



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

Amanda N. Shelton<sup>1</sup> · Erica C. Seth<sup>1</sup> · Kenny C. Mok<sup>1</sup> · Andrew W. Han<sup>1,2</sup> · Samantha N. Jackson<sup>3,4</sup> · David R. Haft<sup>3</sup> · Michiko E. Taga<sup>1</sup>

Received: 7 June 2018 / Revised: 14 September 2018 / Accepted: 4 October 2018 / Published online: 14 November 2018  
© The Author(s) 2018. This article is published with open access

## Abstract

The vitamin B<sub>12</sub> 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, but 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. Bacteria with a partial cobamide biosynthesis pathway include those in 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 [1]. By

understanding which organisms require nutrients and which can produce them, we can predict specific metabolic interactions between members of a microbial community [2]. 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<sub>12</sub> 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 [1, 3, 4]. 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 [5] (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 [6].

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41396-018-0304-9>) contains supplementary material, which is available to authorized users.

---

✉ Michiko E. Taga  
taga@berkeley.edu

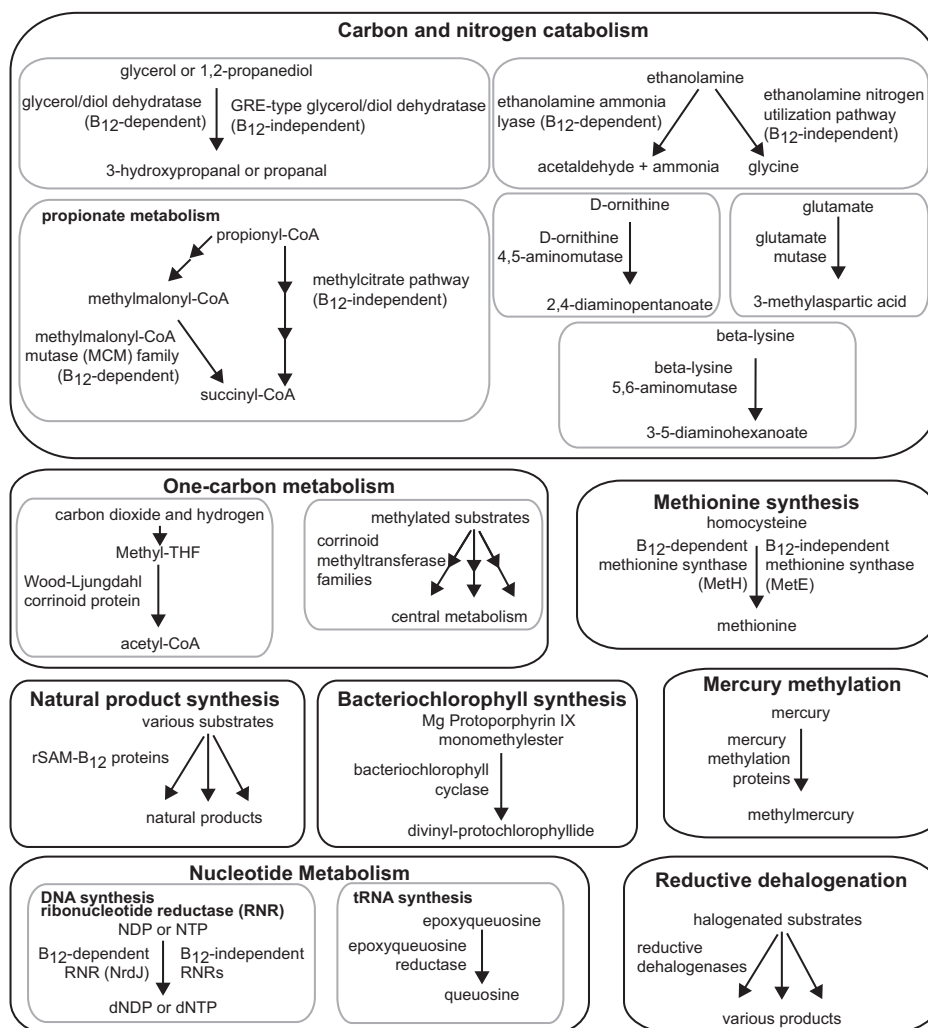
<sup>1</sup> Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

<sup>2</sup> Second Genome, Inc., South San Francisco, CA, USA

<sup>3</sup> J. Craig Venter Institute, Rockville, MD, USA

<sup>4</sup> Present address: Department of Biological & Environmental Engineering, Cornell University, Ithaca, NY, USA

**Fig. 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 Supplementary Table 4. Tetrahydrofolate (THF), radical S-adenosylmethionine (rSAM)

