | Chapter 2[221] |
| *Clostridium tetanomorphium* | cobamide producer | 156 |
| *Clostridium kluyveri* | cobamide producer | Chapter 2[221] |
| *Acetobacterium woodii* | cobamide producer | 151 |
| *Clostridioides (Clostridium) difficile* | ALA salvager | Chapter 3 |
| *Lachnoclostridium (Clostridium) phytofermentans* | cobamide producer | Chapter 2[221] |
| *Anaerobutyricum (Eubacterium) hallii* | cobamide producer | 163 |
| *Blautia hydrogenotrophica* | cobamide producer | Chapter 2[221] |
| *Lachnoclostridium (Clostridium) scindens* | ALA salvager | Chapter 2[221] |


tetrapyrrole precursors
corrrin ring
nucleotide loop assembly
adenosyltransferase
hemA
hemL
hemB
hemC
hemD
cysG/cobA
cbiK/cbiX
cbiL
cbiH
cbiF
cbiG
cbiD
cbiJ
cbiE
cbiC
cbiA
cbiP
cbiB
cobT
cobU/cobP
cobU/cobP/cobY
cobS/cobV
cobC/cobZ
Family
Clostridiaceae i.s. Family XVII
Peptococcaceae
Heliobacteriaceae
Syntrophomonadaceae
Hungateiclostridiaceae
Clostridiaceae sensu stricto
Lactobacillales
Eubacteriaceae
Clostridiaceae i.s. Family XIII
Peptostreptococcaceae
Lachnospiraceae
Tree scale: 1
Coprothermobacter proteolyticus DSM 5255
Sulfobacillus acidophilus TPY
Sulfobacillus thermosulfidooxidans Cutipay
Thermaerobacter marianensis 7p75a
Symbiobacterium thermophilum IAM 14863
Clostridium drakei SL1
Catonella morbi ATCC 51271
Eubacterium sp. 3_1_31
Clostridium sp. HGF2
Clostridiales bacterium VE202-18
Clostridiales bacterium VE202-01
Pelotomaculum thermopropionicum SI
Desulfofundulus thermocisternus DSM 10259
Peptococcaceae bacterium SCADC1_2_3
Desulfallas gibsoniae Groll
Desulfotomaculum alcoholivorax DSM 16058
Desulfofarcimen acetoxidans DSM 771
Desulfotomaculum ruminis DL
Desulfotomaculum reducens MI-1
Desulfotomaculum nigrificans DSM 574
Desulfotomaculum hydrothermale Lam5
Thermincola potens JR
Heliobacterium modesticaldum Ice1
Syntrophobotulus glycolicus FlGlyR
Dehalobacter sp. CF
Desulfosporosinus meridiei S10
Desulfosporosinus youngiae JW/YJL-B18
Desulfosporosinus acidiphilus SJ4
Desulfosporosinus orientis Singapore I
Desulfitobacterium metallireducens 853-15A
Desulfitobacterium dichloroeliminans LMG P-21439
Desulfitobacterium dehalogenans JW/IU-DC1
Desulfitobacterium hafniense TCP-A
Desulfitobacterium hafniense DCB-2
Dethiobacter alkaliphilus AHT1
Syntrophomonas wolfei Goettingen
Desulfitibacter alkalitolerans DSM 16504
Thermoclostridium stercorarium stercorarium DSM 8532
Mageeibacillus indolicus 0009-5 S7-24-11
Ruminococcaceae bacterium AE2021
Ruminococcaceae bacterium AB4001
Clostridium josui JCM 17888
Ruminiclostridium papyrosolvens C7
Hungateiclostridium thermocellum AD2
Hungateiclostridium clariflavum 4-2a
Sporobacter termitidis DSM 10068
Papillibacter cinnamivorans DSM 12816
Oscillibacter valericigenes Sjm18-20
Oscillibacter sp. 40911
Flavonifractor plautii 1_3_50AFAA
Clostridium viride DSM 6836
Ruminococcaceae bacterium D16
Agathobaculum desmolans ATCC 43058
Butyricicoccus pullicaecorum 1.2
Clostridium sp. KNHs216
Clostridium sp. MSTE9
Ruminococcaceae bacterium P7
Ruminococcus bromii YE282
Clostridium methylpentosum R2
Anaerotruncus sp. G3(2012)
Anaerotruncus colihominis DSM 17241
Ethanoligenens harbinense YUAN-3T
Faecalibacterium prausnitzii A2-165
Subdoligranulum sp. 4_3_54A2FAA
Eubacterium siraeum V10Sc8a
Ruminococcus sp. HUN007
Ruminococcus sp. 18P13
Ruminococcus flavefaciens 007c
Ruminococcus flavefaciens YL228
Ruminococcus sp. YE71
Ruminococcus sp. NK3A76
Ruminococcus albus SY3
Ruminococcus sp. FC2018
Ruminococcus bicirculans 80/3
Caldicoprobacter oshimai DSM 21659
Fervidicella Metallireducens AeB
Thermobrachium celere JW/YL-NZ35
Hathewaya proteolytica DSM 3090
Clostridium sp. KNHs214
Clostridium botulinum 84
Clostridium tetani Massachusetts E88
Clostridium tetanomorphum DSM 665
Clostridium acetobutylicum EA 2018
Clostridium pasteurianum DSM 525
Clostridium carboxidivorans P7
Clostridium kluyveri DSM 555
Clostridium autoethanogenum DSM 10061
Clostridium tyrobutyricum UC7086
Clostridium tyrobutyricum DSM 2637
Clostridium senegalense JC122
Clostridium tunisiense TJ
Clostridium botulinum type C Stockholm
Clostridium bornimense M2/40
Clostridium cellulovorans 743B
Candidatus Arthromitus sp. SFB-5
Clostridium algidicarnis B3
Clostridium hydrogeniformans DSM 21757
Clostridium intestinale URNW
Youngiibacter fragilis 232.1
Clostridium cavendishii DSM 21758
Clostridium sartagoforme AAU1_buffalo
Clostridium celatum DSM 1785
Clostridium botulinum KAPB-3
Clostridium butyricum 5521
Clostridium perfringens type C - JGS1495
Clostridium sp. Ade.TY
Clostridium colicanis 209318
Clostridiisalibacter paucivorans DSM 22131
Fenollaria massiliensis 9401234

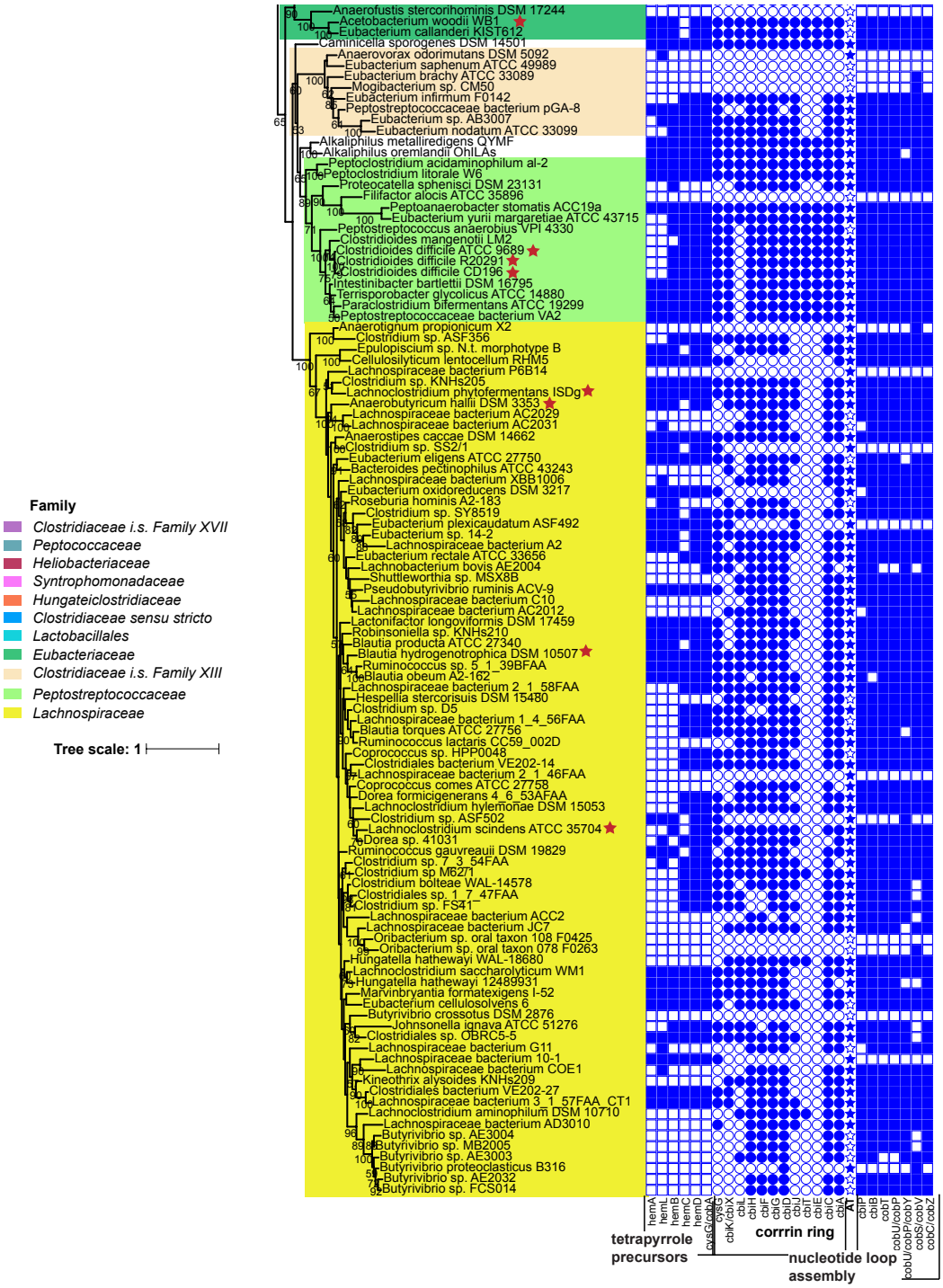
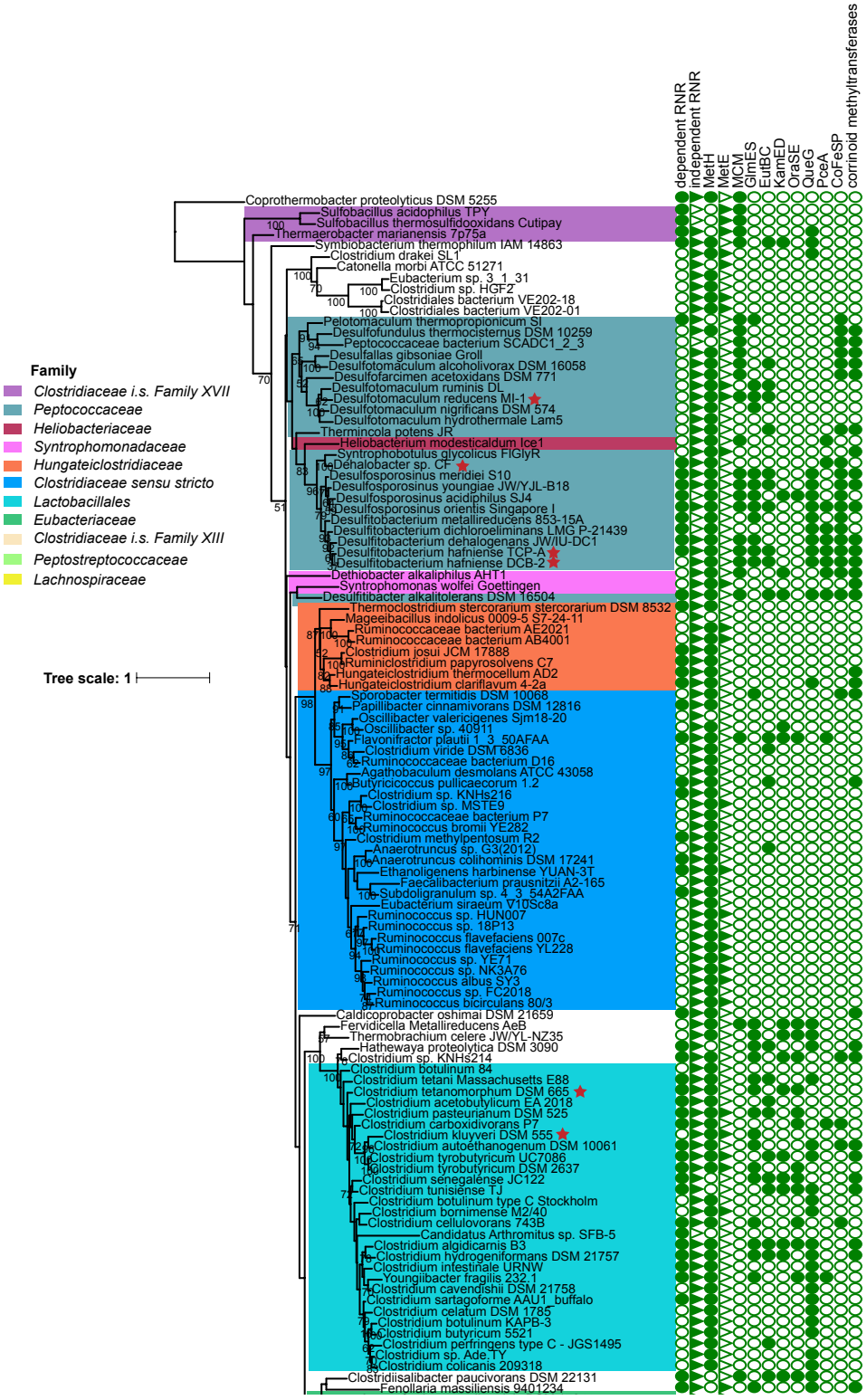



**Figure 2**. **Anaerobic cobamide biosynthesis in the Clostridiales.** A 16S rRNA tree of a subset of Clostridiales genomes on IMG was made using RAxML, with 100 bootstraps. Only branches with bootstrap values over 50 are labeled. There are 217 Clostridiales plus one outgroup (*Coprothermobacter proteolyticus*). Annotations for cobamide biosynthesis as in Fig. 1 are mapped

onto the tree. Filled shapes indicate the presence of an annotation. Red stars indicate species that have been experimentally validated to produce cobamides or are cobamide precursor-salvaging bacteria. Branch length is number of substitutions per site. The colored boxes under the genome names are the families to which they belong. No color indicates that the classification was unclear to the author. Tree is split over two pages. Abbreviations: AT, adenosyltransferase, i.s.: *incertae sedis*.


dependent RNR
independent RNR
MetH
MetE
MCM
GlmES
EutBC
KamED
OraSE
QueG
PceA
CoFeSP
corrinoid methyltransferases
Family
Clostridiaceae i.s. Family XVII
Peptococcaceae
Heliobacteriaceae
Syntrophomonadaceae
Hungateiclostridiaceae
Clostridiaceae sensu stricto
Lactobacillales
Eubacteriaceae
Clostridiaceae i.s. Family XIII
Peptostreptococcaceae
Lachnospiraceae
Tree scale: 1
Coprothermobacter proteolyticus DSM 5255
Sulfobacillus acidophilus TPY
Sulfobacillus thermosulfidooxidans Cutipay
Thermaerobacter marianensis 7p75a
Symbiobacterium thermophilum IAM 14863
Clostridium drakei SL1
Catonella morbi ATCC 51271
Eubacterium sp. 3_1_31
Clostridium sp. HGF2
Clostridiales bacterium VE202-18
Clostridiales bacterium VE202-01
Pelotomaculum thermopropionicum SI
Desulfofundulus thermocisternus DSM 10259
Peptococcaceae bacterium SCADC1_2_3
Desulfallas gibsoniae Groll
Desulfotomaculum alcoholivorax DSM 16058
Desulfofarcimen acetoxidans DSM 771
Desulfotomaculum ruminis DL
Desulfotomaculum reducens MI-1
Desulfotomaculum nigrificans DSM 574
Desulfotomaculum hydrothermale Lam5
Thermincola potens JR
Heliobacterium modesticaldum Ice1
Syntrophobotulus glycolicus FlGlyR
Dehalobacter sp. CF
Desulfosporosinus meridiei S10
Desulfosporosinus youngiae JW/YJL-B18
Desulfosporosinus acidiphilus SJ4
Desulfosporosinus orientis Singapore I
Desulfitobacterium metallireducens 853-15A
Desulfitobacterium dichloroeliminans LMG P-21439
Desulfitobacterium dehalogenans JW/IU-DC1
Desulfitobacterium hafniense TCP-A
Desulfitobacterium hafniense DCB-2
Dethiobacter alkaliphilus AHT1
Syntrophomonas wolfei Goettingen
Desulfitibacter alkalitolerans DSM 16504
Thermoclostridium stercorarium stercorarium DSM 8532
Mageeibacillus indolicus 0009-5 S7-24-11
Ruminococcaceae bacterium AE2021
Ruminococcaceae bacterium AB4001
Clostridium josui JCM 17888
Ruminiclostridium papyrosolvens C7
Hungateiclostridium thermocellum AD2
Hungateiclostridium clariflavum 4-2a
Sporobacter termitidis DSM 10068
Papillibacter cinnamivorans DSM 12816
Oscillibacter valericigenes Sjm18-20
Oscillibacter sp. 40911
Flavonifractor plautii 1_3_50AFAA
Clostridium viride DSM 6836
Ruminococcaceae bacterium D16
Agathobaculum desmolans ATCC 43058
Butyricicoccus pullicaecorum 1.2
Clostridium sp. KNHs216
Clostridium sp. MSTE9
Ruminococcaceae bacterium P7
Ruminococcus bromii YE282
Clostridium methylpentosum R2
Anaerotruncus sp. G3(2012)
Anaerotruncus colihominis DSM 17241
Ethanoligenens harbinense YUAN-3T
Faecalibacterium prausnitzii A2-165
Subdoligranulum sp. 4_3_54A2FAA
Eubacterium siraeum V10Sc8a
Ruminococcus sp. HUN007
Ruminococcus sp. 18P13
Ruminococcus flavefaciens 007c
Ruminococcus flavefaciens YL228
Ruminococcus sp. YE71
Ruminococcus sp. NK3A76
Ruminococcus albus SY3
Ruminococcus sp. FC2018
Ruminococcus bicirculans 80/3
Caldicoprobacter oshimai DSM 21659
Fervidicella Metallireducens AeB
Thermobrachium celere JW/YL-NZ35
Hathewaya proteolytica DSM 3090
Clostridium sp. KNHs214
Clostridium botulinum 84
Clostridium tetani Massachusetts E88
Clostridium tetanomorphum DSM 665
Clostridium acetobutylicum EA 2018
Clostridium pasteurianum DSM 525
Clostridium carboxidivorans P7
Clostridium kluyveri DSM 555
Clostridium autoethanogenum DSM 10061
Clostridium tyrobutyricum UC7086
Clostridium tyrobutyricum DSM 2637
Clostridium senegalense JC122
Clostridium tunisiense TJ
Clostridium botulinum type C Stockholm
Clostridium bornimense M2/40
Clostridium cellulovorans 743B
Candidatus Arthromitus sp. SFB-5
Clostridium algidicarnis B3
Clostridium hydrogeniformans DSM 21757
Clostridium intestinale URNW
Youngiibacter fragilis 232.1
Clostridium cavendishii DSM 21758
Clostridium sartagoforme AAU1_buffalo
Clostridium celatum DSM 1785
Clostridium botulinum KAPB-3
Clostridium butyricum 5521
Clostridium perfringens type C - JGS1495
Clostridium sp. Ade.TY
Clostridium colicanis 209318
Clostridiisalibacter paucivorans DSM 22131
Fenollaria massiliensis 9401234

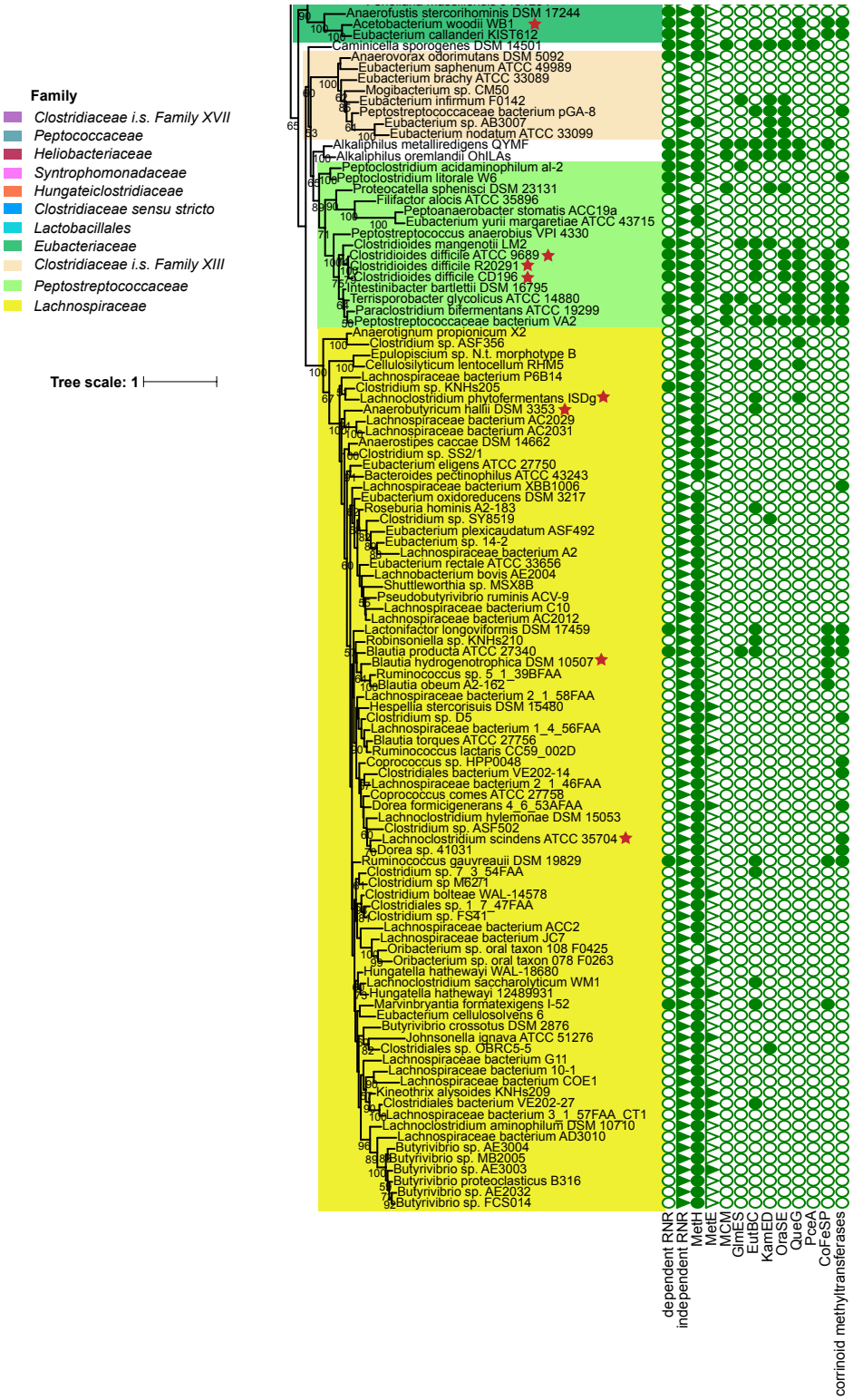



**Figure 3. Cobamide dependent gene families in the Clostridiales.** The same 16S tree in Fig. 2 is annotated with the cobamide dependent gene families. A filled shape indicates the presence of one or more genes in the family as in Fig. 1. Abbreviations: RNR, ribonucleotide reductase; MetH, cobamide-

dependent methionine synthase; MetE, cobamide-independent methionine synthase; MCM, methylmalonyl-CoA mutase; GlmES, glutamate mutase; EutBC, ethanolamine ammonia lyase; KamED, 5,6-beta-lysine aminomutase; OraSE, D-ornithine aminomutase; QueG, epoxyqueuosine reductase; PceA, reductive dehalogenases; CoFeSP, acetogenesis corrinoid protein, i.s: *incertae sedis.* Tree is split over two pages.

An example of a clade that consists of nearly all cobamide producers is the family *Clostridiaceae sensu stricto* (Fig. 2,3, teal), some of which have been experimentally confirmed to produce cobamides. Members of this clade have many cobamide-dependent enzyme families. A few species have lost the genetic ability to produce cobamides, including *Clostridium celatum*, *Clostridium sartagoforme, Clostridium cavendishii,* and *Youngiibacter fragilis*. There is not an obvious trend of cobamide-dependence in these species relative to the rest of the clade (Fig. 3). *C. cavendishii* and *Y. fragilis* are sister to each other, but the other genomes with losses are more distantly related to each other, suggesting multiple independent losses of the pathway. The genome of Candidatus *Arthromitus* sp. SFB-5 consists of many small scaffolds, so it is difficult to assess the presence and absence of particular genes. Additional interesting exceptions to containing the complete cobamide biosynthesis pathway in this clade are *Clostridium botulinum* and *Clostridium hydrogeniformans*, both of which are missing the first two genes to make the early precursor 5-aminolevulinic acid (ALA). Since they are distantly related, it suggests that there have been multiple *hemAL* loss events in the clade. A strain of *C. sporogenes*, closely related to *C. botulinum*, has also been confirmed to be an ALA salvager (Chapter 2)[221].