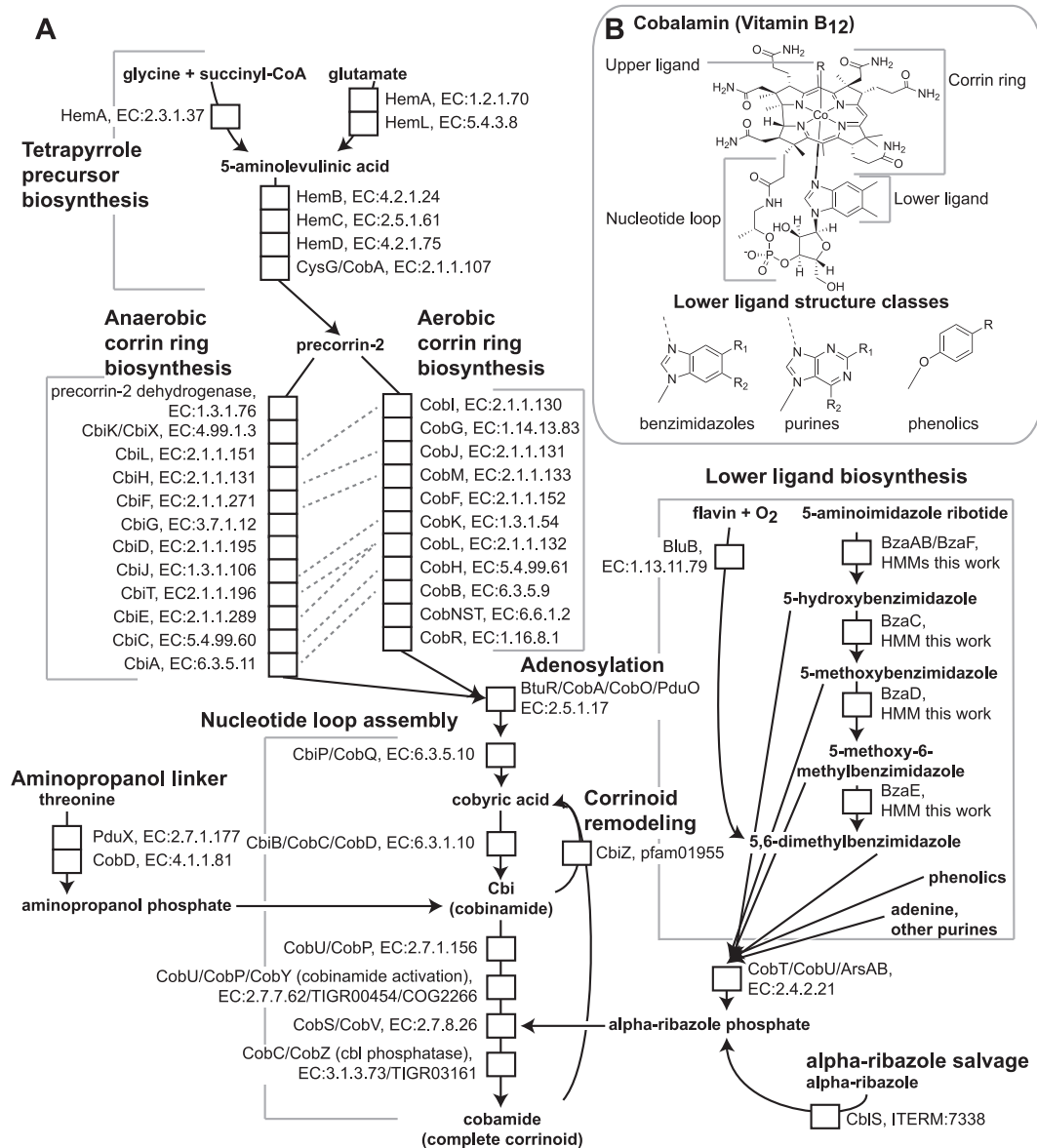


Cobamides are also used for environmentally and industrially important processes such as reductive dehalogenation and natural product synthesis [7, 8].

De novo cobamide biosynthesis involves approximately 30 steps [9], 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 adenylation 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 [10–16]. Recent work has identified many of the genetic determinants for the biosynthesis of the benzimidazole class of lower ligands [17–21] and attachment of phenolic lower ligands [22, 23] (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 [24, 25], suggesting that cobamide salvaging may be widespread in microbial communities. Analyses of cobamide biosynthesis in the human gut [10, 26] 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



**Fig. 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 a 5'-deoxyadenosyl or methyl group. Classes of possible lower ligand structures are also shown. Benzimidazoles: R<sub>1</sub>, R<sub>2</sub> = H, OH, CH<sub>3</sub>, OCH<sub>3</sub>; Purines: R<sub>1</sub> = H, CH<sub>3</sub>, NH<sub>2</sub>; R<sub>2</sub> = H, NH<sub>2</sub>, OH, O. Phenolics: R = H, CH<sub>3</sub>.

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.

## 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/IMG/M ER, <https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>) [27] classified as “finished” (accessed 11 January 2017) or “permanent draft” (accessed 23 February 2017) were downloaded (Supplementary Table 1, Sheet 1). To assess genome completeness, we searched for 55 single copy gene annotations [28, 29] using the “function profile: genomes vs functions” tool in each genome (Supplementary Table 1, Sheet 4). 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) (Supplementary Table 3). We removed 2776 genomes with fewer than 45 out of 55 unique single copy genes (Supplementary Fig. 1). 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 that 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 (Supplementary Table 2).

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

Annotations from Enzyme Commission (EC) numbers (<http://www.sbc.su.se/iubmb/enzyme/>), Pfam, TIGRFAM, Clusters of Orthologous Groups (COG), and IMG Terms [27, 30–33] for cobamide biosynthesis, cobamide-dependent enzymes, and cobamide-independent alternative annotations were chosen. These included annotations used by Degnan et al. [10], but in other cases alternative annotations were chosen to improve specificity of the identified genes (Supplementary Table 4). For example, EC: 4.2.1.30 for glycerol dehydratase identifies both cobamide-dependent and -independent isozymes, and hence 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) (Supplementary Table 1, 2 sheet 2).

For genes without functional annotations in the IMG/M ER database, we chose sequences that were genetically or biochemically characterized [34–37] to use as the query genes in one-way BLASTP [38] against the filtered genomes using the IMG/M ER “gene profile: genomes vs genes” tool, accessed Jan–May 2017 (Supplementary Table 4).

Output files for the cobamide genes were combined into a master file in Microsoft Excel (Supplementary Table 1, 2 sheet 2). 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 (Supplementary Table 5) based on the annotations in Supplementary Table 4. We then created criteria for seven cobamide biosynthesis phenotypes based on the presence of certain sets of cobamide biosynthesis genes (Supplementary Table 7): 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 (Supplementary Table 5). 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 [22, 39]. To distinguish putative *arsAB* homologs from other *cobT* homologs that are not known to produce phenolyl cobamides, IMG/M ER 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 (AA) residues, indicating annotation errors (CobT is approximately 350 AA residues) (Supplementary Table 9). 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 [38, 40]. 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 [41]. All six HMMs have not been assigned TIGRFAM accessions at the time of publication, but will be included in the next TIGRFAM release, and are included as Supplementary HMM Files. Details for each protein are listed in the Supplementary Materials and Methods. Protein sequences for 10,591 of the filtered genomes were queried for each *bza* HMM using *hmm3search* (HMMER3.1)[96].

Hits are only reported above the trusted cutoff defined for each HMM (Supplementary Table 8). 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 [19].

We used BLASTP on IMG/M ER 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 open reading frames near any BLASTP hits in the IMG/M ER genome browser. After this analysis, 180 species remained (Supplementary Table 10). Genomes were classified as a particular type of tetrapyrrole precursor salvager only if they were missing all genes upstream of a precursor.

### Strains and growth conditions

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

*Desulfotomaculum reducens* MI-1, *Listeria monocytogenes*, *Blautia hydrogenotrophica* DSM 10507, *Clostridium kluyveri* DSM 555, and *Clostridium phytofermentans* ISDg were grown anaerobically. Details of the growth conditions are listed in the Supplementary Materials and Methods.