The family *Lachnospiraceae* is a group of anaerobes, often host-associated[263,264], and many members of this clade exhibit partial cobamide biosynthesis pathways (Fig. 2, yellow), suggesting multiple independent gains or losses of genes in the pathway. For many of these species, the only cobamide-dependent enzyme family is the methionine synthase, MetH, which could make cobamide biosynthesis dispensable if exogenous methionine or cobamide is available (Fig. 3, yellow).

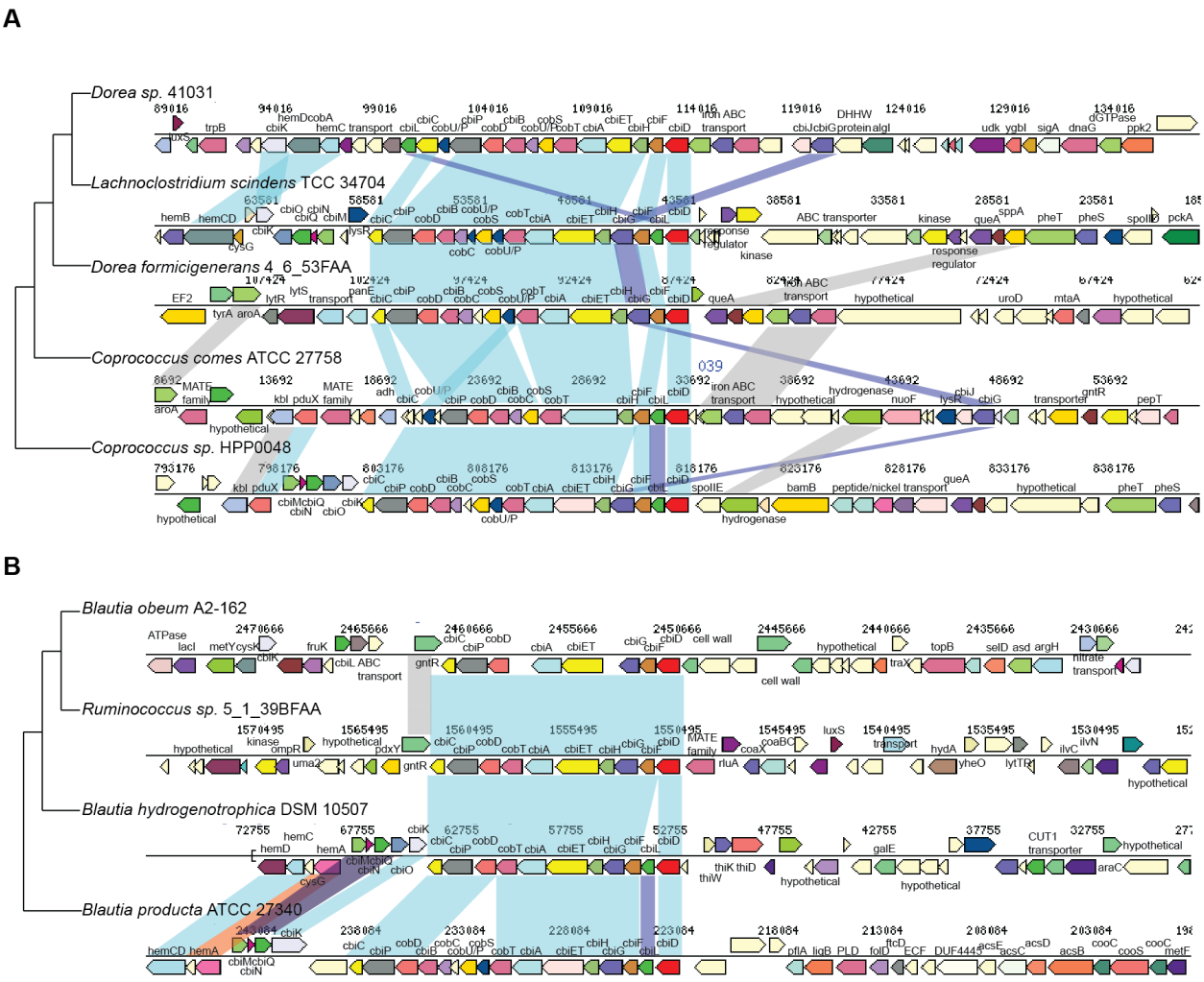



**Figure 4. Synteny of the cobamide biosynthesis gene cluster in the *Lachnospiraceae*.** The *cbiD* gene of *Acetobacterium woodii,* a confirmed cobamide producer, was used as the query for the IMG gene neighborhood best hits search. The resulting homolog hits were then arranged and labeled based on their annotation. Genes of the same color are homologs, white are pseudogenes, light yellow are genes without COG assignment. Colored pairwise connections between genomes indicate regions of homology. Blue is *cysGhemDcobAcbiKLJHGFTEDCpduXcobDcbiBAP*, green is *cobCSUT*, dark blue is *cbiMNQO*, orange is *hemAL*, gray are shared loci. Brackets indicate the end of the scaffold. All genomes are found in the tree in Fig 2 and cladograms are from that analysis. **A.** *Dorea/Coprococcus* group of precursor auxotrophs. **B.** *Blautia* group of cobamide producers.

To further explore gains and losses of cobamide biosynthesis in this family, the cobamide biosynthesis operons in a subset of the genomes were visualized using the IMG gene neighborhood tool with best bidirectional hits to *cbiD* (Fig. 4). While the relative locations of the *cbi* genes are generally conserved in those members of the *Lachnospiraceae* that have them, even closely related species have different genomic contexts for their *cbi* gene clusters. There is often rearrangement of the *cbiG, cbiL,* and *cbiJ* genes encoding methyltransferases that catalyze the corrin ring substitutions. The genes *cbiMNOQ*, encoding the cobalt importer, is located near the corrin ring genes in some genomes, but are upstream of the cluster in some cases and downstream in others. Likewise, the tetrapyrrole precursor biosynthesis genes,

*hemALBCD* and *cysG*, are not always near the *cbi* genes. Loss of these tetrapyrrole precursor genes has occurred in the genomes depicted in Fig. 4A, but the relatively low confidence of the relationship between these species (Fig. 2) makes it difficult to annotate gene loss events without additional information. For example, in this cladogram (Fig. 4A), the most parsimonious explanation is the loss of *hemALB* basal to this group, and then a gain of *hemB* in *Lachnoclostridium (Clostridium) scindens*, a strain that has been validated to be an ALA salvager (Chapter 2)[221]. Alternatively, there have been independent multiple loss events of *hemAL,* and also *hemB* in this clade.

One clade in the genus *Blautia* also exhibits retention of cobamide biosynthesis relative to sister clades, possibly due to the cobamide-dependent acetogenesis pathway restricted to this clade (Fig. 2,3). The cobamide biosynthesis operons of this clade are visualized in Fig. 4B. The biosynthesis genes are split from the tetrapyrrole precursor genes and the lower ligand attachment genes, *cobCUST*. For example, in *Blautia obeum*, the tetrapyrrole precursor and corrin ring biosynthesis genes are located on different scaffolds (data not shown).

Additionally, large parts of the *Lachnospiraceae* clade are missing *cbiE* and *cbiT* annotations, suggesting that the search with the EC number used was too narrow. Using BLASTP with a query of *Acetobacterium woodii cbiT,* homologs can be found in this family (data not shown). *cbiET* appear to be fused in a single gene in this clade (Fig. 4), which explains why some of these genes are not accurately annotated.

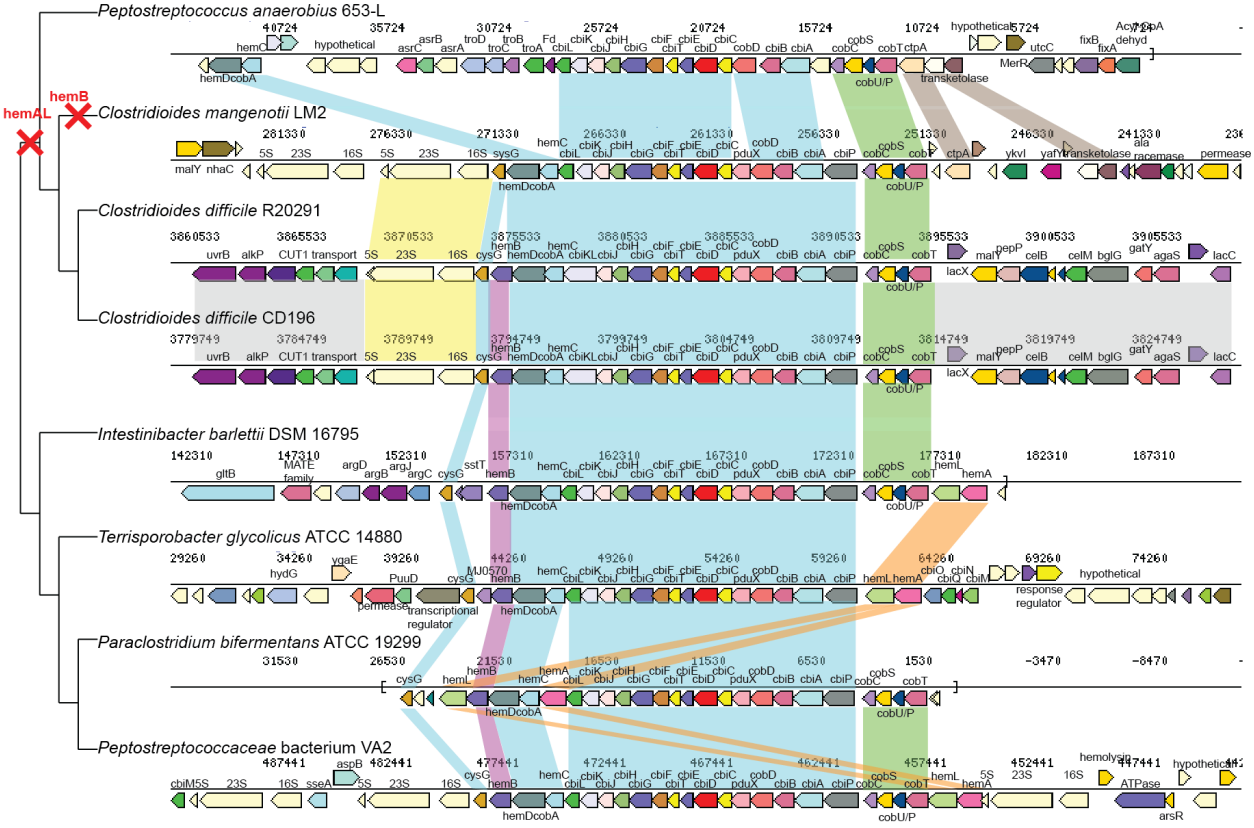



**Figure 5. Synteny of the cobamide biosynthesis gene cluster in selected *Peptostreptococcaceae*.** The *cbiD* of *Clostridium difficile*, a confirmed ALA salvager (Chapter 3), was used as query for the search. *P. anaerobius* 653-L was used instead of VPI 4330 because its scaffold with the *cbi* operon was more complete, and all other genomes are in Fig. 2. Genes of the same color are homologs, white

are pseudogenes, light yellow are genes without COG assignment. Colored pairwise connections between genomes indicate regions of homology. Blue is *cysGhemDcobAcbiKLJHGFTEDCpduXcobDcbiBAP*, green is *cobCSUT*, purple is *hemB*, orange is *hemAL*, yellow is rRNA genes, gray are *C. difficile* shared loci, brown are loci shared between *P. anaerobius* and *C. mangenotii*. Brackets on the genome indicate the end of the scaffold. A cladogram connects the species names based on the relationships in Fig 2.