### Corrinoid extraction and analysis

Corrinoid extractions were performed as previously described [16]. 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  $\mu$ m, 4.6  $\times$  150 mm) at 35  $^{\circ}$ 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  $^{\circ}$ 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  $\mu$ m, 4.6  $\times$  150 mm) with 1 mL/min flow rate at 30  $^{\circ}$ C. The samples were separated with a gradient of 25 to 34% acidified (0.1% formic acid) methanol over 11 min, followed by 34 to 50% over 2 min, and 50 to 75% over 9 min.

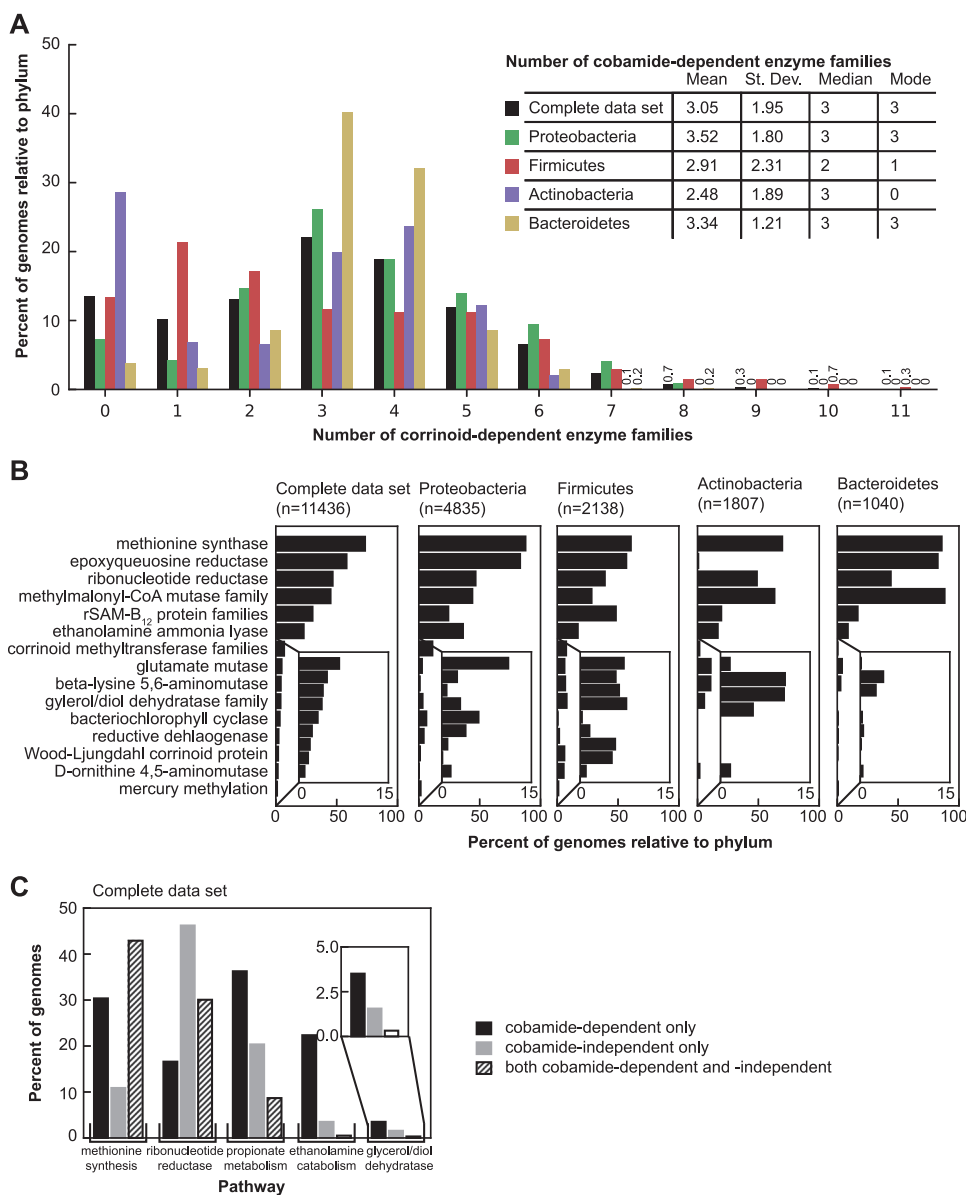
## Results

### 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 and Supplementary Table 4, and 88% of these species have more than one family (Fig. 3a). This is consistent with previous analyses of smaller data sets [10, 24, 25]. 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, Supplementary Table 5). 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).



**Fig. 3** Cobamide dependence in bacteria. **a** Histogram of the number of cobamide-dependent enzyme families (shown in Fig. 1, Supplementary Table 4) 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 (St. Dev.), 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



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 (Supplementary Figure 2).

### 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 [9], 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 [42, 43], 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 1, Supplementary Table 6), including five tested in this study (Table 1, bold names, Supplementary Figure 3). We identified a core set of eight annotations shared by all or all except one of the genomes of cobamide-producing bacteria (Table 1, 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 1, Fig. 2a). An additional five nucleotide loop assembly annotations were also highly abundant in these genomes (Table 1).

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 1; 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 1). 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 [44].

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 (Supplementary Table 7). Annotations that are absent from the majority of genomes of experimentally verified cobamide producers (*cobR*, *pduX*, and *cobD*) (Table 1, Fig. 2a), as well as one whose role in cobamide biosynthesis has not been determined (*cobW*) [45], were excluded from these threshold-based definitions. We did not exclusively use the small set of core annotations identified in Table 1 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. 4, 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 (Supplementary Figure 4).

To assess whether the core corrin ring annotations (Table 1, 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 2). The presence of two or all three marker annotations matches the threshold-based predictions even more closely (Table 2). The corrin ring markers chosen in Table 2 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 [46]; 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 (Supplementary Table 5).

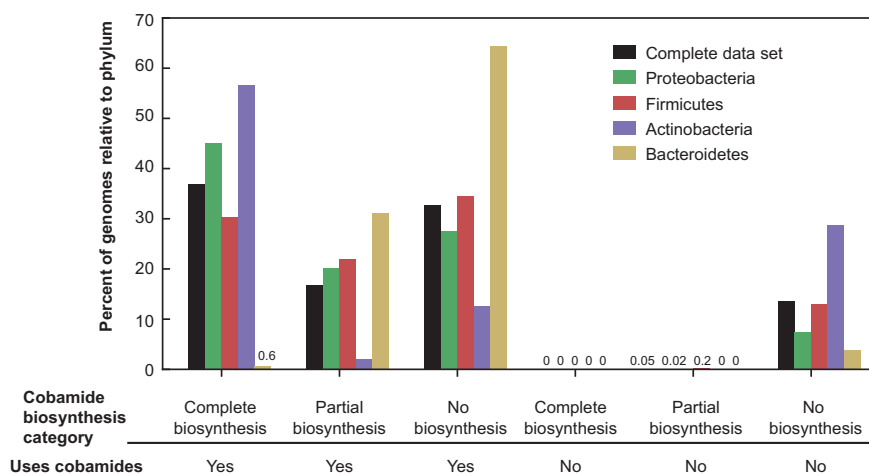
As with the cobamide-dependent enzyme families, the four major phyla in the data set have notable differences in their predicted cobamide biosynthesis phenotypes (Fig. 4). Around half of Actinobacteria (57%) and Proteobacteria (45%) and 30% of Firmicutes are predicted to be cobamide producers. In contrast, 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; Supplementary Table 10), whereas very few Actinobacterial species are predicted to salvage precursors (Fig. 4). These divisions reveal potential cobamide and cobamide precursor requirements across phyla.

## 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* [17–19, 21, 22, 39, 47, 48]. We first defined predictions for the







**Fig. 4** 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 cobamide biosynthesis phenotypes based on the criteria listed in Supplementary Table 7. The “Partial biosynthesis” category includes cobinamide (Cbi) salvagers

and tetrapyrrole precursor salvagers. The “Uses cobamides” category is defined as having one or more of the cobamide-dependent enzyme families shown in Fig. 1. The numbers are given for bars that are not visible

production [17, 21, 49] 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. 5a), particularly in Proteobacteria (Fig. 5b).

Anaerobic biosynthesis of DMB and three other benzimidazoles requires different combinations of the *bza* genes as shown in Figs. 2a and 5c [19, 20]. Because annotations for the majority of the *bza* genes were not available, we developed profile HMMs to search for them (see Supplementary Materials and Methods, Supplementary Files). 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. 5c, Supplementary Table 8). As seen with *bluB*, anaerobic benzimidazole biosynthesis genes are highly enriched in cobamide producers (Fig. 5a). 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. 5c). 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 [22]. 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 [50, 51] (Fig. 5a, b, Supplementary Table 9).

Forty-two percent of predicted cobamide producers in the data set do not have any of the benzimidazole biosynthesis or phenolic attachment genes (Fig. 5a). However, bacteria that have the  $\alpha$ -ribazole kinase CblS (Fig. 5a, 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,  $\alpha$ -ribazole salvaging); we found CblS in 363 species (3.2%), mostly in the Firmicutes phylum (Fig. 5a, b, inner rings) [42, 52]. A higher proportion of bacteria, 1041 species (9.1%), have a CbiZ annotation (Fig. 5a, b, outer rings), an amidohydrolase that cleaves the nucleotide loop, allowing cells to rebuild a cobamide with a different lower ligand [53] (Fig. 2a, corrinoid remodeling). CbiZ is found in genomes of predicted cobamide producers and Cbi auxotrophs (see following section) (Fig. 5a), as expected based on experimental studies [16, 54–56]. The reliance of some bacteria on exogenous lower ligands or  $\alpha$ -ribazoles produced by other organisms precludes prediction of cobamide structure in all cases.

### 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. 6a). As demonstrated in *Escherichia coli* [57], *Thermotoga lettingae* [58], and

**Table 2** 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 <sup>a</sup>	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	

<sup>a</sup>Numbers represent the percent of genomes containing each marker annotation and combinations of annotations within each cobamide biosynthesis category

*Dehalococcoides mccartyi* [16], and predicted in human gut microbes [10], Cbi salvagers can take up the late intermediate Cbi, assemble the nucleotide loop, and attach a lower ligand.

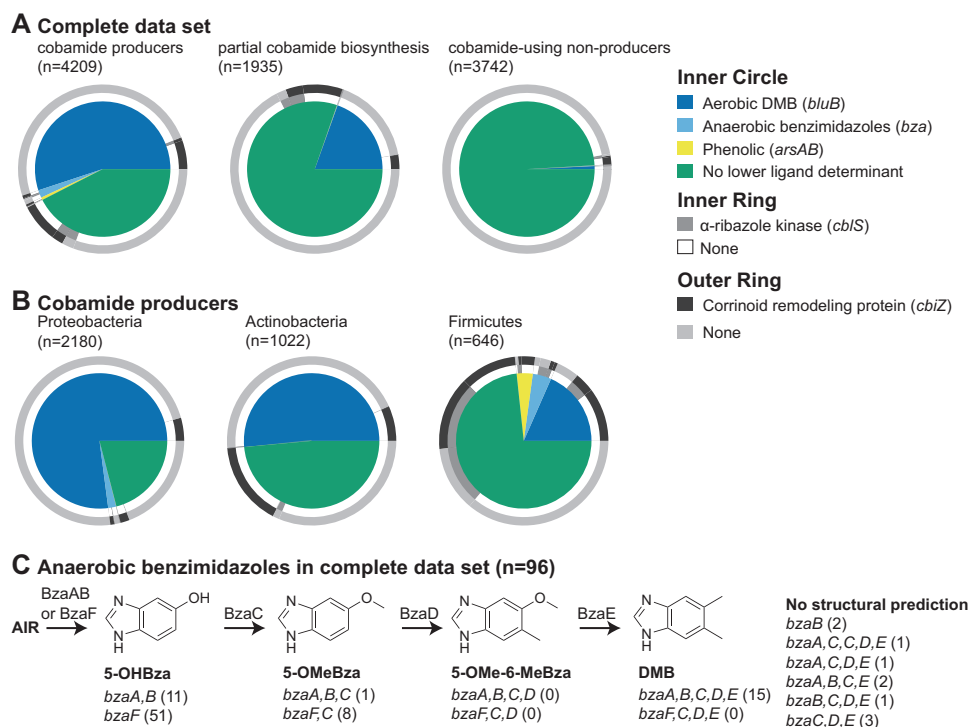
We observed an additional 201 genomes (1.7%) that lack one or more initial steps in tetrapyrrole precursor biosynthesis but have complete corrin ring biosynthesis and nucleotide loop assembly pathways, primarily in the Firmicutes (Supplementary Table 7). After searching these genomes manually for genes missing from the pathway, we designated 180 of these species as tetrapyrrole precursor salvagers, a new classification of cobamide intermediate auxotrophs (Fig. 6a, Supplementary Table 10). These organisms are predicted to produce cobamides only when provided with a tetrapyrrole precursor or a later intermediate in the pathway.