The *Peptostreptococcaceae* family (Fig. 2,3 light green) includes the opportunistic pathogens *Clostridioides difficile, Clostridium sordellii*, and *Peptostreptococus anaerobius*[209,265,266]. While many members of this clade have annotations for the complete cobamide biosynthesis pathway, there are a few notable exceptions that are missing *hemAL* to make ALA or *hemALB* to make porphobilinogen. In the case of *C. difficile* and *Clostridioides mangenotii,* the most parsimonious explanation is that the ancestor lost both *hemA* and *hemL*, and that *C. mangenotii* also lost *hemB* (Fig. 5, crosses). Both species have at least 6 cobamide-dependent gene families in this analysis, so they must take up cobamide precursors or cobamides to use these metabolisms (see Chapter 3 on *C. difficile)* (Fig. 3). In contrast, I found no cobamide-dependent genes in *P. anaerobius*, which is missing *hemA* and *hemL*. In the *Peptostreptococcaceae*, gene order appears to be more conserved (Fig. 5) than in the Lachnospiraceae examined (Fig. 4). Again, the tetrapyrrole precursor genes appear to be most variable in their position relative to the *cbi* operon. For four of the strains in Figure 5, the ribosomal rRNA genes are located adjacent to the *cbi* genes, but as with the *Lachnospiraceae*, the genomic context of the locus is highly variable.

**Discussion and Future Directions**

I speculate that losses of cobamide biosynthesis genes occur when there is sufficient cobamide or cobamide precursors in the environment to allow use of the cobamide-dependent pathways. One lineage, the Lachnospiraceae, has incomplete cobamide biosynthesis pathways and primarily only MetH as their cobamide-requiring pathway, discussed above, and may either have access to enough cobamides to meet their methionine requirement through MetH or have plentiful exogenous methionine. These bacteria are largely animal host-associated, primarily from the mammalian digestive tract[263] and cobamides are likely to be present in their habitats. Cobamides have been detected in human feces at tens of nanograms per gram feces[187], and have been shown to be synthesized at hundreds of milligrams per day in dairy cow rumens[188].

While no eukaryotic algae are known to produce cobamides, studies on the occurrence of MetH and MetE have shown that there have been multiple independent losses of both genes. When serially passaged with added vitamin $B_{12}$, *Chlamydomonas reinhardtii* can lose MetE and exclusively use MetH[35,267]. Alternatively, lineages may lose cobamide biosynthetic capability because of a relaxed requirement for cobamide-dependent functions. The case of *P. anaerobius* in the *Peptostreptococcaceae* discussed above is a possible example of this. In addition to RNRs and methionine synthase, there are at least five other cobamide-dependent enzymes that have cobamide-independent alternatives. Given that many bacteria can use multiple carbon and energy sources, they may preferentially use those that do not require cobamides for catabolism.

This chapter provides a number of interesting observations of cobamide-related genes in the phylum Firmicutes that can be further explored with statistical techniques to better understand the evolutionary history of the pathway, cobamide biosynthesis genetic architecture, mechanistic explanations for gene gain and loss. A major limitation of the approach in this chapter is that, while meeting the tests used for genome completeness, a few

genomes are highly fragmented. The automated, annotation-based approach may overestimate gene loss in these genomes. Whole genome alignments and manual checking of the scaffolds for genes that may not have been called by the automated analysis due to truncation may improve confidence in the lack of a gene in any given genome.

To further understand the evolutionary relationship between cobamide biosynthesis genes, gene trees of all the cobamide biosynthesis genes would help define the relationship between the corrin ring, lower ligand attachment, and the tetrapyrrole precursor genes. The differences in gene order in both *Lachnospiraceae* and *Peptostreptococcaceae* may be a result of internal gene rearrangement, or in rare cases, horizontal gene transfer. To further support the 16S rRNA trees presented here and to better infer instances of lateral gene transfer, concatenated ribosomal protein trees should also be used to determine the phylogenetic relationship between the species. Neither method may reflect the true species tree, but clades found in both trees would be assumed to be well supported. Previously published species trees for certain clades could also be used for analysis if the genomes are available to be analyzed for their cobamide biosynthesis and dependence gene content. Only a well-supported phylogenetic relationship will help uncover the apparent order of gene gain and loss in the lineage.

One question that remains unanswered is what causes the apparent gene loss. This chapter does not attempt to definitively find pseudogenes, but we would expect to see them if the entire cobamide pathway is under relaxed selection. In an example visualized in Fig. 5, *hemA* and *hemL* are found in different locations relative to the corrin ring genes, and are lost in the *C. difficile* and *C. mangenotii* genomes. Additional sampling of very closely related strains may help further resolve the timeline of apparent loss.

Further work to understand the habitat and physiology of bacteria with missing or partial cobamide biosynthesis pathways that have cobamide-dependent enzymes will improve our understanding of what selective forces allow these organisms to dispense with the ability to produce cobamides. One framework is the Black Queen Hypothesis, which describes the adaptive advantage of losing a function as long as the function is sufficiently performed by the other members of the community[119]. By making observations on the availability of cobamides and cobamide precursors in extant habitats, we can speculate on the availability of these nutrients in the evolutionary history of cobamide auxotrophs.

# Chapter 5

## Summary and Future Directions

### Summary

Prior to my work, bioinformatic studies of hundreds of genomes observed that only some bacteria and archaea that use cobamides can synthesize them *de novo*, suggesting that there is widespread sharing of cobamides in microbial communities[7–11]. Since then, our knowledge of cobamide biosynthesis genes and lineage-specific variants has increased. A number of new cobamide dependent enzymes have been described, as have new cobamide-independent alternative pathways, and advances in sequencing have led to tens of thousands of available microbial genomes.

By using our experimental knowledge to constrain genomic analysis of the cobamide biosynthetic pathways, with newly confirmed cobamide producers by Kenny Mok, in over 11,000 bacterial genomes, Erica Seth, Andrew Han, and I found that over 80% of bacteria use cobamides yet only 37% are predicted to produce them *de novo*, consistent with previous observations. I collaborated with Samantha Jackson and David Haft at the J. Craig Venter Institute to develop profile hidden Markov models (HMMs) for the newly discovered anaerobic benzimidazole biosynthesis genes to better identify these genes in genomes[68,72], and with these and other known genetic lower ligand determinants, I was able to predict the lower ligand class for 58% of putative cobamide producing bacteria.

The comparative genomics analysis also revealed that 17% of bacterial species have only partial cobamide biosynthetic pathways, but that they may be capable of salvaging cobinamide (Cbi) or other cobamide precursors. Cobamide precursor salvaging in bacteria had been underexplored, and the extent of potential precursor salvaging bacteria was unexpected. The distribution of these predicted phenotypes is uneven, for example, the phylum Bacteroidetes was predicted to contain fewer than 1% cobamide producers, but up to 30% of the species were predicted to salvage Cbi. While few in proportion, predicted tetrapyrrole precursor salvagers were enriched in the phylum Firmicutes. Erica Seth experimentally validated three bacteria missing the 5-aminolevulinic acid (ALA) synthesis genes as being capable of taking up ALA and producing cobamides.

One of these predicted ALA-salvaging bacteria was *Clostridioides (Clostridium) difficile*, which encodes seven cobamide-dependent pathways in its genome, higher than the average number of cobamide-dependent pathways for Firmicutes. I confirmed that *C. difficile* can only make cobamides with exogenous cobamide precursors, and that this cobamide is pseudocobalamin unless exogenous benzimidazoles are available. I explored the ability of nine different cobamides to support cobamide-dependent growth, and under the methionine synthase (MetH) dependent condition, *C. difficile* could use all of them, making it one of the most promiscuous cobamide-dependent metabolisms that has been observed. I developed deletion mutants with the help of Xun Lyu to interrogate other aspects of its cobamide-dependent metabolism. A deletion of the cobamide-independent ribonucleotide reductase (NrdDG) caused cells to grow only in a cobamide-dependent manner, relying on NrdJ. NrdJ was unable to use phenolyl cobamides, as expected. A deletion of the putative cobamide importer BtuFCD confirmed that it was necessary for uptake of physiological relevant concentrations of cobamides and Cbi. Because RT027 strains, including the hypervirulent strain R20291, have transposon insertions in *btuC*[242], it is likely that these strains are unable to acquire sufficient

cobamides to support cobamide-dependent metabolism, but were able to grow in the presence of ALA. Based on these experiments, *C. difficile*'s cobamide biology is just as flexible as the rest of its metabolism, and shows the adaptability and thus success of this opportunistic pathogen.

The uneven taxonomic distribution of the cobamide biosynthesis pathway at the phylum level led to questions about when and how it is gained or lost in bacterial lineages. Observations of cobamide biosynthesis in the Clostridiales showed multiple independent losses of all or part of the pathway, but specific retention in some clades. Further analysis of cobamide biosynthesis gene trees would resolve some uncertainty in the occurrence of gain and loss events. Comparing the cobamide biosynthesis operon and surrounding gene neighborhood in selected *Lachnospiraceae* and *Peptostreptococcaceae* revealed high conservation of the order of corrin ring biosynthesis genes, but occasional rearrangement of cobalt importer genes, tetrapyrrole precursor genes, and a subset of corrin ring methyltransferases. These observations show that even closely related species can have divergent cobamide biosynthesis gene content. For example, the taxonomic integration as performed in Chapter 2 at the phylum-level might obscure details in specific systems.

**Future directions**

Genomics is a powerful tool for predicting bacterial phenotypes and potential metabolic interactions in communities. However, these predictions need to be validated by experiment. Genomics also cannot yet predict which cobamides an organism requires based on the sequence of its cobamide-dependent enzymes, cobamide riboswitches, or cobamide transporter, and must be empirically determined for each strain. Additional experiments to discover sequence determinants for cobamide specificity for these functions will be necessary to be able to use genomics to predict cobamide selectivity for an organism. There are a number of intriguing hypotheses that could be tested from my comparative genomics analysis, and I will list a few here.

We predicted that up to 17% of bacteria might potentially salvage cobamide precursors, including tetrapyrrole precursors. The tetrapyrrole precursor salvaging activities for precursors other than ALA have not been confirmed, due to the limited commercial availability of some of these compounds in high quantities due to their instability. While Cbi salvaging has been previously observed, it might be more common in bacteria than previously appreciated. For example, the observation that many Bacteroidetes may be able to salvage Cbi or other late intermediates could be important to their physiology in the host. The relative abundance of *Bacteroides* spp. affect human health, and these bacteria often rely on cobamides in the function of methionine synthase and methylmalonyl-CoA mutase[9].

Further understanding of the role of cobamide and cobamide precursors *in situ* will require additional measurements of the abundance of these compounds in a variety of ecosystems to establish a steady-state baseline for cobamide availability in habitats. So far, only human stool[187], bovine rumen[188], and a trichloroethene-degrading enrichment cultures[185] have been analyzed for their cobamide diversity. Additional studies could reveal the temporal dynamics of cobamides and cobamide precursors, particularly in the human gut under dysbiosis, where change in the abundance of community members may affect the total amount and specific cobamides available to gut microbes, and this in turn may affect host health. For example, the ability to use ethanolamine via the cobamide-dependent enzyme EutBC[222], which requires a

base-on cobamide[13], or other cobamide-dependent processes could affect the colonization and disease state due to infection by *C. difficile* or other pathogens.

My work has focused on primarily single bacterial isolates, but to better understand cobamide cross-feeding, analysis of cobamide metabolism must also include mixed synthetic and natural communities. Synthetic co-cultures and enrichments should prove useful in the exploration of the mechanistic basis of cobamide sharing. This includes how cobamides are released from cells, the temporal dynamics of the interaction, and reciprocity of the interaction. Metagenomics, metatranscriptomics, metaproteomics, and metabolomics of communities may reveal cobamide sharing *in situ*. My pipeline to analyze isolate genomes requires genomes to be near complete, so a different prediction method for cobamide biosynthesis is needed to have higher confidence in predictions of biosynthesis and auxotrophy in incomplete genomes, such as those generated by metagenomics and single-cell genomics. I proposed marker genes for cobamide biosynthesis in Chapter 2, and I think additional phenotypic validation of both predicted cobamide producers and cobamide non-producers would increase the confidence of using these markers in diverse systems. Being able to confidently assess potential cobamide biosynthesis phenotypes in incomplete genomes will increase the diversity of microbes and ecosystems in which we can predict cobamide cross-feeding.

Cobamide cross-feeding is expected to be widespread in microbial communities. My work has created a resource of predicted cobamide biosynthesis and dependence phenotypes that can be used to test cobamide specificity and cobamide sharing in diverse environments.

## References


1. D'Souza, G. *et al.* Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* **35**, 455–488 (2018).
2. Zengler, K. & Zaramela, L. S. The social network of microorganisms - How auxotrophies shape complex communities. *Nat. Rev. Microbiol.* **16**, 383–390 (2018).
3. D'Souza, G. *et al.* Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution (N. Y).* **68**, 2559–2570 (2014).
4. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2149-56 (2014).
5. Price, M. N. *et al.* Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet.* **14**, 1–23 (2018).
6. D'Souza, G. & Kost, C. Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet.* **12**, 1–27 (2016).
7. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. *J. Biol. Chem.* **278**, 41148–41159 (2003).
8. Zhang, Y., Rodionov, D. A., Gelfand, M. S. & Gladyshev, V. N. Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. *BMC Genomics* **10**, 78 (2009).
9. Degnan, P. H., Barry, N. A., Mok, K. C., Taga, M. E. & Goodman, A. L. Human gut microbes use multiple transporters to distinguish vitamin B 12 analogs and compete in the gut. *Cell Host Microbe* **15**, 47–57 (2014).
10. Magnúsdóttir, S., Ravcheev, D., De Crécy-Lagard, V. & Thiele, I. Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes. *Front. Genet.* **6**, 148 (2015).
11. Helliwell, K. E. *et al.* Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12. *Curr. Biol.* **26**, 999–1008 (2016).
12. Matthews, R. G., Koutmos, M. & Datta, S. Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr. Opin. Struct. Biol.* **18**, 658–666 (2008).
13. Banerjee, R. & Ragsdale, S. W. The Many Faces of Vitamin B 12 : Catalysis by Cobalamin-Dependent Enzymes. *Annu. Rev. Biochem.* **72**, 209–247 (2003).
14. Bridwell-Rabb, J. & Drennan, C. L. Vitamin B12 in the spotlight again. *Curr. Opin. Chem. Biol.* **37**, 63–70 (2017).
15. Miles, Z. D., McCarty, R. M., Molnar, G. & Bandarian, V. Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7368–7372 (2011).
16. Kräutler, B. Thermodynamic trans - Effects of the Nucleotide Base in the B12 Coenzymes. *Helv. Chim. Acta* **70**, 1268–1278 (1987).
17. Renz, P. Biosynthesis of the 5,6-dimethylbenzimidazole moiety of cobalamin and of the other bases found in natural corrinoids. in *Chemistry and Biochemistry of B12* (ed. Banerjee, R.) 557–576 (John Wiley & Sons, Inc, 1999).
18. Sokolovskaya, O. M. *et al.* Cofactor selectivity in methylmalonyl-CoA mutase, a model cobamide-dependent enzyme. *bioRxiv* 1–28 (2019). doi:http://dx.doi.org/10.1101/637140
19. Anderson, P. J. *et al.* One pathway can incorporate either adenine or dimethylbenzimidazole as an α-axial ligand of B12 cofactors in Salmonella enterica. *J.*

*Bacteriol.* **190**, 1160–71 (2008).
20. Yan, J. *et al.* The corrinoid cofactor of reductive dehalogenases affects dechlorination rates and extents in organohalide-respiring Dehalococcoides mccartyi. *ISME J.* **10**, 1092–1101 (2016).
21. Keller, S. *et al.* Selective Utilization of Benzimidazolyl-Norcobamides as Cofactors by the Tetrachloroethene Reductive Dehalogenase of Sulfurospirillum multivorans. *J. Bacteriol.* **200**, e00584-17 (2018).
22. Tanioka, Y. *et al.* Methyladeninylcobamide functions as the cofactor of methionine synthase in a Cyanobacterium, Spirulina platensis NIES-39. *FEBS Lett.* **584**, 3223–3226 (2010).
23. Roth, J. R., Lawrence, J. G. & Bobik, T. A. Cobalamin (Coenzyme B12): Synthesis and Biological Significance. *Annu. Rev. Microbiol.* **50**, 137–181 (1996).
24. Hamilton, F. D. Ribonucleotide Reductase from Euglena gracilis, a Deoxyadenosylcobalamin-dependent Enzyme*. *J. Biol. Chem.* **249**, 4428–4434 (1974).
25. Torrents, E. *et al.* Euglena gracilis ribonucleotide reductase: The eukaryote class II enzyme and the possible antiquity of eukaryote B12 dependence. *J. Biol. Chem.* **281**, 5604–5611 (2006).
26. Kräutler, B. *et al.* The Cofactor of Tetrachloroethene Reductive Dehalogenase of Dehalospirillum multivorans Is Norpseudo-B12, a New Type of a Natural Corrinoid. *Helv. Chim. Acta* **86**, 3698–3716 (2003).
27. Parks, J. M. *et al.* The Genetic Basis for Bacterial Mercury Methylation. *Science* **220**, 173–174 (2013).
28. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L. & Liu, H. A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature* **544**, 322–326 (2017).
29. Blaszczyk, A. J. *et al.* Spectroscopic and Electrochemical Characterization of the Iron-Sulfur and Cobalamin Cofactors of TsrM, an Unusual Radical S-Adenosylmethionine Methylase. *J. Am. Chem. Soc.* **138**, 3416–3426 (2016).
30. Yamamoto, H., Fang, M., Dragnea, V. & Bauer, C. E. Differing isoforms of the cobalamin binding photoreceptor AerR oppositely regulate photosystem expression. *Elife* **7**, 1–26 (2018).
31. Ortiz-Guerrero, J. M., Polanco, M. C., Murillo, F. J., Padmanabhan, S. & Elías-Arnanz, M. Light-dependent gene regulation by a coenzyme B12-based photoreceptor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7565–7570 (2011).
32. Romine, M. F. *et al.* Elucidation of roles for vitamin B 12 in regulation of folate, ubiquinone, and methionine metabolism . *Proc. Natl. Acad. Sci.* **114**, E1205–E1214 (2017).
33. Fontecave, M. Ribonucleotide reductases and radical reactions. *Cell. Mol. Life Sci.* **54**, 684–695 (1998).
34. Savvi, S. *et al.* Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium tuberculosis: Implications for propionate metabolism during growth on fatty acids. *J. Bacteriol.* **190**, 3886–3895 (2008).
35. Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E. & Smith, A. G. Insights into the evolution of vitamin B 12 auxotrophy from sequenced algal genomes. *Mol. Biol. Evol.* **28**, 2921–2933 (2011).
36. Xie, B. *et al.* Chlamydomonas reinhardtii thermal tolerance enhancement mediated by a mutualistic interaction with vitamin B12-producing bacteria. *ISME J.* **7**, 1544–1555

(2013).
37. Gonzalez, J. C., Banerjee, R. V., Huang, S., Sumner, J. S. & Matthews, R. G. Comparison of Cobalamin-Independent and Cobalamin-Dependent Methionine Synthases from Escherichia coli: Two Solutions to the Same Chemical Problem. *Biochemistry* **31**, 6045–6056 (1992).
38. Aretakis, J. R., Gega, A. & Schrader, J. M. Absolute Measurements of mRNA Translation in Caulobacter crescentus Reveal Important Fitness Costs of Vitamin B12. *mSystems* **4**, 1–14 (2019).
39. Taga, M. E. & Walker, G. C. Sinorhizobium meliloti Requires a Cobalamin-Dependent Ribonucleotide Reductase for Symbiosis With Its Plant Host. *MPMI* **23**, 1643–1654 (2010).
40. Vetting, M. W. *et al.* Experimental strategies for functional annotation and metabolism discovery: Targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **54**, 909–931 (2015).
41. Zallot, R. *et al.* Identification of a Novel Epoxyqueuosine Reductase Family by Comparative Genomics. *ACS Chem. Biol.* **12**, 844–851 (2017).
42. LaMattina, J. W. *et al.* 1,2-propanediol Dehydration in Roseburia inulinivorans; Structural Basis for Substrate and Enantiomer Selectivity. *J. Biol. Chem.* jbc.M116.721142 (2016). doi:10.1074/jbc.M116.721142
43. Raynaud, C., Sarçabal, P., Meynial-Salles, I., Croux, C. & Soucaille, P. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5010–5015 (2003).
44. Daniel, R., Bobik, T. A. & Gottschalk, G. Biochemistry of coenzyme B12-dependent glycerol and diol dehydratases and organization of the encoding genes. *FEMS Microbiol. Rev.* **22**, 553–566 (1998).
45. Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J. & Kerfeld, C. A. In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. *Sci. Rep.* **7**, 42757 (2017).
46. Warren, M. J., Raux, E., Schubert, H. L. & Escalante-Semerena, J. C. The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* **19**, 390–412 (2002).
47. Martens, J. H., Barg, H., Warren, M. J. & Jahn, D. Microbial production of vitamin B12. *Appl. Microbiol. Biotechnol.* **58**, 275–285 (2002).
48. Tavares, N. K., Zayas, C. L. & Escalante-Semerena, J. C. The Methanosarcina mazei MM2060 Gene Encodes a Bifunctional Kinase/Decarboxylase Enzyme Involved in Cobamide Biosynthesis. *Biochemistry* **57**, 4478–4495 (2018).
49. Jeter, V. L., Mattes, T. A., Beattie, N. R. & Escalante-Semerena, J. C. A New Class of Phosphoribosyltransferases Involved in Cobamide Biosynthesis Is Found in Methanogenic Archaea and Cyanobacteria. *Biochemistry* **58**, 951–964 (2019).
50. McGoldrick, H. M. *et al.* Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from Rhodobacter capsulatus, containing flavin, heme, and Fe-S cofactors. *J. Biol. Chem.* **280**, 1086–1094 (2005).
51. Otte, M. M. & Escalante-Semerena, J. C. Biochemical characterization of the GTP:Adenosylcobinamide-phosphate guanylyltransferase (CobY) enzyme of the hyperthermophilic archaeon Methanocaldococcus jannaschii. *Biochemistry* **48**, 5882–5889 (2009).
52. Neidle, E. L. & Kaplan, S. Expression of the *Rhodobacter sphaeroides hemA* and *hemT*

genes, encoding two 5-aminolevulinic acid synthase isozymes. *J. Bacteriol.* **175**, 2292–2303 (1993).
53. Raux, E., Schubert, H. L. & Warren, M. J. Biosynthesis of cobalamin (vitamin B12): A bacterial conundrum. *Cell. Mol. Life Sci.* **57**, 1880–1893 (2000).
54. Avissar, Y. J. & Beale, S. I. Identification of the enzymatic basis for Δ-aminolevulinic acid auxotrophy in a hemA mutant of Escherichia coli. *J. Bacteriol.* **171**, 2919–2924 (1989).
55. Bali, S., Palmer, D. J., Schroeder, S., Ferguson, S. J. & Warren, M. J. Recent advances in the biosynthesis of modified tetrapyrroles: The discovery of an alternative pathway for the formation of heme and heme d 1. *Cell. Mol. Life Sci.* **71**, 2837–2863 (2014).
56. Fan, C. & Bobik, T. A. The PduX enzyme of Salmonella enterica is an L-threonine kinase used for coenzyme B12 synthesis. *J. Biol. Chem.* **283**, 11322–11329 (2008).
57. Fan, C., Fromm, H. J. & Bobik, T. A. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of Salmonella enterica. *J. Biol. Chem.* **284**, 20240–20248 (2009).
58. Grabau, C. & Roth, J. R. A *Salmonella typhimurium* cobalamin-deficient mutant blocked in 1-amino-2-propanol synthesis. *J. Bacteriol.* **174**, 2138–44 (1992).
59. Brushaber, K. R., O'Toole, G. A. & Escalante-Semerena, J. C. CobD , a Novel Enzyme with L-Threonine-O-3-phosphate Decarboxylase Activity, Is Responsible for the Synthesis of (R)-1-Amino-2-propanol O-2-Phosphate, a Proposed New Intermediate in Cobalamin Biosynthesis in Salmonella typhimurium LT2*. *J. Biol. Chem.* **273**, 2684–2691 (1998).
60. Tavares, N. K., VanDrisse, C. M. & Escalante-Semerena, J. C. Rhodobacterales use a unique L-threonine kinase for the assembly of the nucleotide loop of coenzyme B12. *Mol. Microbiol.* **110**, 239–261 (2018).
61. Keller, S., Treder, A., Von Reuss, S. H., Escalante-Semerena, J. C. & Schubert, T. The SMUL_1544 gene product governs norcobamide biosynthesis in the tetrachloroethene-respiring bacterium Sulfurospirillum multivorans. *J. Bacteriol.* **198**, 2236–2243 (2016).
62. Trzebiatowski, J. R. & Escalante-Semerena, J. C. Purification and Characterization of CobT, the Nicotinate-mononucleotide:5,6-Dimethylbenzimidazole Phosphoribosyltransferase Enzyme from Salmonella typhimurium LT2*. *J. Biol. Chem.* **272**, 17662–17667 (1997).
63. Zayas, C. L. & Escalante-Semerena, J. C. Reassessment of the late steps of coenzyme B12 synthesis in Salmonella enterica: Evidence that dephosphorylation of adenosylcobalamin-5'-phosphate by the CobC phosphatase is the last step of the pathway. *J. Bacteriol.* **189**, 2210–2218 (2007).
64. Chan, C. H. & Escalante-Semerena, J. C. ArsAB, a novel enzyme from Sporomusa ovata activates phenolic bases for adenosylcobamide biosynthesis. *Mol. Microbiol.* **81**, 952–967 (2011).
65. Newmister, S. A., Chan, C. H., Escalante-Semerena, J. C. & Rayment, I. Structural insights into the function of the nicotinate mononucleotide:phenol/p-cresol phosphoribosyltransferase (ArsAB) enzyme from Sporomusa ovata. *Biochemistry* **51**, 8571–8582 (2012).
66. Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T. & Walker, G. C. BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* **446**, 449–453 (2007).