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

The identification of putative tetrapyrrole precursor salvagers suggests that either 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. *C. scindens* and *C. sporogenes*, which are predicted to require ALA, produced corrinoids in defined media only when ALA was supplied, suggesting that they do not have a novel ALA biosynthesis pathway (Fig. 6b). 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. 6b). 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. 6c). Together, these results support the hypothesis that predicted ALA salvagers synthesize cobamides by taking up ALA from the environment.

### Discussion

Vitamin B<sub>12</sub> 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



**Fig. 5** Lower ligand structure predictions. **a, b** Proportion of genomes containing the indicated lower ligand structure determinants (inner circle),  $\alpha$ -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.

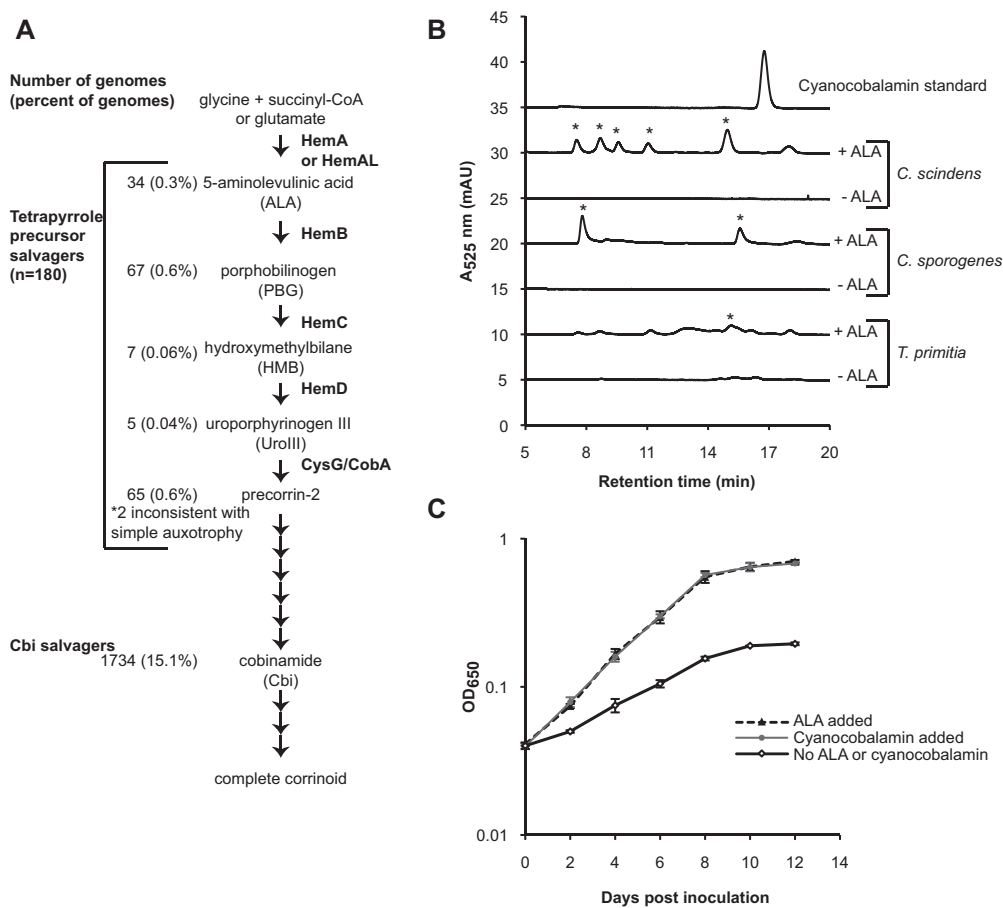
The genes required to produce each benzimidazole are shown below each structure, with the number of genomes in the complete filtered data set containing each combination of genes in parentheses. The sets of *bza* genes that do not have a predicted structure are listed on the right. Aminoimidazole ribotide (AIR), 5-hydroxybenzimidazole (5-OHBza), 5-methoxybenzimidazole (5-OMeBza), 5-methoxy-6-methylbenzimidazole (5-OMe-6-MeBza)

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 [10, 24, 25]. 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 highlight 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 [59], 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 [60–62] and that cobamide-independent RNRs are active in a limited range of oxygen concentrations [63, 64].

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 [65]. The use of experimental data gives confidence to our predictions and allowed identification of marker genes for cobamide biosynthesis. Nevertheless, our predictions likely



**Fig. 6** Characterization of putative tetrapyrrole precursor salvagers **a** Steps in cobamide biosynthesis. The enzymes 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 (\*). Specific tetrapyrrole precursor salvager genomes are listed in Supplementary Table 10. **b** HPLC analysis of corrinoid extracts from *Clostridium scindens*, *Clostridium sporogenes*,

and *Treponema primitia* grown with and without added ALA. A cyanocobalamin standard (200 pmoles) is shown for comparison. Asterisks denote peaks with ultraviolet-visible (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

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 [5], and cobamide biosynthesis operons are commonly subjected to negative regulation by riboswitches [24, 66]. The abundance of cobamide importers [10, 24, 25], 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 [67, 68], 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 [10–16], 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 signature genes specific to nor-cobamide biosynthesis [69, 70]. 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, 39, 71]. 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 [39, 48, 72]. The presence of free benzimidazoles and  $\alpha$ -ribazoles in microbial communities [73–75] and the ability of bacteria to take up and incorporate these compounds into cobamides [13, 72, 76, 77] 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 trichloroethylene (TCE)-contaminated groundwater enrichment culture, 5-hydroxybenzimidazolyl cobamide and *p*-cresolyl cobamide were the most abundant cobamides [50], compared to cobalamin in bovine rumen [78] and 2-methyladeninyl cobamide in human stool [71]. 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 [10].

Dependence on biosynthetic precursors has been observed or predicted for amino acids, nucleotides, and the cofactors thiamin and folate [79–82]. Here, we describe genomic evidence for dependence on cobamide precursors, namely Cbi or tetrapyrrole precursors. The prevalence of Cbi-salvaging bacteria (Fig. 6a) 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 [50], and

represented up to 12.8% of the total corrinoids detected in human stool samples [71].

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 [83]. 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 [11, 24–26]. 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 [84, 85] and ALA has been found in swine manure extract [86]. 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 [87] or bioassay with *Rhodobacter sphaeroides hemAT1* [88], which lacks ALA synthase, suggesting that either 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 data set. 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 [89, 90] 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 [91]. Genome analysis suggests that *Candidatus Hodgkinia cicadicola*, a predicted UroIII salvager [92], 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 [10].