67. Campbell, G. R. O. *et al.* Sinorhizobium meliloti bluB is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. *Proc. Natl. Acad. Sci.* **103**, 4634–4639 (2006).
68. Hazra, A. B. *et al.* Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc. Natl. Acad. Sci.* **112**, 10792–7 (2015).
69. Yan, J. *et al.* Purinyl-cobamide is a native prosthetic group of reductive dehalogenases. *Nat. Chem. Biol.* **14**, 8–14 (2018).
70. Gray, M. J. & Escalante-Semerena, J. C. Single-enzyme conversion of FMNH2 to 5, 6-dimethylbenzimidazole the lower ligand of B12. *Proc. Natl. Acad. Sci.* **104**, 2921–2926 (2007).
71. Hazra, A. B., Ballou, D. P. & Taga, M. E. Unique Biochemical and Sequence Features Enable BluB to Destroy Flavin and Distinguish BluB from the Flavin Monooxygenase Superfamily. *Biochemistry* **57**, 1748–1757 (2018).
72. Mehta, A. P. *et al.* Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis. *J. Am. Chem. Soc.* **137**, 10444–10447 (2015).
73. Stupperich, E. & Eisinger, H. J. Biosynthesis of para-cresolyl cobamide in Sporomusa ovata. *Arch. Microbiol.* **151**, 372–377 (1989).
74. Crofts, T. S., Seth, E. C., Hazra, A. B. & Taga, M. E. Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chem. Biol.* **20**, 1265–1274 (2013).
75. Hazra, A. B., Tran, J. L. A., Crofts, T. S. & Taga, M. E. Analysis of substrate specificity in CobT homologs reveals widespread preference for DMB, the lower axial ligand of vitamin B12. *Chem. Biol.* **20**, 1275–1285 (2013).
76. Mireku, S. A. *et al.* Conformational Change of a Tryptophan Residue in BtuF Facilitates Binding and Transport of Cobinamide by the Vitamin B12 Transporter BtuCD-F. *Sci. Rep.* **7**, 41575 (2017).
77. Borths, E. L., Locher, K. P., Lee, A. T. & Rees, D. C. The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. *Proc. Natl. Acad. Sci.* **99**, 16642–16647 (2002).
78. Agarwal, S., Dey, S., Ghosh, B., Biswas, M. & Dasgupta, J. Mechanistic basis of vitamin B12 and cobinamide salvaging by the Vibrio species. *Biochim. Biophys. Acta - Proteins Proteomics* **1867**, 140–151 (2019).
79. Di Girolamo, P. M. & Bradbeer, C. Transport of vitamin B12 in Escherichia coli. *J. Bacteriol.* **106**, 745–750 (1976).
80. Santos, J. A. *et al.* Functional and structural characterization of an ECF-type ABC transporter for vitamin B12. *Elife* **7**, 1–16 (2018).
81. Gopinath, K. *et al.* A vitamin B12 transporter in Mycobacterium tuberculosis. *Open Biol.* **3**, 120175 (2013).
82. Rodionov, D. A. *et al.* A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **91**, 42–51 (2009).
83. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043–1049 (2002).
84. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *Rna* **9**, 1084–1097 (2003).

85. Gallo, S., Oberhuber, M., Sigel, R. K. O. & Kräutler, B. The corrin moiety of coenzyme B12 is the determinant for switching the btuB riboswitch of E. coli. *ChemBioChem* **9**, 1408–1414 (2008).
86. Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32**, 143–150 (2004).
87. Choudhary, P. K. *et al.* Diversity of cobalamin riboswitches in the corrinoid-producing organohalide respirer Desulfitobacterium hafniense. *J. Bacteriol.* **195**, 5186–95 (2013).
88. Perez, A. A., Rodionov, D. A. & Bryant, D. A. Identification and regulation of genes for cobalamin transport in the cyanobacterium Synechococcus sp. strain PCC 7002. *J. Bacteriol.* **198**, 2753–2761 (2016).
89. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a Protein Binds B12: A 3.0 A X-ray Structure of B12-Binding Domains of Methionine Synthase. *Science (80-. ).* **266**, 1669–1674 (1994).
90. Dowling, D. P., Croft, A. K. & Drennan, C. L. Radical Use of Rossmann and TIM Barrel Architectures for Controlling Coenzyme B 12 Chemistry . *Annu. Rev. Biophys.* **41**, 403–427 (2012).
91. Yan, J., Im, J., Yang, Y. & Löffler, F. E. Guided cobalamin biosynthesis supports dehalococcoides mccartyi reductive dechlorination activity. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120320 (2013).
92. Kruse, S. *et al.* Guided cobamide biosynthesis for heterologous production of reductive dehalogenases. *Microb. Biotechnol.* (2018). doi:10.1111/1751-7915.13339
93. Mok, K. C. & Taga, M. E. Growth inhibition of Sporomusa ovata by incorporation of benzimidazole bases into cobamides. *J. Bacteriol.* **195**, 1902–11 (2013).
94. Keller, S. *et al.* Exogenous 5,6-dimethylbenzimidazole caused production of a non-functional tetrachloroethene reductive dehalogenase in Sulfurospirillum multivorans. *Environ. Microbiol.* **16**, 3361–3369 (2013).
95. Crofts, T. S., Men, Y., Alvarez-Cohen, L. & Taga, M. E. A bioassay for the detection of benzimidazoles reveals their presence in a range of environmental samples. *Front. Microbiol.* **5**, 592 (2014).
96. Lawrence, J. G. & Roth, J. R. The Cobalamin (Coenzyme B12) Biosynthetic Genes of Escherichia coli. *Microbiology* **177**, 6371–6380 (1995).
97. Butzin, N. C., Secinaro, M. A., Swithers, K. S., Gogarten, J. P. & Noll, K. M. Thermotoga lettingae can salvage cobinamide to synthesize vitamin B12. *Appl. Environ. Microbiol.* **79**, 7006–12 (2013).
98. Swithers, K. S. *et al.* Vitamin B(12) synthesis and salvage pathways were acquired by horizontal gene transfer to the Thermotogales. *Genome Biol. Evol.* **4**, 730–9 (2012).
99. Mattes, T. A. & Escalante-Semerena, J. C. Salmonella enterica synthesizes 5,6-dimethylbenzimidazolyl-(DMB)-α-riboside. Why some Firmicutes do not require the canonical DMB activation system to synthesize adenosylcobalamin. *Mol. Microbiol.* **103**, 269–281 (2017).
100. Gray, M. J. & Escalante-Semerena, J. C. A new pathway for the synthesis of α-ribazole-phosphate in Listeria innocua. *Mol. Microbiol.* **77**, 1429–1438 (2010).
101. Wienhausen, G., Noriega-Ortega, B. E., Niggemann, J., Dittmar, T. & Simon, M. The Exometabolome of Two Model Strains of the Roseobacter Group: A Marketplace of Microbial Metabolites. *Front. Microbiol.* **8**, 1985 (2017).
102. Johnson, W. M., Kido Soule, M. C. & Kujawinski, E. B. Evidence for quorum sensing

and differential metabolite production by a marine bacterium in response to DMSP. *ISME J.* **10**, 2304–2316 (2016).
103. Gray, M. J., Tavares, N. K. & Escalante-Semerena, J. C. The genome of Rhodobacter sphaeroides strain 2.4.1 encodes functional cobinamide salvaging systems of archaeal and bacterial origins. *Mol. Microbiol.* **70**, 824–836 (2008).
104. Gray, M. J. & Escalante-Semerena, J. C. In vivo analysis of cobinamide salvaging in Rhodobacter sphaeroides strain 2.4.1. *J. Bacteriol.* **191**, 3842–51 (2009).
105. Woodson, J. D. & Escalante-Semerena, J. C. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B12 precursor cobinamide in archaea. *Proc. Natl. Acad. Sci.* **101**, 3591–3596 (2004).
106. Gray, M. J. & Escalante-Semerena, J. C. The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: A critical activity of the cobamide remodelling system of Rhodobacter sphaeroides. *Mol. Microbiol.* **74**, 1198–1210 (2009).
107. Yi, S. *et al.* Versatility in corrinoid salvaging and remodeling pathways supports corrinoid-dependent metabolism in Dehalococcoides mccartyi. *Appl. Environ. Microbiol.* **78**, 7745–7752 (2012).
108. Men, Y. *et al.* Sustainable growth of Dehalococcoides mccartyi 195 by corrinoid salvaging and remodeling in defined lactate-fermenting consortia. *Appl. Environ. Microbiol.* **80**, 2133–2141 (2014).
109. Romine, M. F., Rodionov, D. A., Maezato, Y., Osterman, A. L. & Nelson, W. C. Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J.* **11**, 1434–1446 (2017).
110. Sañudo-Wilhelmy, S. a *et al.* Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14041–5 (2012).
111. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–3 (2005).
112. Wagner-Döbler, I. *et al.* The complete genome sequence of the algal symbiont Dinoroseobacter shibae: a hitchhiker's guide to life in the sea. *ISME J.* **4**, 61–77 (2010).
113. Grant, M. A. A., Kazamia, E., Cicuta, P. & Smith, A. G. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME J.* **8**, 1418–27 (2014).
114. Kazamia, E. *et al.* Mutualistic interactions between vitamin B12 -dependent algae and heterotrophic bacteria exhibit regulation. *Environ. Microbiol.* **14**, 1466–76 (2012).
115. Helliwell, K. E. *et al.* Quantitative proteomics of a B12-dependent alga grown in coculture with bacteria reveals metabolic tradeoffs required for mutualism. *New Phytol.* **217**, 599–612 (2018).
116. Cooper, M. B. *et al.* Cross-exchange of B-vitamins underpins a mutualistic interaction between Ostreococcus tauri and Dinoroseobacter shibae. *ISME J.* **13**, 334–345 (2019).
117. Yan, J., Ritalahti, K. M., Wagner, D. D. & Löffler, F. E. Unexpected specificity of interspecies cobamide transfer from Geobacter spp. to organohalide-respiring Dehalococcoides mccartyi strains. *Appl. Environ. Microbiol.* **78**, 6630–6636 (2012).
118. Bonnet, S. *et al.* Vitamin B12 excretion by cultures of the marine cyanobacteria Crocosphaera and Synechococcus. *Limnol. Oceanogr.* **55**, 1959–1964 (2010).
119. Morris, J. J., Lenski, R. E. & Zinser, E. R. The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *MBio* **3**, e00036-12 (2012).

120. Morris, J. J. Black Queen evolution: The role of leakiness in structuring microbial communities. *Trends Genet.* **31**, 475–482 (2015).
121. Jeffrey Morris, J., Papoulis, S. E. & Lenski, R. E. Coexistence of evolving bacteria stabilized by a shared Black Queen function. *Evolution (N. Y).* **68**, 2960–2971 (2014).
122. Pande, S. *et al.* Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* **8**, 953–62 (2014).
123. Seth, E. C. & Taga, M. E. Nutrient cross-feeding in the microbial world. *Front. Microbiol.* **5**, 350 (2014).
124. Abreu, N. A. & Taga, M. E. Decoding molecular interactions in microbial communities. *FEMS Microbiol. Rev.* **40**, 648–663 (2016).
125. Degnan, P. H., Taga, M. E. & Goodman, A. L. Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab.* **20**, 769–778 (2014).
126. Helliwell, K. E. The roles of B vitamins in phytoplankton nutrition: New perspectives and prospects. *New Phytol.* **216**, 62–68 (2017).
127. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl Pathway of CO2 Fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008).
128. Broderick, J. B., Du, B. R., Duschene, K. S. & Shepard, E. M. Radical S-Adenosylmethionine Enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
129. Markowitz, V. M. *et al.* IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
130. Raes, J., Korbel, J. O., Lercher, M. J., von Mering, C. & Bork, P. Prediction of effective genome size in metagenomic samples. *Genome Biol.* **8**, R10 (2007).
131. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
132. Cornish-Bowden, A. Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspect. Sci.* **1**, 74–87 (2014).
133. Finn, R. D. *et al.* The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
134. Haft, D. H. *et al.* TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* **41**, 387–395 (2012).
135. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–D269 (2015).
136. Parks, J. M. *et al.* The genetic basis for bacterial mercury methylation. *Science (80-. ).* **339**, 1332–5 (2013).
137. Brien, J. R. O. *et al.* Insight into the Mechanism of the B12-Independent Glycerol Dehydratase from Clostridium butryicum: Preliminary Biochemical and Structural Characterization. *Biochemistry* **43**, 4635–4645 (2004).
138. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
139. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
140. Haft, D. R. & Haft, D. H. A comprehensive software suite for protein family construction and functional site prediction. *PLoS One* **12**, e0171758 (2017).
141. Eddy, S. R. HMMER: biosequence analysis using profile hidden Markov models

(v3.1b2) [software]. (2015). Available at: http://hmmer.org/.
142. Textor, S. *et al.* Propionate oxidation in Escherichia coli: Evidence for operation of a methylcitrate cycle in bacteria. *Arch. Microbiol.* **168**, 428–436 (1997).
143. Gough, S. P., Petersen, B. O. & Duus, J. O. Anaerobic chlorophyll isocyclic ring formation in Rhodobacter capsulatus requires a cobalamin cofactor. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6908–13 (2000).
144. He, J., Holmes, V. F., Lee, P. K. H. & Alvarez-Cohen, L. Influence of vitamin B12 and cocultures on the growth of Dehalococcoides isolates in defined medium. *Appl. Environ. Microbiol.* **73**, 2847–2853 (2007).
145. Widdel, F. & Bak, F. Gram-negative mesophilic sulfate-reducing bacteria. in *The Prokaryotes, vol 4* 3352–3378 (1992).
146. Lovitt, R. W., Kell, D. B. & Morris, J. G. The physiology of Clostridium sporogenes NCIB 8053 growing in defined media. *J. Appl. Bacteriol.* **62**, 81–92 (1987).
147. Wolin, E. A., Wolin, M. J. & Wolfe, R. S. Formation of methane by Bacterial Extracts. *J. Biol. Chem.* **238**, 2882–2886 (1963).
148. Graber, J. R. & Breznak, J. A. Physiology and Nutrition of Treponema primitia, an H2/CO2-Acetogenic Spirochete from Termite Hindguts. *Appl. Environ. Microbiol.* **70**, 1307–1314 (2004).
149. Costa, J. C., Barbosa, S. G., Alves, M. M. & Sousa, D. Z. Thermochemical pre- and biological co-treatments to improve hydrolysis and methane production from poultry litter. *Bioresour. Technol.* **111**, 141–147 (2012).
150. Warnick, T. A., Methé, B. A. & Leschine, S. B. Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* **52**, 1155–1160 (2002).
151. Stupperich, E., Eisinger, H. J. & Kräutler, B. Diversity of corrinoids in acetogenic bacteria. P-cresolylcobamide from Sporomusa ovata, 5-methoxy-6-methylbenzimidazolylcobamide from Clostridium formicoaceticum and vitamin B12 from Acetobacterium woodii. *Eur. J. Biochem.* **172**, 459–464 (1988).
152. Cameron, B., Briggs, K., Pridmore, S., Brefort, G. & Crouzet, J. Cloning and analysis of genes involved in coenzyme B12 biosynthesis in Pseudomonas denitrificans. *J. Bacteriol.* **171**, 547–557 (1989).
153. Miyamoto, E. *et al.* Purification and characterization of a corrinoid-compound in an edible cyanobacterium Aphanizomenon flos-aquae as a nutritional supplementary food. *J. Agric. Food Chem.* **54**, 9604–9607 (2006).
154. Wolf, J. B. & Brey, R. N. Isolation and genetic characterizations of Bacillus megaterium cobalamin biosynthesis-deficient mutants. *J. Bacteriol.* **166**, 51–58 (1986).
155. Hoffmann, B. *et al.* Native corrinoids from Clostridium cochlearium are adeninylcobamides: Spectroscopic analysis and identification of pseudovitamin B12 and factor A. *J. Bacteriol.* **182**, 4773–4782 (2000).
156. Barker, H. A., Weissbach, H. & Smyth, R. D. A coenzyme containing pseudovitamin B(12). *Proc. Natl. Acad. Sci. U. S. A.* **44**, 1093–7 (1958).
157. Sophie Bonnet *et al.* Vitamin B12 excretion by cultures of the marine cyanobacteria Crocosphaera and Synechococcus. *Limnol. Oceanogr.* **55**, 1959–1964 (2010).
158. Wang, P. H. *et al.* Refined experimental annotation reveals conserved corrinoid autotrophy in chloroform-respiring Dehalobacter isolates. *ISME J.* **11**, 626–640 (2017).
159. Kräutler, B., Kohler, H. P. & Stupperich, E. 5'-Methylbenzimidazolyl-cobamides are

the corrinoids from some sulfate-reducing and sulfur-metabolizing bacteria. *Eur. J. Biochem.* **176**, 461–469 (1988).
160. Choi, S. C. & Bartha, R. Cobalamin-mediated mercury methylation by Desulfovibrio desulfuricans LS. *Appl. Environ. Microbiol.* **59**, 290–295 (1993).
161. Guimarães, D. H., Weber, A., Klaiber, I., Vogler, B. & Renz, P. Guanylcobamide and hypoxanthylcobamide-Corrinoids formed by Desulfovibrio vulgaris. *Arch. Microbiol.* **162**, 272–276 (1994).
162. Höllriegl, V., Lamm, L., Rowold, J., Hörig, J. & Renz, P. Biosynthesis of vitamin B12 - Different pathways in some aerobic and anaerobic microorganisms. *Arch. Microbiol.* **132**, 155–158 (1982).
163. Engels, C., Ruscheweyh, H. J., Beerenwinkel, N., Lacroix, C. & Schwab, C. The common gut microbe Eubacterium hallii also contributes to intestinal propionate formation. *Front. Microbiol.* **7**, 1–12 (2016).
164. Belzer, C. *et al.* Microbial Metabolic Networks at the Mucus Layer Lead to Diet-Independent Butyrate and Vitamin B 12 Production by Intestinal Symbionts. *MBio* **8**, 1–14 (2017).
165. Schulze, B., Vogler, B. & Renz, P. Biosynthesis of vitamin B12 in anaerobic bacteria--experiments with Eubacterium limosum on the transformation of 5-hydroxy-6-methyl-benzimidazole, its nucleoside, its cobamide, and of 5-hydroxybenzimidazolylcobamide in vitamin B12. *Eur. J. Biochem.* **254**, 620–5 (1998).
166. Torres, A. C. *et al.* Cobalamin production by Lactobacillus coryniformis: biochemical identification of the synthetized corrinoid and genomic analysis of the biosynthetic cluster. *BMC Microbiol.* **16**, 240 (2016).
167. Bhushan, B., Tomar, S. K. & Mandal, S. Phenotypic and genotypic screening of human-originated lactobacilli for vitamin B12 production potential: process validation by micro-assay and UFLC. *Appl. Microbiol. Biotechnol.* **100**, 6791–6803 (2016).
168. De Angelis, M. *et al.* Lactobacillus rossiae, a vitamin B12 producer, represents a metabolically versatile species within the genus lactobacillus. *PLoS One* **9**, 1–11 (2014).
169. Iguchi, H., Yurimoto, H. & Sakai, Y. Stimulation of methanotrophic growth in cocultures by cobalamin excreted by rhizobia. *Appl. Environ. Microbiol.* **77**, 8509–15 (2011).
170. Ivanova, E. G., Fedorov, D. N., Doronina, N. V & Trotsenko, Y. A. Production of vitamin B12 in aerobic methylotrophic bacteria. *Microbiology* **75**, 494–496 (2006).
171. Irion, E. & Ljungdahl, L. Isolation of factor IIIm coenzyme and cobyric acid coenzyme plus other B12 factors from Clostridium thermoaceticum. *Biochemistry* **4**, 2780–90 (1965).
172. Stupperich, E. & Eisinger, H. J. Function and the biosynthesis of unusual corrinoids by a novel activation mechanism of aromatic compounds in anaerobic bacteria. *Adv. Sp. Res.* **9**, 117–125 (1989).
173. Heal, K. R. *et al.* Two distinct pools of B 12 analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci.* **114**, 364–369 (2017).
174. Stupperich, E., Eisinger, H. J. & Schurr, S. Corrinoids in anaerobic bacteria. *FEMS Microbiol. Rev.* **87**, 355–359 (1990).
175. Deptula, P. *et al.* BluB/CobT2 fusion enzyme activity reveals mechanisms responsible for production of active form of vitamin B12 by Propionibacterium freudenreichii. *Microb. Cell Fact.* **14**, 186 (2015).
176. McGoldrick, H., Deery, E., Warren, M. & Heathcote, P. Cobalamin (vitamin B12)

biosynthesis in Rhodobacter capsulatus. *Biochem. Soc. Trans.* **30**, 646–648 (2002).
177. Keck, B. & Renz, P. Salmonella typhimurium forms adenylcobamide and 2-methyladenylcobamide, but no detectable cobalamin during strictly anaerobic growth. *Arch. Microbiol.* **173**, 76–77 (2000).
178. Takano, H., Hagiwara, K. & Ueda, K. Fundamental role of cobalamin biosynthesis in the developmental growth of Streptomyces coelicolor A3 (2). *Appl. Microbiol. Biotechnol.* **99**, 2329–2337 (2015).
179. Perlman, D., O'Brien, E., Bayan, A. P. & Greenfield, R. B. Antibiotic and vitamin B12 production by a steroid oxidizing actinomycete. *J. Bacteriol.* **69**, 347–352 (1955).
180. Prentice, M. B. *et al.* Cobalamin synthesis in Yersinia enterocolitica 8081. Functional aspects of a putative metabolic island. in *The Genus Yersinia. Advances in Experimental Medicine and Biology* (eds. Shurnik, M., Bengoechea, J. A. & Granfors, K.) **529**, 43–46 (2004).
181. Schubert, H. L., Raux, E., Wilson, K. S. & Warren, M. J. Common chelatase design in the branched tetrapyrrole pathways of heme and anaerobic cobalamin synthesis. *Biochemistry* **38**, 10660–10669 (1999).
182. Haas, C. E. *et al.* A subset of the diverse COG0523 family of putative metal chaperones is linked to zinc homeostasis in all kingdoms of life. *BMC Genomics* **10**, 470 (2009).
183. Bertrand, E. M., Saito, M. A., Jeon, Y. J. & Neilan, B. A. Vitamin B12 biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B12 biosynthesizers. *Environ. Microbiol.* **13**, 1285–1298 (2011).
184. Crofts, T. S. *et al.* Regiospecific Formation of Cobamide Isomers Is Directed by CobT. *Biochemistry* **53**, 7805–7815 (2014).
185. Men, Y. *et al.* Identification of specific corrinoids reveals corrinoid modification in dechlorinating microbial communities. *Environ. Microbiol.* **17**, 4873–4884 (2014).
186. Hondorp, E. R. & Matthews, R. G. Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in Escherichia coli. *PLoS Biol.* **2**, e336 (2004).
187. Allen, R. H. & Stabler, S. P. Identification and quantitation of cobalamin and cobalamin analogues in human feces. *Am. J. Clin. Nutr.* **87**, 1324–1335 (2008).
188. Girard, C. L., Santschi, D. E., Stabler, S. P. & Allen, R. H. Apparent ruminal synthesis and intestinal disappearance of vitamin B12 and its analogs in dairy cows. *J. Dairy Sci.* **92**, 4524–4529 (2009).
189. Sloan, D. B. & Moran, N. A. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* **29**, 3781–3792 (2012).
190. Kilstrup, M., Hammer, K., Jensen, P. R. & Martinussen, J. Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol. Rev.* **29**, 555–590 (2005).
191. Paerl, R. W. *et al.* Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *ISME J.* **11**, 753–765 (2016).
192. de Crécy-Lagard, V., El Yacoubi, B., de la Garza, R. D., Noiriel, A. & Hanson, A. D. Comparative genomics of bacterial and plant folate synthesis and salvage: Predictions and validations. *BMC Genomics* **8**, (2007).
193. Roper, J. M. *et al.* The enigma of cobalamin (vitamin B12) biosynthesis in Porphyromonas gingivalis: Identification and characterization of a functional corrin pathway. *J. Biol. Chem.* **275**, 40316–40323 (2000).
194. Dobriner, K. Porphyrin Excretion in the Feces in Normal and Pathological Conditions.