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 [93]. 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 [94, 95]. 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.

**Acknowledgements** This work was supported by the NIH grant A1117984 and a Hellman Family Faculty Fund award to MET, a Grace Kase Graduate Fellowship to ANS, an NSF Graduate Fellowship to ECS, and NSF grant No. 1458808 to the J. Craig Venter Institute. We thank Alison Smith, Johan Kudahl, Jared Leadbetter, Amrita Hazra, Yamini Mathur, and members of the Taga lab for helpful advice. We thank Olga Sokolovskaya, Sebastian Gude, Zachary Hallberg, and Alexa Nicolas for critical reading for this manuscript. We thank Jared Leadbetter, Rizlan Bernier-Latmani, Daniel Portnoy, Rolf Thauer, Susan Leschine, and Jill Zeilstra-Ryalls for providing bacterial strains.

**Author contributions** ANS, ECS, and AWH performed the bioinformatic analysis. SNJ and DRH created the BzaABCDEF HMMs. KCM, ECS, and ANS performed growth assays and corrinoid extractions.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Seth EC, Taga ME. Nutrient cross-feeding in the microbial world. *Front Microbiol.* 2014;5:350.
- Abreu NA, Taga ME. Decoding molecular interactions in microbial communities. *FEMS Microbiol Rev.* 2016;40:648–63.
- Degnan PH, Taga ME, Goodman AL. Vitamin B<sub>12</sub> as a modulator of gut microbial ecology. *Cell Metab.* 2014b;20:769–78.
- Helliwell KE. The roles of B vitamins in phytoplankton nutrition: new perspectives and prospects. *New Phytol.* 2017;216:62–68.
- Roth JR, Lawrence JG, Bobik TA. Cobalamin (coenzyme B<sub>12</sub>): synthesis and biological significance. *Annu Rev Microbiol.* 1996; 50:137–81.
- Ragsdale SW, Pierce E. Acetogenesis and the Wood-Ljungdahl pathway of CO<sub>2</sub> fixation. *Biochim Biophys Acta.* 2008;1784: 1873–98.
- Banerjee R, Ragsdale SW. The many faces of vitamin B<sub>12</sub>: catalysis by cobalamin-dependent enzymes. *Annu Rev Biochem.* 2003;72:209–47.
- Broderick JB, Du BR, Duschene KS, Shepard EM. Radical S-adenosylmethionine enzymes. *Chem Rev.* 2014;114:4229–317.
- Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC. The biosynthesis of adenosylcobalamin (vitamin B<sub>12</sub>). *Nat Prod Rep.* 2002;19:390–412.
- Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. Human gut microbes use multiple transporters to distinguish vitamin B<sub>12</sub> analogs and compete in the gut. *Cell Host Microbe.* 2014;15:47–57.
- Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, et al. Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B<sub>12</sub>. *Curr Biol.* 2016;26:999–1008.
- Keller S, Kunze C, Bommer M, Paetz C, Menezes RC, Svatoš A, et al. Selective utilization of benzimidazolyl-norcobamides as cofactors by the tetrachloroethene reductive dehalogenase of *Sulfurospirillum multivorans*. *J Bacteriol.* 2018;200:e00584–17.
- Mok KC, Taga ME. Growth inhibition of *Sporomusa ovata* by incorporation of benzimidazole bases into cobamides. *J Bacteriol.* 2013;195:1902–11.
- Yan J, Ritalahti KM, Wagner DD, Löffler FE. Unexpected specificity of interspecies cobamide transfer from *Geobacter* spp. to organohalide-respiring *Dehalococcoides mccartyi* strains. *Appl Environ Microbiol.* 2012;78:6630–6.
- Yan J, Şimşir B, Farmer AT, Bi M, Yang Y, Campagna SR, et al. The corrinoid cofactor of reductive dehalogenases affects dechlorination rates and extents in organohalide-respiring *Dehalococcoides mccartyi*. *ISME J.* 2016;10:1092–101.
- Yi S, Seth E, Men Y, Stabler SP, Allen RH, Alvarez-Cohen L, et al. Versatility in corrinoid salvaging and remodeling pathways supports corrinoid-dependent metabolism in *Dehalococcoides mccartyi*. *Appl Environ Microbiol.* 2012;78:7745–52.
- Campbell GRO, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, et al. *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B<sub>12</sub>. *Proc Natl Acad Sci USA.* 2006;103:4634–9.
- Gray MJ, Escalante-Semerena JC. Single-enzyme conversion of FMNH<sub>2</sub> to 5, 6-dimethylbenzimidazole the lower ligand of B<sub>12</sub>. *Proc Natl Acad Sci USA.* 2007;104:2921–6.
- Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy V, Begley TP, et al. Anaerobic biosynthesis of the lower ligand of vitamin B<sub>12</sub>. *Proc Natl Acad Sci USA.* 2015;112:10792–7.
- Mehta AP, Abdelwahed SH, Fenwick MK, Hazra AB, Taga ME, Zhang Y, et al. Anaerobic 5-hydroxybenzimidazole formation from aminoimidazole ribotide: an unanticipated intersection of thiamin and vitamin B<sub>12</sub> biosynthesis. *J Am Chem Soc.* 2015; 137:10444–7.
- Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC. BluB cannibalizes flavin to form the lower ligand of vitamin B<sub>12</sub>. *Nature.* 2007;446:449–53.
- Chan CH, Escalante-Semerena JC. ArsAB, a novel enzyme from *Sporomusa ovata* activates phenolic bases for adenosylcobamide biosynthesis. *Mol Microbiol.* 2011;81:952–67.
- Newmister SA, Chan CH, Escalante-Semerena JC, Rayment I. Structural insights into the function of the nicotinate mononucleotide:phenol/p-cresol phosphoribosyltransferase (ArsAB) enzyme from *Sporomusa ovata*. *Biochemistry.* 2012;51:8571–82.

24. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of the vitamin B<sub>12</sub> metabolism and regulation in prokaryotes. *J Biol Chem.* 2003;278:41148–59.
25. Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN. Comparative genomic analyses of nickel, cobalt and vitamin B<sub>12</sub> utilization. *BMC Genomics.* 2009;10:78.
26. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet.* 2015;6:148.
27. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 2012;40:D115–22.
28. Raes J, Korbil JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. *Genome Biol.* 2007;8:R10.
29. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature.* 2015;523:208–11.
30. Cornish-Bowden A. Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspect Sci.* 2014;1:74–87.
31. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44:D279–85.
32. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 2015;43:D261–9.
33. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 2012;41:387–95.
34. Parks JM, Johs A, Podar M, Bridou R, Hurt RA, Smith SD, et al. The genetic basis for bacterial mercury methylation. *Science.* 2013;339:1332–5.
35. Brien JRO, Raynaud C, Croux C, Girbal L, Soucaille P, Lanzilotta WN, et al. Insight into the mechanism of the B<sub>12</sub>-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry.* 2004;43:4635–45.
36. Vetting MW, Al-Obaidi N, Zhao S, San Francisco B, Kim J, Wichelecki DJ, et al. Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry.* 2015;54:909–31.
37. Zarzycki J, Sutter M, Cortina NS, Erb TJ, Kerfeld CA. In vitro characterization and concerted function of three core enzymes of a glycyl radical enzyme - associated bacterial microcompartment. *Sci Rep.* 2017;7:42757
38. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–402.
39. Yan J, Bi M, Bourdon AK, Farmer AT, Wang P-H, Molenda O, et al. Purinyl-cobamide is a native prosthetic group of reductive dehalogenases. *Nat Chem Biol.* 2018;14:8–14.
40. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25:1189–91.
41. Haft DR, Haft DH. A comprehensive software suite for protein family construction and functional site prediction. *PLoS ONE.* 2017;12:e0171758.
42. Gray MJ, Escalante-Semerena JC. A new pathway for the synthesis of alpha-ribazole-phosphate in *Listeria innocua*. *Mol Microbiol.* 2010;77:1429–38.
43. McGoldrick HM, Roessner CA, Raux E, Lawrence AD, McLean KJ, Munro AW, et al. Identification and characterization of a novel vitamin B<sub>12</sub> (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. *J Biol Chem.* 2005;280:1086–94.
44. Schubert HL, Raux E, Wilson KS, Warren MJ. Common chelatase design in the branched tetrapyrrole pathways of heme and anaerobic cobalamin synthesis. *Biochemistry.* 1999;38:10660–9.
45. Haas CE, Rodionov DA, Kropat J, Malasarn D, Merchant SS, de Crécy-Lagard V. A subset of the diverse COG0523 family of putative metal chaperones is linked to zinc homeostasis in all kingdoms of life. *BMC Genom.* 2009;10:470.
46. Bertrand EM, Saito MA, Jeon YJ, Neilan BA. Vitamin B<sub>12</sub> biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B<sub>12</sub> biosynthesizers. *Environ Microbiol.* 2011;13:1285–98.
47. Crofts TS, Hazra AB, Tran JLA, Sokolovskaya OM, Osadchiy V, Ad O, et al. Regiospecific formation of cobamide isomers is directed by CobT. *Biochemistry.* 2014;53:7805–15.
48. Hazra AB, Tran JLA, Crofts TS, Taga ME. Analysis of substrate specificity in CobT homologs reveals widespread preference for DMB, the lower axial ligand of vitamin B<sub>12</sub>. *Chem Biol.* 2013;20:1275–85.
49. Hazra AB, Ballou DP, Taga ME. Unique biochemical and sequence features enable BluB to destroy flavin and distinguish BluB from the flavin monooxygenase superfamily. *Biochemistry.* 2018;57:1748–57.
50. Men Y, Seth EC, Yi S, Crofts TS, Allen RH, Taga ME, et al. Identification of specific corrinoids reveals corrinoid modification in dechlorinating microbial communities. *Environ Microbiol.* 2014;17:4873–84.
51. Stupperich E, Eisinger HJ. Biosynthesis of para-cresolyl cobamide in *Sporomusa ovata*. *Arch Microbiol.* 1989;151:372–7.
52. Mattes TA, Escalante-Semerena JC. *Salmonella enterica* synthesizes 5,6-dimethylbenzimidazolyl-(DMB)- $\alpha$ -riboside. Why some Firmicutes do not require the canonical DMB activation system to synthesize adenosylcobalamin. *Mol Microbiol.* 2017;103:269–81.
53. Woodson JD, Escalante-Semerena JC. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B<sub>12</sub> precursor cobinamide in archaea. *Proc Natl Acad Sci USA.* 2004;101:3591–6.
54. Gray MJ, Escalante-Semerena JC. In vivo analysis of cobinamide salvaging in *Rhodobacter sphaeroides* strain 2.4.1. *J Bacteriol.* 2009;191:3842–51.
55. Gray MJ, Escalante-Semerena JC. The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: a critical activity of the cobamide remodelling system of *Rhodobacter sphaeroides*. *Mol Microbiol.* 2009;74:1198–210.
56. Men Y, Seth EC, Yi S, Allen RH, Taga ME, Alvarez-Cohen L. Sustainable growth of *Dehalococcoides mccartyi* 195 by corrinoid salvaging and remodeling in defined lactate-fermenting consortia. *Appl Environ Microbiol.* 2014;80:2133–41.
57. Di Girolamo PM, Bradbeer C. Transport of vitamin B<sub>12</sub> in *Escherichia coli*. *J Bacteriol.* 1976;106:745–50.
58. Butzin NC, Secinaro MA, Swithers KS, Gogarten JP, Noll KM. *Thermotoga lettingae* can salvage cobinamide to synthesize vitamin B<sub>12</sub>. *Appl Environ Microbiol.* 2013;79:7006–12.
59. Zallot R, Ross R, Chen WH, Bruner SD, Limbach PA, de Crécy-Lagard V. Identification of a novel epoxyqueuosine reductase family by comparative genomics. *ACS Chem Biol.* 2017;12:844–51.
60. Gonzalez JC, Banerjee RV, Huang S, Sumner JS, Matthews RG. Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: two solutions to the same chemical problem. *Biochemistry.* 1992;31:6045–56.