*J. Biol. Chem.* **120**, 115–128 (1937).
195. Watson, C. J., Schwartz, S. & Hawkinson, V. Studies of the Uroporphyrins II. Further Studies of the Porphyrins of the Urine, Feces, Bile, and Liver in Cases of Porphyria, With Particular Reference To a Waldenström Type Porphyrin Behaving As an Entity on the Tswett Column. *J. Biol. Chem.* **157**, 345–362 (1945).
196. Kanto, U. *et al.* Quantification of 5-aminolevulinic acid in swine manure extract by HPLC-fluorescence. *J. Liq. Chromatogr. Relat. Technol.* **36**, 2731–2748 (2013).
197. Giuntini, F., Bourré, L., MacRobert, A. J., Wilson, M. & Eggleston, I. M. Quantitative determination of 5-aminolaevulinic acid and its esters in cell lysates by HPLC-fluorescence. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **875**, 562–566 (2008).
198. Sangwan, I. & O'Brian, M. R. Evidence for an Inter-organismic heme biosynthetic pathway in symbiotic soybean root nodules. *Science (80-. ).* **251**, 1220–1222 (1991).
199. Lyell, N. L. *et al.* An expanded transposon mutant library reveals that Vibrio fischeri δ-aminolevulinate auxotrophs can colonize Euprymna scolopes. *Appl. Environ. Microbiol.* **83**, e02470-16 (2017).
200. Qi, M. *et al.* Genomic differences between Fibrobacter succinogenes S85 and Fibrobacter intestinalis DR7, identified by suppression subtractive hybridization. *Appl. Environ. Microbiol.* **74**, 987–993 (2008).
201. McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci.* **106**, 15394–15399 (2009).
202. Morita, H. *et al.* Comparative genome analysis of Lactobacillus reuteri and Lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. *DNA Res.* **15**, 151–161 (2008).
203. Lawrence, J. & Roth, J. Evolution of coenzyme B12 synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics* **142**, 11–24 (1996).
204. Neumann-Schaal, M., Jahn, D. & Schmidt-Hohagen, K. Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile. *Front. Microbiol.* **10**, (2019).
205. Hryckowian, A. J., Pruss, K. M. & Sonnenburg, J. L. The emerging metabolic view of Clostridium difficile pathogenesis. *Curr. Opin. Microbiol.* **35**, 42–47 (2017).
206. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
207. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–9 (2012).
208. Libertucci, J. & Young, V. B. The role of the microbiota in infectious diseases. *Nat. Microbiol.* **4**, 35–45 (2019).
209. Lessa, F. C. *et al.* Burden of Clostridium difficile infection in the United States. *N. Engl. J. Med.* **372**, 825–34 (2015).
210. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nat. Commun.* **5**, 3114 (2014).
211. Theriot, C. M. & Young, V. B. Microbial and metabolic interactions between the gastrointestinal tract and Clostridium difficile infection. *Gut Microbes* **5**, 86–95 (2014).
212. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut

Microbiomes. *mSystems* **2**, e00063-17 (2017).
213. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. Clostridium difficile Alters the Structure and Metabolism of Distinct Cecal Microbiomes during Initial Infection To Promote Sustained Colonization. *mSphere* **3**, 1–15 (2018).
214. Carlucci, C. *et al.* Effects of defined gut microbial ecosystem components on virulence determinants of Clostridioides difficile. *Sci. Rep.* **9**, 1–11 (2019).
215. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. *Cell Host Microbe* **16**, 770–777 (2014).
216. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–9 (2013).
217. Darkoh, C., Plants-Paris, K., Bishoff, D. & DuPont, H. L. Clostridium difficile Modulates the Gut Microbiota by Inducing the Production of Indole, an Interkingdom Signaling and Antimicrobial Molecule. *mSystems* **4**, e00346-18 (2019).
218. Valdés-Varela, L., Hernández-Barranco, A. M., Ruas-Madiedo, P. & Gueimonde, M. Effect of Bifidobacterium upon Clostridium difficile growth and toxicity when co-cultured in different prebiotic substrates. *Front. Microbiol.* **7**, 1–9 (2016).
219. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**, 205–208 (2014).
220. Kang, J. D. *et al.* Bile Acid 7α-Dehydroxylating Gut Bacteria Secrete Antibiotics that Inhibit Clostridium difficile: Role of Secondary Bile Acids. *Cell Chem. Biol.* **26**, 27-34.e4 (2019).
221. Shelton, A. N. *et al.* Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J.* **13**, 789–804 (2019).
222. Nawrocki, K. L., Wetzel, D., Jones, J. B., Woods, E. C. & McBride, S. M. Ethanolamine is a valuable nutrient source that impacts *Clostridium difficile* pathogenesis. *Environ. Microbiol.* **20**, 1419–1435 (2018).
223. Dannheim, H., Will, S. E., Schomburg, D. & Neumann-Schaal, M. Clostridioides difficile 630Δerm in silico and in vivo – quantitative growth and extensive polysaccharide secretion. *FEBS Open Bio* **7**, 602–615 (2017).
224. Fletcher, J. R., Erwin, S., Lanzas, C. & Theriot, C. M. Shifts in the gut metabolome and Clostridium difficile transcriptome throughout colonization and infection in a mouse model. *mSphere* **3**, e00089-18 (2018).
225. Köpke, M., Straub, M. & Dürre, P. Clostridium difficile Is an Autotrophic Bacterial Pathogen. *PLoS One* **8**, 1–7 (2013).
226. Williams, D. R., Young, D. I. & Young, M. Conjugative plasmid transfer from Escherichia coli to Clostridium acetobutylicum. *J. Gen. Microbiol.* **136**, 819–826 (1990).
227. O'Connor, J. R. *et al.* Construction and analysis of chromosomal Clostridium difficile mutants. *Mol. Microbiol.* **61**, 1335–1351 (2006).
228. Ng, Y. K. *et al.* Expanding the Repertoire of Gene Tools for Precise Manipulation of the Clostridium difficile Genome: Allelic Exchange Using pyrE Alleles. *PLoS One* **8**, (2013).
229. Sorg, J. A. & Dineen, S. S. Laboratory maintenance of Clostridium difficile. *Curr. Protoc. Microbiol.* 1–10 (2009). doi:10.1002/9780471729259.mc09a01s12
230. Cartman, S. T. & Minton, N. P. A mariner -Based Transposon System for In Vivo Random Mutagenesis of Clostridium difficile. *Appl. Environ. Microbiol.* **76**, 1103–1109 (2010).
231. Karasawa, T., Ikoma, S., Yamakawa, K. & Nakamura, S. A defined growth medium for

Clostridium difficile. *Microbiology* **141**, 371–375 (1995).
232. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
233. Bouillaut, L., McBride, S. M. & Sorg, J. A. Genetic manipulation of Clostridium difficile. *Curr. Protoc. Microbiol.* 1–17 (2011). doi:10.1002/9780471729259.mc09a02s20
234. Chen, I. M. A. *et al.* IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
235. Griffiths, D. *et al.* Multilocus sequence typing of clostridium difficile. *J. Clin. Microbiol.* **646**, 77–90 (2010).
236. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–7 (2004).
237. Okonechnikov, K. *et al.* Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
238. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
239. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
240. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gatew. Comput. Environ. Work. GCE 2010* 1–8 (2010). doi:10.1109/GCE.2010.5676129
241. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
242. Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**, R102 (2009).
243. Cadieux, N. *et al.* Identification of the Periplasmic Cobalamin-Binding Protein BtuF of Escherichia coli. *J. Bacteriol.* **184**, 706–717 (2002).
244. Lawrence, C. C. & Stubbe, J. A. The function of adenosylcobalamin in the mechanism of ribonucleoside triphosphate reductase from Lactobacillus leichmannii. *Curr. Opin. Chem. Biol.* **2**, 650–655 (1998).
245. Sintchak, M. D., Arjara, G., Kellogg, B. A., Stubbe, J. A. & Drennan, C. L. The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nat. Struct. Biol.* **9**, 293–300 (2002).
246. Herbert, V. Vitamin B12: plant sources, requirements, and assay. *Am. J. Clin. Nutr.* **48**, 852–858 (1988).
247. Janoir, C. *et al.* Adaptive strategies and pathogenesis of clostridium difficile from In vivo transcriptomics. *Infect. Immun.* **81**, 3757–3769 (2013).
248. Kansau, I. *et al.* Deciphering adaptation strategies of the epidemic clostridium difficile 027 strain during infection through in vivo transcriptional analysis. *PLoS One* **11**, 1–13 (2016).
249. Scaria, J. *et al.* Clostridium difficile transcriptome analysis using pig ligated loop model reveals modulation of pathways not modulated in vitro. *J. Infect. Dis.* **203**, 1613–1620 (2011).
250. Ren, M. *et al.* Phylogenomics suggests oxygen availability as a driving force in

Thaumarchaeota evolution. *ISME J.* (2019). doi:10.1038/s41396-019-0418-8
251. Santos, F. *et al.* The complete coenzyme B12 biosynthesis gene cluster of Lactobacillus reuteri CRL1098. *Microbiology* **154**, 81–93 (2008).
252. Buchrieser, C. *et al.* Comparison of the genome sequences of Listeria monocytogenes and Listeria innocua: Clues for evolution and pathogenicity. *FEMS Immunol. Med. Microbiol.* **35**, 207–213 (2003).
253. Young, D. B., Comas, I. & de Carvalho, L. P. S. Phylogenetic analysis of vitamin B12-related metabolism in Mycobacterium tuberculosis. *Front. Mol. Biosci.* **2**, 1–14 (2015).
254. Reuter, S. *et al.* Parallel independent evolution of pathogenicity within the genus Yersinia. *Proc. Natl. Acad. Sci.* **111**, 6768–6773 (2014).
255. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
256. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).
257. Zhang, W. & Lu, Z. Phylogenomic evaluation of members above the species level within the phylum Firmicutes based on conserved proteins. *Environ. Microbiol. Rep.* **7**, 273–281 (2015).
258. Fonknechten, N. *et al.* Clostridium sticklandii, a specialist in amino acid degradation:Revisiting its metabolism through its genome sequence. *BMC Genomics* **11**, 555 (2010).
259. Stadtman, T. C. & Renz, P. Anaerobic Degradation of Lysine V. Some properties of the cobamide-dependent beta-Lysine mutase of Clostridium sticklandii. *Arch. Biochem. Biophys.* **125**, 226–239 (1968).
260. Somack, R. & Costilow, R. N. Purification and Properties of a Pyridoxal Phosphate and Coenzyme B12 Dependent D-α-Ornithine 5,4-Aminomutase. *Biochemistry* **12**, 2597–2604 (1973).
261. Ljungdahl, L., Irion, E. & Wood, H. G. Total Synthesis of Acetate from CO,. I. Co-Methylcobyric Acid and Co- (Methyl) -5-methoxybenzimidazolylcobamide as Intermediates with Clostridium thermoaceticum. *Biochemistry* **4**, 2771–2780 (1965).
262. Pavan, M. E. *et al.* Proposal for a new classification of a deep branching bacterial phylogenetic lineage: Transfer of Coprothermobacter proteolyticus and Coprothermobacter platensis to Coprothermobacteraceae fam. nov., within Coprothermobacterales ord. nov., Coprothermobacte. *Int. J. Syst. Evol. Microbiol.* **68**, 1627–1632 (2018).
263. Meehan, C. J. & Beiko, R. G. A phylogenomic view of ecological specialization in the lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* **6**, 703–713 (2014).
264. De Vos, P. *et al. Bergey's Manual of Systematic Bacteriology - Vol 3: The Firmicutes*. *Springer-Verlag New York Inc.* (2009). doi:10.1007/b92997
265. Scaria, J. *et al.* Comparative genomic and phenomic analysis of Clostridium difficile and Clostridium sordellii, two related pathogens with differing host tissue preference. *BMC Genomics* **16**, (2015).
266. Murphy, E. C. & Frick, I. M. Gram-positive anaerobic cocci - commensals and opportunistic pathogens. *FEMS Microbiol. Rev.* **37**, 520–553 (2013).
267. Helliwell, K. E. *et al.* Fundamental shift in vitamin B12 eco-physiology of a model alga demonstrated by experimental evolution. *ISME J.* **9**, 1446–1455 (2014).