61. Hondorp ER, Matthews RG. Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in *Escherichia coli*. PLoS Biol. 2004;2:e336.
62. Xie B, Bishop S, Stessman D, Wright D, Spalding MH, Halverson LJ. *Chlamydomonas reinhardtii* thermal tolerance enhancement mediated by a mutualistic interaction with vitamin B<sub>12</sub>-producing bacteria. ISME J. 2013;7:1544–55.
63. Fontecave M. Ribonucleotide reductases and radical reactions. Cell Mol Life Sci. 1998;54:684–95.
64. Taga ME, Walker GC. *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. MPMI. 2010;23:1643–54.
65. Tavares NK, Zayas CL, Escalante-Semerena JC. The *Methanosarcina mazei* MM2060 gene encodes a bifunctional kinase/decarboxylase enzyme involved in cobamide biosynthesis. Biochemistry. 2018;57:4478–95.
66. Nahvi A. Coenzyme B<sub>12</sub> riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. 2004;32:143–50.
67. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. Algae acquire vitamin B<sub>12</sub> through a symbiotic relationship with bacteria. Nature. 2005;438:90–3.
68. Kazamia E, Czesnick H, Van Nguyen TT, Croft MT, Sherwood E, Sasso S, et al. Mutualistic interactions between vitamin B<sub>12</sub>-dependent algae and heterotrophic bacteria exhibit regulation. Environ Microbiol. 2012;14:1466–76.
69. Keller S, Treder A, Von Reuss SH, Escalante-Semerena JC, Schubert T. The SMUL\_1544 gene product governs norcobamide biosynthesis in the tetrachloroethene-respiring bacterium *Sulfurospirillum multivorans*. J Bacteriol. 2016;198:2236–43.
70. Kräutler B, Fieber W, Ostermann S, Fasching M, Ongania KH, Gruber K, et al. The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is nNorpseudo-B<sub>12</sub>, a new type of a natural corrinoid. Helv Chim Acta. 2003;86:3698–716.
71. Allen RH, Stabler SP. Identification and quantitation of cobalamin and cobalamin analogues in human feces. Am J Clin Nutr. 2008;87:1324–35.
72. Crofts TS, Seth EC, Hazra AB, Taga ME. Cobamide structure depends on both lower ligand availability and CobT substrate specificity. Chem Biol. 2013;20:1265–74.
73. Crofts TS, Men Y, Alvarez-Cohen L, Taga ME. A bioassay for the detection of benzimidazoles reveals their presence in a range of environmental samples. Front Microbiol. 2014;5:592.
74. Johnson WM, Kido Soule MC, Kujawinski EB. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. ISME J. 2016;10:2304–16.
75. Wienhausen G, Noriega-Ortega BE, Niggemann J, Dittmar T, Simon M. The exometabolome of two model strains of the *Roseobacter* group: a marketplace of microbial metabolites. Front Microbiol. 2017;8:1985.
76. Anderson PJ, Lango J, Carkeet C, Britten A, Kräutler B, Hammock BD, et al. One pathway can incorporate either adenine or dimethylbenzimidazole as an alpha-axial ligand of B<sub>12</sub> cofactors in *Salmonella enterica*. J Bacteriol. 2008;190:1160–71.
77. Keller S, Ruetz M, Kunze C, Kräutler B, Diekert G, Schubert T. Exogenous 5,6-dimethylbenzimidazole caused production of a non-functional tetrachloroethene reductive dehalogenase in *Sulfurospirillum multivorans*. Environ Microbiol. 2013;16:3361–9.
78. Girard CL, Santschi DE, Stabler SP, Allen RH. Apparent ruminal synthesis and intestinal disappearance of vitamin B<sub>12</sub> and its analogs in dairy cows. J Dairy Sci. 2009;92:4524–9.
79. de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noiriell A, Hanson AD. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics. 2007;8:245.
80. Kilstrup M, Hammer K, Jensen PR, Martinussen J. Nucleotide metabolism and its control in lactic acid bacteria. FEMS Microbiol Rev. 2005;29:555–90.
81. Paerl RW, Bouget F-Y, Lozano J-C, Vergé V, Schatt P, Allen EE, et al. Use of plankton-derived vitamin B<sub>1</sub> precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. ISME J. 2016;11:753–65.
82. Sloan DB, Moran NA. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Mol Biol Evol. 2012;29:3781–92.
83. Roper JM, Raux E, Brindley AA, Schubert HL, Gharbia SE, Shah HN, et al. The enigma of cobalamin (vitamin B<sub>12</sub>) biosynthesis in *Porphyromonas gingivalis*: Identification and characterization of a functional corrin pathway. J Biol Chem. 2000;275:40316–23.
84. Dobriner K. Porphyrin excretion in the feces in normal and pathological conditions. J Biol Chem. 1937;120:115–28.
85. Watson CJ, 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. 1945;157:345–62.
86. Kanto U, Jutamanee K, Osotsapar Y, Chai-arree W, Jintanawich W, Promdang S, et al. Quantification of 5-aminolevulinic acid in swine manure extract by HPLC-fluorescence. J Liq Chromatogr Relat Technol. 2013;36:2731–48.
87. Giuntini F, Bourré L, MacRobert AJ, Wilson M, Eggleston IM. Quantitative determination of 5-aminolaevulinic acid and its esters in cell lysates by HPLC-fluorescence. J Chromatogr B Anal Technol Biomed Life Sci. 2008;875:562–6.
88. Neidle EL, Kaplan S. Expression of the *Rhodobacter sphaeroides* *hemA* and *hemT* genes, encoding two 5-aminolevulinic acid synthase isozymes. J Bacteriol. 1993;175:2292–303.
89. Lyell NL, Septer AN, Dunn AK, Duckett D, Stoudenmire JL, Stabb EV. An expanded transposon mutant library reveals that *Vibrio fischeri* δ-aminolevulinic acid auxotrophs can colonize *Euprymna scolopes*. Appl Environ Microbiol. 2017;83:e02470–16.
90. Sangwan I, O'Brian MR. Evidence for an Inter-organismic heme biosynthetic pathway in symbiotic soybean root nodules. Science. 1991;251:1220–2.
91. Qi M, Nelson KE, Daugherty SC, Nelson WC, Hance IR, Morrison M, et al. Genomic differences between *Fibrobacter succinogenes* S85 and *Fibrobacter intestinalis* DR7, identified by suppression subtractive hybridization. Appl Environ Microbiol. 2008;74:987–93.
92. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA. 2009;106:15394–9.
93. Morris JJ, Lenski RE, Zinser ER. The black queen hypothesis: evolution of dependencies through adaptive gene loss. MBio. 2012;3:e00036–12.
94. Lawrence J, Roth J. Evolution of coenzyme B<sub>12</sub> synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. Genetics. 1996;142:11–24.
95. Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, et al. Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Res. 2008;15:151–61.
96. Eddy SR. HMMER: biosequence analysis using profile hidden Markov models (v3.1b2) [software]. 2015. Available from <http://hmmerr.org/>.